Scientific Human Capital and Institutional Incentives: The Roles of Topic Knowledge and Research Skills

Academic scientists develop deep topic knowledge in highly specialized niches. However, they also acquire extensive skills to undertake advanced research in their fields. Prior research examining how firms use scientists' human capital typically focuses on a generalist-specialist distinction in breadth of topic knowledge or a basic-applied distinction in research orientation. This paper emphasizes an alternative distinction between two dimensions of scientific human capital: topic knowledge and scientific research skills. I build on prior literature on scientific careers and the logic of scientific inquiry to analyze how each dimension provides value in corporate research. I test my arguments using longitudinal data on researchers in regenerative medicine. While scientists moving to industry have more conceptually diverse subsequent research output, there is little evidence of other changes in their research productivity, impact, or basic-applied focus. This is consistent with industry more highly valuing the wider research skills of scientists hired from academia. It contrasts with a view in which firms primarily focus on exploiting highly specialized, but narrow, topic knowledge for commercial purposes. The findings also provide insights into how the division of scientific labor between academia and industry may shape the knowledge accumulation process in areas where academic and commercial research overlap.

INTRODUCTION

An extensive body of research has documented how firms can benefit from developing advanced internal scientific capabilities (Cohen and Levinthal, 1990; Rosenberg, 1990; Gambardella, 1992; Cockburn and Henderson, 1998). Specialized knowledge can facilitate corporate innovation by guiding search processes and identifying valuable new information (Fleming and Sorenson, 2004; Arora, Belenzon, and Sheer, 2017). Scientists' knowledge can also help them develop relationships with external knowledge holders to broker valuable ideas from beyond the organizations' boundaries (Sorenson, Rivkin, and Fleming, 2006; Fleming, Mingo, and Chen 2007; Fabrizio, 2009). To facilitate knowledge acquisition, firms can hire scientists with extensive specialized training and experience from academia to build the human capital resources that provide the foundation for internal scientific capabilities (Zucker and Darby, 1997; Zucker, Darby, and Torero, 2002; Gittelman, 2005; Ejsing et al., 2013). While research has shown why firms hire and collaborate with scientists trained in academia, we know far less about how the different types of knowledge and skills that academic scientists develop contribute to research programs in corporate science and how firms exploit the human capital of these hires.

Academic scientists offer human capital that is specialized on multiple dimensions. On the one hand, academic scientists have deep knowledge in a relatively narrow niche in which they seek to contribute to advancing the frontier of knowledge in their field (Jones, 2009; Leahey, Beckman, and Stanko, 2017). However, scientists in academia also develop extensive advanced skills – often rooted in tacit knowledge – to carry out research projects that advance the knowledge frontier (Senker, 1995). These skills enable them to understand the relationships between objects of investigation through scientific experiments. A scientist with narrow topic expertise may use a wide range of different tools to analyze a narrow set of phenomena in their area of interest. Conversely, another scientist may use a narrow set of analytical tools but apply these to a wide range of topics. Past research has primarily focused on the first dimension of specialization of scientific human capital – topic knowledge – when examining the value scientific human capital contributes to firms' research and development processes (Jones, 2009; Agrawal, Goldfarb, and Teodoridis, 2016; Teodoridis, Vakili, and Bikard, 2019). This literature is driven by the question of how firms can internalize valuable topic knowledge by hiring a scientist from academia or collaborating with external experts.

Nonetheless, at the heart of our understanding of absorptive capacity is the ability to identify, evaluate, and re-use external knowledge in ways that are valuable for the firm (Cohen and Levinthal, 1990; Arora and Gambardella, 1994). Understanding and applying external scientific knowledge to new internal uses is an inherently experimental process applying internal scientists' knowledge and skills to decompose and recombine a range of ideas (Henderson, 1994; Iansiti and Clark, 1994; Senker, 1995; Fabrizio, 2009). Firms can achieve organization-level absorptive capacity across scientific topics in different ways. First, firms may employ individual scientists as internal experts in their specialized niche. In this case, individual specialization across diverse specialties provides the necessary knowledge diversity at the organization level). On the other hand, firms may apply individual scientists' human capital to research projects across a wider range of topics of interest to the firm. Here, individual diversification supports organization-level capabilities across a range of scientific topics. This paper examines how the different dimensions of scientific human capital are valued and incentivized across academic and industry institutional settings. In particular, it examines how the nature of scientists' research changes as their careers take them through transitions between academic and industry employment.

An additional literature has sought to explain why the different institutional features of academia and industry provide an appropriate setting for different types of scientific research (Bush, 1945; Nelson, 1959; Arrow, 1962; Aghion, Dewatripont, and Stein, 2008; Lacetera, 2009). Academia is seen to offer a research environment that is suited for basic research because its reward structure lowers scientists' exposure to immediate commercial pressures and emphasizes contributions to fundamental scientific understanding. Conversely, industry – as an institution – is seen to support applied research because firms' and scientists' incentives can be better aligned where there is less ex ante uncertainty about projects' value and risk that scientists pursue their private scientific interests rather than commercial value. Such higher-powered financial incentive schemes are viewed as more difficult to design, and monitor on an ongoing basis, for basic research with more unpredictable outcomes. Nonetheless, extensive past research shows that, in many fields, academic and industry scientists have significant overlap in research topics (Rosenberg and Nelson, 1994; Sauermann and Stephan, 2013; Bikard, 2018). In 'Pasteur's Quadrant' scientific discoveries may both contribute to fundamental scientific understanding of phenomena and offer promise for commercial technologies (Stokes, 1997; Bikard, 2015). Yet, if, as prior research suggests, the incentives and institutional logics governing scientific inquiry vary so much between academia and industry, this raises the question of why we see scientists working on similar research questions across both settings – and what, if any, differences there are in how they exploit and further develop their human capital.

This paper seeks to provide insights into scientists' knowledge accumulation and firms' scientific human capital strategies by analyzing how the research direction of scientists moving from academia to industry in the stem cell field changes after acquiring an industry affiliation. If firms value academic scientists' human capital to take advantage of the deep topic knowledge developed through their prior research, we would expect them to be applied to research projects on the same topics they worked on in academia. Conversely, to the extent firms value scientists because of their more general scientific research skills that can be flexibly applied across a range of research questions, we would expect to see their human capital applied more widely to new topics after transitions to industry employment. This paper also sheds light on the important implications for how the division of scientific labor between academia and industry in areas where both institutions have interests in scientific research. The role institutional factors play in the type of research scientists produce – even when working on the same type of research questions – will affect how knowledge is accumulated in those areas of science. In particular, if scientists need deep topic knowledge to push forward the knowledge frontier and make scientific breakthroughs (Jones, 2009; Kaplan and Vakili, 2015; Agrawal et al., 2016), these may be less likely to occur if the balance of scientific labor shifts heavily towards industry.

SCIENTIFIC HUMAN CAPITAL

To understand how different aspects of academic scientists' human capital are valuable to employers, we must start by understanding how academia incentivizes scientists to develop their potentially valuable knowledge and skills. The credit-based reward system in academic science prioritizes early contributions that significantly advance the scientific community's understanding of the fundamental properties of the phenomena that are of interest to a field (Merton, 1957; Dasgupta and David, 1994). As the 'burden of knowledge' required to make these types of contributions increases, scientists are incentivized to focus their research more narrowly to develop highly specialized topic knowledge that is deeper but narrower than before (Leahey, 2007; Jones, 2009; Agrawal et al., 2016). Past studies of academic scientists have largely focused on differences in scientists' human capital according to the extent to which they are generalists or specialists in the breadth of their topic knowledge (Agrawal et al., 2016; Leahey et al., 2017; Teodoridis, 2018; Teodoridis et al., 2019). Alternatively, prior studies have examined scientists' relative experience with commercial or patenting activities as distinct parts of their human capital from pure academic research (Toole and Czarnitzki, 2009; Baba, Shichijo, and Sedita, 2009; Subramanian, Lim, and Soh, 2013).

However, academic scientists also develop specialized research skills to carry out advanced research in their discipline. Some scientists may be topic generalists but apply a similar set of research skills across topics. Others may be topic specialists but use a variety of methods in their area of specialization. For example, in the social sciences, an economist with advanced research skills in microeconometric analysis could specialize in applying advanced microeconometric research skills within one sub-field of economics, such as labor economics, in which she is a topic specialist. She could apply these skills to research across multiple areas, such as labor, development, and public economics – as a researcher with wider topic knowledge but a more specialized research skill toolkit. She may focus within one sub-field of economics, as a topic specialist, and apply many different analytical skills to her research. She may have wider breadth in both her topic knowledge and research skills. Finally, she may focus her research on theoretical microeconometrics and the development of new microeconometric research tools. Similarly, in sociology, a researcher with advanced skills in social network analysis may focus on applying these in one niche or across a wide range of topics; use social network analysis alongside multiple other research skills in research on her topic of interest; or focus on developing new theory or methods in social network analysis. Topic knowledge and scientific research skills are both highly specialized forms of human capital developed by academic scientists. However, the research skills to design, execute, and analyze complex scientific projects skills may be more readily transferrable across topics in a field than the highly specialized knowledge an academic scientist has developed in a particular topic.

Prior research suggests that the logic of scientific enquiry can be different in firms and academia due divergence in the respective nature and goals of scientific research in each type of institution (Evans, 2010; Sauermann and Stephan, 2013). Differences in the nature of work between academia and industry may have important implications for the ways in which knowledge and skills are valued in each setting. Whereas the reward system of academic science privileges major discoveries about fundamental phenomena (Merton, 1957; Latour, 1987; Dasgupta and David, 1994), the returns to firms are based more on understanding relationships between objects that can subsequently be developed for commercial ends (Evans, 2010; Sauermann and Stephan, 2013). This may shape the relative returns to scientific topic knowledge and research skills in academia and industry. To be able to push the frontier of knowledge on a phenomenon, a scientist may need increasingly specialized knowledge to be at the forefront of a narrow field (Jones, 2009; Agrawal, Goldfarb, and Teodoridis, 2016). Conversely, uncovering the relational possibilities of objects may still require substantial subject knowledge but rely more heavily on applying research skills across a range of scientific objects and phenomena.

To the extent that a large part of the value scientific human capital is as a resource that provides firms with absorptive capacity, this may affect how firms value and exploit the different dimensions of scientific human capital. Identifying, interpreting, and experimenting with new ideas from outside the firm's boundary may require less deep topic knowledge than advancing the knowledge frontier within a given topic. Firms and their internal scientists may also contract or build relationships with academic scientists to access specific novel ideas as they develop in academic science, rather than hiring the holders of this knowledge (Liebeskind et al., 1996; Cockburn and Henderson, 1998; Zucker, Darby, and Armstrong, 2002; Fabrizio, 2009). Indeed, it may be very costly to acquire the wide range of topic specialists that may be needed to get organization-level diversification across the relevant range of scientific knowledge (Jones, 2009; Agrawal et al., 2016). Instead, having individual-level research diversity may be a more cost-effective way of covering the necessary range of knowledge at the organization level

Why Do Firms Hire Scientists?

An extensive literature has developed to understand why firms choose to invest in basic scientific research – despite the value of research outputs being difficult to internalize fully (Rosenberg, 1990; Hicks, 1995; Arora et al., 2017). Investment in scientific research of interest to the academic community can offer firms access to valuable external knowledge through scientific networks (Powell, Koput, and Smith-Doerr, 1996) or help them attract talent (Stern, 2004). Firms that develop their own scientific capabilities have superior capacity to absorb and apply potential valuable knowledge from beyond the firm's boundaries in academic science (Cockburn and Henderson, 1998; Gittelman and Kogut, 2003) and acquire insights into new knowledge through engagement with specialists in academia (Zucker, Darby, and Armstrong, 2002; Fabrizio, 2009). Basic research can also be used as a means to develop valuable new ideas internally that can be applied in products (Arora et al., 2017). A central element of firms' basic scientific capabilities is their human capital. In order to absorb external knowledge, firms need scientists able to digest and apply knowledge from the frontier of science and collaborate with academic scientists. Similarly, to build valuable relationships and participate in networks with academic scientists, firms need scientists to be sufficiently engaged in the wider scientific community – for example, through publishing, co-authoring, and attending scientific conferences (Liebeskind et al., 1996; Cockburn and Henderson, 1998; Simeth and Raffo, 2013).

Hiring workers who have human capital in relevant domains can provide firms scientific or technological knowledge in new areas (Zucker and Darby, 1997; Almeida and Kogut, 1999; Zucker, Darby, and Torero, 2002; Rosenkopf and Almeida, 2003; Song, Almeida, and Wu, 2003; Singh and Fleming, 2010; Singh and Agrawal, 2011). Past research has detailed the ways in which scientific knowledge can help organizations overcome technological barriers and identify valuable new technological opportunities (Fleming and Sorenson, 2004; Fabrizio, 2009; Arora et al., 2017). To take advantage of these opportunities, firms need internal capabilities in basic science to understand and utilize the new ideas emerging beyond their boundaries (Dyer and Singh, 1998; Lane and Lubatkin, 1998; Zahra and George, 2002; Lewin, Massini, and Peeters, 2011). One way in which firms can build these capabilities to access valuable external knowledge at the forefront of research is by hiring scientists from academia (Henderson, 1994; Lacetera, Cockburn, and Henderson, 2004). While there has been extensive prior research on why firms build internal scientific capabilities (Cohen and Levinthal, 1990; Cockburn and Henderson, 1998; Arora et al., 2017), we know less about firms' strategies to exploit the human capital of the scientists carrying out basic research in the firm.

Firms can also obtain value from their employees' human capital through collaborations with external scientists, which provides access to valuable knowledge and expertise beyond the boundaries of the firm (Liebeskind et al., 1998; Cockburn and Henderson, 1998; Fabrizio, 2009). To do this successfully, firms must have the internal scientific capabilities to identify and interpret potentially valuable knowledge either to build on it internally or to find productive opportunities for external collaboration (Cohen and Levinthal, 1990). Beyond the firms' internal routines and stocks of knowledge, scientists who are able to access networks beyond the firm can have greater access to information, particularly if this has a tacit dimension (Rappert, Webster, and Charles, 1999; Zucker, Darby, and Armstrong, 2002). Moreover, to subsequently use external ideas, firms need scientists who have the human capital needed to assimilate these ideas and translate them internally for relevant projects inside their organization (Zahra and George, 2002). Numerous prior studies have examined complementarities between internal scientific research and external knowledge sourcing strategies (Powell et al., 1996;

Cassiman and Veugelers, 2006; Fabrizio, 2009). This literature highlights the dual role of firms' scientific human capital strategies. One role is to produce valuable knowledge internally through scientific research (Arora et al., 2017); another is to access and build on external knowledge that may be valuable to the firm (Cohen and Levinthal, 1990; Cockburn and Henderson, 1998).

Reward Systems in Science

While firms may benefit from scientists' participating on research projects of interest to the academic community (Hicks, 1995; Cockburn and Henderson, 1998). Nonetheless, the nature and goals of commercial science differs from academia in some important ways (Dasgupta and David, 1994; Gittelman and Kogut, 2003; Evans, 2010; Sauermann and Stephan, 2013). Firms are ultimately trying to develop valuable products based on their scientific research, whereas academia places more emphasis on the credit-based rewards for furthering scientific understanding. Even where firms and academic scientists are simultaneously working on very similar topics, corporate researchers are more likely to subsequently focus on the range of applications whereas academic scientists publish more subsequent research on the fundamental attributes of the phenomenon (Bikard, 2015).

Scientific credit – as a reward system – is allocated according to the extent to which scientific research improves knowledge of the fundamental properties of a phenomenon (Merton, 1957; Dasgupta and David, 1994). As the knowledge frontier advances, scientists seeking to make these types of contributions are incentivized to focus their research more narrowly to develop highly specialized knowledge sets that are deeper but narrower than before (Jones, 2009). Firms benefit from rapidly identifying relational applications of new knowledge that can patented and form the basis of products (Evans, 2010). Scientific human capital is important in understanding and testing these potential relationships but may require less specialized knowledge than research on more fundamental properties of an object. While domain knowledge will be important for this, and for maintaining effective external collaborations, the scientific skills to rapidly test potential relationships are likely to be applicable across a wider range of topics than the specialized knowledge required for fundamental academic contributions in a given research area. Evans (2010) shows that academia-industry collaboration tends to produce research that is less central to existing hubs of knowledge in a scientific field.

This divergence has implications for the ways in which firms will exploit the knowledge and skills of experienced scientific hires and how industry transitions may affect scientists' research trajectories. First, firms may be more interested in the range of potential applications of knowledge in a domain in which a scientist has expertise, rather than in narrow niches corresponding to highly specific knowledge. Second, insofar as firms value science more as a tool to discover potentially commercializable relationships between objects (rather than furthering fundamental understanding as a research goal) the value of the scientists' skillset may extend more widely than it would in academia. The first of these regards the differential value for a firm of developing new knowledge within a scientific area, relative to scientists in academia. The second suggests that, given this, scientists skills can be more widely applied within the scientific area to generate outputs with commercial value, than they could to generate academic value. Where the goal is furthering fundamental understanding in a topic, the scientist may need to already be at the knowledge frontier in that topic (in addition to having the necessary research skills). However, if much of the value of their scientific human capital is identifying, interpreting, and experimenting with new ideas from outside the firm's boundary, advanced research skills that can be flexibly deployed may be more important than exploiting deep topic knowledge.

MARKETS FOR SCIENTIFIC HUMAN CAPITAL

There is a significant gap between industry and academia salaries, on average, which implies that firms have to pay a compensating differential to attract and motivate scientists to forgo the non-pecuniary benefits of academia (Agarwal and Ohyama, 2013; Sauermann and Stephan, 2013). For example, many scientists have a preference for participating in the wider scientific community through publishing – albeit with heterogeneity between scientists (Stern, 2004; Roach and Sauermann, 2010). Nonetheless, despite differences on average in the type of research pursued by scientists in industry and academia (Sauermann and Stephan, 2013), there is still substantial overlap in research between corporate and academic research in science-intensive industries (Stokes, 1997; Bikard, 2015, 2018). Survey data from the National Science Foundation shows that more than one-quarter of industry scientists in the United States work on basic research and a sizeable minority of academic scientists engage primarily in applied projects (Agarwal and Ohyama, 2013).

Despite this, we know relatively little on how career transitions between the two types of institutions occur during the course of scientists' careers and how transitions impact the type of research scientists pursue. Industry employment may be becoming more attractive to scientists as volumes of public funding in academia have stalled and competition for these resources increased (National Science Foundation, 2017; Vallas and Kleinman, 2008). As these changes occur in scientific labor markets, it becomes all the more important to understand how the knowledge and skills developed by academic scientists can be a valuable human capital resource for firms. Most empirical research regarding how scientists choose industry or academic employment and the impact it has on their research outputs, has focused on surveys of graduating doctoral students (e.g. Stern, 2004; Roach and Sauermann, 2010; Sauermann and Cohen, 2010). This research typically finds higher earnings profiles for industry scientists and that some scientists with a greater 'taste for science' are willing to accept lower initial salaries to work in academia. It also finds that academic scientists have significantly greater publication volumes, on average, but are far less likely to patent than scientists in industry (Sauermann and Stephan, 2013). A second stream of research has analyzed transitions to entrepreneurship from academia, primarily in the context of commercializing the outputs of their academic research (Stuart and Ding, 2006; Roach and Sauermann, 2015; Azoulay, Liu, and Stuart, 2016). This literature identifies an important role for both local social norms and imprinting in early academic careers in scientists' decisions to become entrepreneurs (Stuart and Ding, 2006; Azoulay et al., 2016). It also typically finds that scientists becoming entrepreneurs do not appear to suffer negative effects on their scientific output from engagement with commercial goals (Azoulay, Ding, and Stuart, 2009; Hvide and Jones, 2018).

Beyond these highly specific contexts, less is known about the impact of transitions on scientists' research during the course of their careers. Growing collaborative activities between industry and academic scientists promoting networks across institution types, changing academic norms, public grants becoming more restrictive, and shifts in relative funding availability may induce greater transition rates to industry (Bozeman and Gaughan, 2007; Vallas and Kleinman, 2008; Roach and Sauermann, 2017). Previous research has also focused on how the human capital of 'star' scientists can positively impact a firm (Lacetera et al., 2004; Zucker, Darby, and Torero, 2002; Rothaermel and Hess, 2007). These scientists are recognized by their peers as highly knowledgeable and influential scientists who can span boundaries inside and outside the firm and enhance its ability to access and utilize external knowledge. Star scientists may be particularly valuable in the scientific labor market and be accorded greater autonomy and resources for the research by the firm (Liu and Stuart, 2010; Stephan, 2012). However, they may also be unrepresentative of the mass of scientists who comprise the bulk of firms' scientific human capital resources.

Division of Labor between Industry and Academia

Prior theoretical research has also analyzed how scientists select into academic or industry careers based on ability and preferences for types of research (Aghion et al., 2008; Lacetera, 2009; Agarwal and Ohyama, 2013). Corporate research, driven by maximizing the expected returns of scientific research, is seen to limit scientific autonomy as firms eschew more speculative research and be more likely to curtail projects that are not showing interim results even if scientists wished to continue with it for its scientific potential (Lacetera, 2009). On the other hand, higher autonomy over projects can be beneficial in incentivizing workers' efforts in settings where research projects' ongoing outputs are more speculative and difficult to contract over ex ante (Eisenhardt, 1985; Foss and Laursen, 2005). Building on these insights, the current literature predicts a division of labor across different types of scientific projects. In particular, the institutions of academia are seen to provide appropriate incentives for more speculative and fundamental research due to greater control rights and rewards for generating scientifically, rather than only commercially, valuable knowledge. Conversely, industry can provide more high-powered incentives to scientists working on more incremental applied research projects where scientists' and firms' interests are less likely to diverge and where there is less ambiguity about a project's ongoing value (Aghion et al., 2008).

An implication of this line of research is that the division of labor between academia and industry should partly be explained by these types of institutional characteristics. The optimal interface for collaboration between industry and academia would be determined by the nature of a scientific project according to which type of institution has advantages in providing the appropriate incentives for the type of research needed. In particular, this theoretical literature suggests that firms will focus on identifying potentially valuable ideas and developing these in new projects, while retaining control rights that are exercised according to the revealed commercial prospects. On the other hand, academia will promote more open-ended exploratory research where the perceived scientific value may diverge more strongly from perceivable commercial rewards.

Scientists' Human Capital and Research Direction

The discussion in the preceding sections leads to the core hypothesis of this paper. The absorptive capacity literature suggests much of the value of scientific human capital in firms is as a more flexible resource that can evaluate and experiment with cutting-edge external knowledge (Cohen and Levinthal, 1990; Arora and Gambardella, 1994). It may be costly to rely on a wide range of topic specialists may be needed to get organization-level diversification across all relevant areas of science (Jones, 2009; Agrawal et al., 2016). Firms and their internal scientists can contract or build relationships with academic scientists to access specific ideas as they develop in academic science, rather than hire holders of this knowledge (Liebeskind et al., 1996; Cockburn and Henderson, 1998; Zucker, Darby, and Armstrong, 2002; Fabrizio, 2009). Moreover, since the goals of corporate research differ from academic research in terms of the focus on fundamental scientific advance, exploiting deep topic knowledge will be less important to understanding the relational properties between phenomena of interest in industry (Evans, 2010). Additionally, if firms are institutionally less able to commit to research projects where the expected value is more ambiguous or takes longer to become apparent (Aghion et al., 2008; Lacetera, 2009), the value of exploiting scientists' topic knowledge in more open-ended deep dives into specific topics may be lower than exploiting their knowledge and skills in discrete projects where evaluation decisions on project continuation can be made more rapidly.

The prior literature typically approaches differences in industry and academic research in the context of the distinction between basic and applied science. However, in Pasteur's Quadrant, where there is alignment between basic scientific value and the potential uses of new scientific knowledge, this distinction may not always be as relevant (Stokes, 1997; Bikard, 2015). Scientists in academia and industry may pursue very different research strategies even when working on basic research. Whereas the reward system of academia incentivizes scientists to develop and exploit deep topic expertise, scientists engaged in basic research in industry may be more likely to exploit (and further develop) their knowledge and skills by applying it to a wider range of research topics.

Hypothesis 1: Scientists will apply their human capital more widely to explore new scientific concepts at higher rates when employed in industry.

To provide novel insights into the drivers of firms' scientific human capital strategies, I analyze the consequences of scientists' transitions to industry employment in the stem cell research field. I examine how scientists' research direction changes when acquiring industry employment using a range of metrics that indicate whether they are applying their scientific skills to new areas or are working in areas in which they already had expertise. I use detailed publication-level data to analyze the changes in research direction among U.S. scientists transitioning to industry employment relative to other scientists who remain employed in academia. If transitions to industry are associated with scientists covering a wider range of topics, this suggests that a major reason for their hire is to acquire the scientists' skillsets and more widely applicable research abilities to search more flexibly within their field. On the other hand, if firms view the specialized topic knowledge aspect of academic scientists' human capital as the core resource of value, we should expect see that new hires' research remain focused in their previous domain of expertise. I take academic scientists as a baseline for this because, as noted above, these scientists are incentivized to deepen knowledge in their niche. Hence to the extent that it is scientists' frontier scientific knowledge (and academic networks) in these niches are valuable to firms, we would expect to see these scientists to continue to develop their research in these areas.

The relationship between scientists' employment in academic or industry institutions and the nature of their research also has relevance to the more policy-focused concerns on how the division of labor between academia and industry affects knowledge accumulation (David, Hall, and O'Toole, 2000; Murray and Stern, 2007; Azoulay et al., 2018, Bikard et al., 2019; Arora et al., 2019). Understanding how the logic of inquiry in academia and industry maps to the use of scientists' human capital in their research helps us to understand how policy changes that affect the locus of research activity will affect knowledge accumulation in a scientific field. The extent to which firms value academic scientists' topic knowledge relative to their scientific research skills will have significant implications to how these scientists develop their knowledge. If scientists in industry focus on exploring topics more widely, rather than more deeply, this will affect both how they further their individual human capital and the research outputs they contribute to the wider scientific community. If deep topic expertise is required to make major breakthroughs (Jones, 2009; Agrawal et al, 2016), a greater (lesser) share of science taking place in industry may decrease (increase) the rate at which these occur. Conversely, to the extent individual-level knowledge diversity promotes making valuable new connections between different ideas (Taylor and Greve, 2006; Nagle and Teodoridis, 2020), higher (lower) relative industry employment may increase (decrease), the rate at which these occur.

RESEARCH DESIGN

Institutional Details

My empirical setting for testing relationship between scientists' employment in academic or industry institutions and the nature of their research is the stem cell science field of biomedical research. Stem cells are a specific type of cell that have the unusual and extremely valuable property of being capable of differentiating into many types of specialized cells, such as skin, nerve, or muscle cells. In 1999, *Science* magazine recognized the immense potential of the stem cell field to generate a vast range of new treatments and therapies, by naming stem cell research as its 'Breakthrough of the Year' (Vogel, 1999). Across industry and academia great attention has been paid to the new opportunities for medical treatments opened up by advances in stem cell science. Stem cell science is an archetype of Pasteur's Quadrant where basic research and potential uses of new scientific knowledge are more closely aligned. Therefore, large amounts of research are carried out by scientists in both academia and industry with significant scope for mobility between settings.

In this paper, I examine how moving to industry changed the research focus of U.S. scientists working in stem cell research compared to those working in academia. This helps shed light on how transitions to industry affect scientists' individual research outputs, the relative value of different dimensions of their human capital in academia and industry, and how the division of scientific labor between the two types of employers affects knowledge accumulation. In particular, I am interested in the extent to which scientists are likely to apply their skillset more widely across a range of topics after moving from academia to industry. This finding would suggest that firms place less emphasis on deep topic knowledge in their human capital strategies than is present in academia. Alternatively, scientists' research may remain focused on exploiting their antecedent knowledge set, which would suggest that this is the core reason for firms to employ scientists from academia.

Data

I begin by using the Scopus database of scientific publications to identify all scientists who had a publication in a stem cell field and were based in the United States during the period from 1996 to 2000 inclusive according to the affiliation information on their publications. To identify stem cell publications, articles were searched for based on mentioning mention “stem cell” or variants in their titles, abstracts, or keywords. For each scientist in the sample, the author information, affiliations, abstracts, and citation statistics were extracted for all papers authored until 2008 inclusive. I excluded those scientists whose first publication appeared before 1976 to mitigate the risk that a reduction in publication rates could be driven by selective attrition of older scientists retiring from active research. This procedure yields a total of 5,348 scientists with an identifiable affiliation during the sample period.

I extend this data by scraping all the keywords associated with the articles from the PubMed database, which includes almost all journal articles published in the life sciences. PubMed contains precisely indexed keywords from a managed dictionary to describe the content of archived papers. These MeSH terms represent medical subject headings and comprise a vocabulary managed by subject-specific experts at the National Library of Medicine (NLM). There are currently over 28,000 descriptors in the MeSH vocabulary along with more than 200,000 supplementary terms that typically indicate specific examples of chemicals, disease, and drug protocols. Indexing is independent of article authors. Terms are assigned by professional indexers at the NLM who select them based on a specific protocol. Indexers therefore place an article at a point in the space of scientific concepts based on its content on the dimensions represented by each index word. For each author in the sample, I use this data to calculate the number of MeSH terms used each year, the cumulative count of unique terms they have used in their careers to date, and the number of terms used for the first time in each year. I also perform textual analysis on the abstracts to provide further evidence on how scientists’ research changes when moving from academia to industry.

Dependent Variables

I am interested in how the conceptual content of scientists' research changes when they change from industry to academic employment. My main dependent variable, *Percent New Terms*, is defined as the percentage of MeSH terms that are indexed to articles published by a focal scientist i in a given year t that had not been previously indexed by the NLM to scientist i's articles from t-1 to the year in which i first published. This reflects the percentage of conceptual terms identified by the NLM used by a scientist in a given year that are new to the scientist's (revealed) antecedent knowledge set as defined by previously used conceptual terms. Thus, this variable represents the extent to which scientists use concepts new to them relative to concepts familiar to them. I also carry out robustness checks in which the definition of a new term excludes subsidiary terms applied to a main stem term (e.g., treating "Antigens, CD34/analysis" and "Antigens, CD34/immunology" as a single "Antigens, CD34" concept). The variable *Percent New Stem Terms* provides a measure of the extent to which scientists are covering more diverse conceptual space across a range of topics, rather than diving more deeply into the range of possibilities within a topic area.

As a robustness check, I use an additional dependent variable, *Total New Terms*, which represents the number of MeSH terms that appear on all articles published by a focal scientist i in a given year t that had not been indexed by the NLM to scientist i's articles from t-1 to the year in which i has a first publication. Due to the skewness of the variable, I use the natural logarithm of one plus the raw count of new terms. *Total New Terms* does not account for the conceptual exploration relative to the exploitation of previously covered conceptual space (as does *Percent New Terms*). However, it does provide a valuable check that the core results on new concept use are not biased by corporate publications systematically being associated with a higher percentage of new terms due to having a smaller denominator of total index terms. Instead, this variable focuses solely on the numerator. I perform unreported robustness checks in which I redefine *Percent New Terms*, *Percent New Stem Terms*, and *Total New Terms* from t-1 to t-8 and from t-1 to t-5 and find qualitatively similar results.

To ensure the robustness of my results, I create a second set of dependent variables using textual analysis of the abstracts of scientists' papers. This allows me to focus on changes in the topics covered by scientists' research in the stem cell field. If firms hiring scientists with knowledge of stem cell research area, the extent to which they subsequently change their rate of working in new areas of stem cell research provides an indication of how far firms are exploiting their antecedent knowledge of topics (relative to hiring them to apply their expertise more widely). I use topic modelling to identify 100 topics used across the stem cell research papers in my sample from 1994 to 2010. I use latent Dirichlet allocation as the basis for an algorithm that infers the posterior distribution of the unobserved topics across the documents in my sample (Blei, Ng, and Jordan, 2003; Blei, 2012). The first output of the topic model algorithm is a vector of words weighted by how important they are in that topic. The second output is a vector of topics for each document that are weighted by how important they are to the document. I follow prior papers in setting the model to identify 100 topics. This is perceived to provide semantically meaningful topics that are not over-fitted to the observed papers (Blei and Lafferty, 2007; Kaplan and Vakili, 2015).

The language scientists use to describe their research provides insights into how they think about the content of their research and how they frame it to be received by the wider scientific community. Scientific paper abstracts are typically 150-300 words requiring the scientist to be parsimonious and careful in their use of language to describe the content of their paper. Abstracts are an important summary of research that indicate to interested authors that an article has specific content and may be of interest to them given the focus of their research. In the life sciences, in particular, abstracts are also very important for the role they play as the key search result for scientists who use the PubMed database to identify relevant literature on scientific topics. It contains almost all publications in the life sciences and is an important resource for scientists searching for prior work in an area of research.

Since my sample only comprises a selection of abstracts from the stem cell field, the topic model algorithm does not model the entire field of stem cell research. However, I use approximately 15,000 abstracts covering 33,000 scientist-publication observations from 1994 to 2010 in the topic model (which include similar scientists outside the U.S. who are not included for purposes of analysis but make the corpus more comprehensive). Although this is not the entire stem cell research field, it is sufficiently large to get reasonably reliable definitions of topics if there is no unobserved selection into the sample by scientists based on how they describe their research. Since topics are defined by the co-occurrence of words if these were used systematically differently by non-included authors, these would either be their own topics (in which case no bias would be introduced) or at the margin they might affect the topic model's allocation of terms to topics if the frequency in which they co-occurred over the entire corpus were significantly altered.

I extend the sample by two years before and after the period on which I focus my empirical analysis to ensure that topics in papers early and late in my sample are at less risk of being misclassified because they are early or late instances of topics and these topics may not be identified if their co-use of concepts is less observed. I also use a more time restricted sample to mitigate the risk that there are within-topic changes in language use over time that affect my topic classification. I do not attempt to model the paper-topic relationships outside the stem cell field. This is because the authors' publications represent such a small sample of the overall space of scientific concepts outside stem cell research. In order to deliver the required number of topics, the topic model algorithm would have to generate topics over larger intellectual distances. Therefore, these would be at a fundamentally different level of abstraction than for the stem cell papers. Moreover, the sample of non-stem cell research papers may be a highly unrepresentative sample of the conceptual space in which the scientists work due to the omission of so many other authors in the wider space. Therefore, in this case, the distribution of concept co-occurrence may be biased by these omissions and so topic modelling would generate flawed outputs.

I use the output of my topic model to generate the variable *New Topic*. This is an indicator variable set to one in years in which a scientist authors an article for which the modal topic in a paper is outside the set of modal topics in which they had previously published research, and zero otherwise. Hence, this variable indicates whether a scientist's publication record reveals that their stem cell research has been on a new topic. The online appendix provides a summary of the model output by showing the ten most frequently occurring terms in a topic generated by the topic model algorithm for five randomly selected topics in stem cell research. When carrying out empirical analysis for which *New Topic* is the dependent variable, I exclude scientists in my sample with a first stem cell publication in a year prior to 1994 (when I begin including publications in the topic model). This is to ensure that the results are not biased by scientists' papers being miscategorized as in a new topic when in fact they had published similar research prior to 1994.

To provide further insights into how employment transitions from academia to industry are associated with changes in scientists' research, I analyze three further dependent variables. These are the scientist's number of publications, the citations per publication, and the median CHI journal ranking of the publications in a given year. The purpose of the variables is to ensure that scientist-level changes in the number of publications or moving to more applied research are not driving the main results. For example, scientists may be exploring more conceptual space per publication but having fewer publications or doing more applied research that covers more topics if these are shorter term. Changes in the rate of publishing – in addition to its direction measured by the core dependent variables – is also important to understanding how the division of labor between academia and industry affects the accumulation of scientific knowledge. I also analyze changes in the number of citations per publication to examine whether this research becomes more or less impactful after scientists acquire corporate affiliations. This is important to understand the consequences of these employment transitions on knowledge accumulation in a scientific field, beyond the directional changes in conceptual space covered by scientists' research.

Finally, the median CHI journal ranking of the scientist's publications indicates whether the results are being driven by changes in the basicness of their research. Even within Pasteur's Quadrant, where basic research is inspired by potential usefulness, scientific research can be more fundamental or applied. In biomedical science, research involving clinical trials may be seen as less basic than laboratory research on the properties and potential of particular substances, for example. The CHI journal ranking captures the extent to which papers published in a journal typically involve more applied clinical research or basic science on a scale of 1-4 points. Where a scientist has a publication in a given journal, I take the CHI score for that journal (if it is ranked in the CHI index) and calculate the median score of all publications in that year (if any are observed). This variable provides insights into whether the same scientist performs more applied research when moving to industry. For example, the scientist may cover more conceptual space if they switch to doing a greater volume of clinical research applying a wider range of substances.

Independent Variables

My main independent variable of interest is *Corporate Affiliation.* This is an indicator variable set equal to one for all scientist-year observations in which scientist *i* has a corporate affiliation on an article, and zero otherwise. Where this is not directly observed from a publication by scientist *i* in year *t* that directly provides the required information, I interpolate between observed affiliations. When examining whether the NLM has indexed new MeSH terms to a scientist's publications in a given year, I control for the natural logarithm of one plus the count of unique terms that have previously been indexed to that scientists' papers. This is intended to control for the stock of prior terms and limit the extent to which the results may be driven by a mechanical correlation in which scientists with more prior publications have less potential to cover additional concepts associated with new terms to be indexed to their papers. For similar reasons, I also control for the natural logarithm of one plus the cumulative number of publications. This is intended to reduce the risk that the rate at which scientists cover new intellectual space (associated with new MeSH terms) is driven by the extent of their prior publishing activity rather than changes in their research strategies.

When using topic-based dependent variables, I incorporate a similar set of variables to control for the risk that the number of prior publications may be related to the stock of prior research. If fewer observed research outputs are associated with the rate of publishing in different topics, this may in part be driven by a difference in the baseline probability that an author is observed as exploring a new topic is linked to the number of observations that reveal past engagement with a topic. First, I control for the lagged count of stem cell publications to control for the risk that the number of publications is mechanically related to the rate at which new topics can be explored. Second, I control for the lagged count of topics that a scientist has used for similar reasons. I include these variables in the regression in the form of the natural logarithm of one plus the value of the variable. Again, the results are robust to including quadratic, rather than logged, versions of these control variables. Finally, I carry out additional analyses in which I examine how the number of stem cell publications a scientist has in a given year interacts with their affiliation. This ensures that the main results are not being driven by scientists with more stem cell publications exploring more topics simply because they have more opportunities for a publication to be in a new topic. In all regressions with controls, I include a quadratic control for the 'age' of the scientist. This is calculated as the number of years between year t and the year in which the scientist had her first publication in Scopus.

Econometric Approach

I compare the change in the research direction of U.S. scientists upon acquiring (and retaining) a corporate affiliation to U.S. scientists compared to those in academia. I use author fixed effects to identify the change in research direction between a scientists pre- and post-corporate affiliation research output compared to matched controls. I follow these scientists from 1996-2008 inclusive. I limit the sample to these years to mitigate the risk that the results are affected by industry scientists being more likely to transition from research to managerial tasks as they advance in their careers in years further after they are observed in the initial inclusion period. To analyze the changes in scientists' research direction in my sample, I use the following specification in my core analysis:

$$Y_{it} = \beta_0 + \beta_1 CorporateAffiliation_{it} + \beta' X_{it} + \tau_t + \delta_i + \varepsilon_{it}$$

Where Y_{it} denotes the dependent variable of interest in each regression for scientist i in year t. The coefficient β_1 is the main variable of interest representing the changes in scientists' research direction following their transition from academia to industry and X_{it} is a vector of control variables as described above. Time period effects affecting all scientists are captured by the year fixed effect, τ_t, and time-invariant individual characteristics are controlled for using scientist fixed effects, δ_i. Thus, the analysis can be seen as identifying the within-scientist changes in scientists' research direction that is associated with acquiring a corporate affiliation. Scientists in my sample can have both academic and corporate affiliations

Instrumental Variables

Scientists transitioning to industry may strategically time transitions or systematically differ in unobserved characteristics to other scientists that make them attractive to industry employers. For example, scientists may switch to industry employment immediately after their research has uncovered a particularly valuable result for commercial exploitation. Alternatively, scientists may switch to industry in order to carry out research in new areas if academic science funding makes it more difficult to acquire resources for research in areas scientists have not previously worked. To address these concerns, I use instrumental variables based on local tax and employment rates to generate plausibly exogenous variation in the assignment of scientists to types of institution. Specifically, I use variation in taxes and employment rates across U.S. states over time as instrumental variables that affect the incentives and opportunities for scientists to take industry employment. In this section of the analysis, I draw on the empirical work of Moretti and Wilson (2017) analyzing scientists' labor market decisions and thank the authors for making their data publicly available.

The instrumental variables focus on changes in the demand side of scientific labor markets in a state at a particular point in time. The first instrument is the state-level tax credit for R&D expenditures. This is based on the tax rates in the state in which scientist *i* worked during the previous year (if it the scientist's first publication year, it is imputed as the rates for the first observed state in the previous year). This variable captures the relative cost to firms in that state of investing in R&D based on the tax credits available to them for these expenditures. Wilson (2009) and Rao (2016) show that that firms in the United States invest more in R&D where tax credits lower the cost to the firm of making marginal R&D investments. This increases firms' incentives to invest in scientific research – where this is a part of their R&D efforts – and thus to hire the scientists with the human capital to carry out this research.

The second instrument is based on the labor market opportunities for scientists in the state in which they reside. The Bureau of Labor Statistics' Census of Employment and Wages records the number of workers in each U.S. state across different occupational groups. The data for NAICS category 54171 records the number of workers employed in scientific research and development in each state over time. Individuals working in the life sciences who are employed by universities and hospitals are not included in this category (these employers have their own respective NAICS categories). I begin by finding a state's pre-sample level of employment (in the year 1995). I then adjust this figure for each year by the percentage growth in employment in NAICS category 54171 for all other U.S. states (excluding the focal state) to create a predicted value of employment opportunities in the state. Finally, I take the natural logarithm of one plus the state's predicted scientific R&D employment. I exclude the scientist's state of residence from the calculation to ensure that changes in the instrument's value is not driven by any endogenous changes in their own state that may increase both employment opportunities and affect a scientist's research strategy. I again construct my instrument for the state in which scientist *i* worked during the previous year and adjust similarly for a scientist's first observed year. This variable is intended to provide variation in the number of employment opportunities for research scientists in that state and thus the relative size of the set of opportunities a scientist would have to transition into industry employment.

The key identifying assumptions are that changes in tax credits or predicted labor market opportunities are only associated with changes in the number of opportunities for scientists to work in industry. It requires that the instruments are not correlated with changes in preferences about which topics scientists choose to research within their existing employment type. I cluster standard errors at the state, rather than scientist, level in these models because the variation in the values of the instrumental variables occurs at the state level. Table 1 below provides summary statistics of the variables used in this paper.

[INSERT TABLE 1 ABOUT HERE]

RESULTS

Scientists' Exploration of New Areas of Conceptual Space

I begin by examining how the acquisition of an industry affiliation is associated with changes in scientists' propensity to explore new areas of conceptual space in their research. The analysis in Table 2 shows how the rate at which scientists use new MeSH terms relative to re-using existing concepts after switching from academia to industry. Models 1 and 2 use my preferred dependent variable *Percent New Terms*. They show that the percentage of terms indexed to a scientists' research in a given year by NLM specialists that had not previously been indexed to their research outputs is greater after scientists' transition to industry. The point estimates in Model 2 suggests that scientists' share of new terms in publications increases by approximately 1.4 percentage points following transitions to industry. This would be equivalent to approximately a 2.5 percent increase in the share of indexed terms applied to a scientists' research that represents new scientific content (which has not been revealed in their antecedent knowledge set through past publications). The results are similar when restricting new terms to the stem terms. Scientists' share of new terms increases by approximately 1.1 percentage points following transitions to industry. This represents an increase of approximately 3.3 percent relative to the baseline rate of new stem terms indexed to scientists' publications.

[INSERT TABLE 2 ABOUT HERE]

As a basic robustness check, Table A.1 in the online appendix uses the *Total New Terms* dependent variable to ensure these results are not being driven by scientists' publications simply covering less conceptual space overall after transitioning to industry. The point estimates in Model 1 and 2 imply that scientists increase their use of new terms in publications by about 13 percent per year following transitions to industry. Models 3 and 4 repeat the analysis of Models 1 and 2 in Table 2 limiting the analysis only to MeSH terms used in the previous eight years to ensure that the results are not being driven by re-use of terms that were last used many years prior by scientists in areas they had previously researched. The point estimates and standard errors of these coefficients are highly consistent with those presented in the main analysis (similar, unreported results are available from the author for models with five-year intervals). Overall, the results suggest that scientists' research covers more new scientific content after moving to industry compared to scientists who remain in academia.

Other Changes in Scientists' Research

The MeSH term-based dependent variables that measure a scientists' research direction are based on observed research outputs. Therefore, the results may be affected by the rate at which scientists' research outputs are observed. For example, there may be changes in the rate at which involvement with industry leads scientists to be subject secrecy or publication delays to protect valuable ideas from leaking to competitors (Czarnitzki, Grimpe, and O'Toole, 2015). As a consequence, I could be under-counting the number of new MeSH terms covered in scientists' research when working in industry (since I would only be observing part of their overall research portfolio). Conversely, if secrecy were systematically applied by firms more to research that exploited scientists' antecedent knowledge sets, this could bias the results by leading me to under-count their repeated use of terms relating to existing knowledge in the denominator of the core variables. The models using the *Total New Terms* variable (which is based on the count, rather than share, of MeSH terms new to the scientist) offers some check on this risk by showing that the crude count of new terms is higher in publications is higher when scientists are affiliated with industry.

To examine whether changes in publication habits or wider changes in scientists' research may be driving the main results, I re-run the core analysis with scientists' number of publications, their number of citations received per publication, and the median CHI ranking of the journals in which they publish in a given year as the dependent variables. The results are presented in Table 3. Models 1 and 2 show that there are no clear changes in the number of publications per year after acquiring industry employment. Models 3 and 4 show that scientists moving to industry appear to receive marginally more citations per paper than those outside industry. Finally, Models 5 and 6 show that the median CHI ranking of journals (a measure of the basicness of the typical research paper a journal publishes) appears to be unchanged after the employment change. Overall, results in Table 3 imply that scientists are not publishing less, having lower impact, or performing less basic research after transitioning to industry employment. This suggests that scientists are not switching to projects that require less skill in basic scientific research nor are they researching topics of more marginal value. However, as the results in Table 2 show, they are applying their scientific human capital across a wider range of the conceptual space of science.

[INERT TABLE 3 ABOUT HERE]

There were also changes in U.S. stem cell funding at the federal level during the sample period, which could have affected scientists' research output and behavior. Federal resources for human embryonic stem cell research were significantly limited from 2001 to 2009 due to restrictions imposed by the Bush administration (although funding for other areas of stem cell research increased). This was particularly pronounced in the immediate period from 2001 to 2005. Over time, foundations and state governments increased funding to provide alternative resources for this area of research. Human embryonic stem cell research was a comparatively small sub-field of stem cell research in terms of the overall volume of publications at the time, but nonetheless had a comparatively high profile in the scientific community. To check whether the changes in federal funds available to academic scientists is driving my results, I replicate my analysis including only the observations from the period before President Bush took office in 2001. The results presented in Table A.2 show a similar pattern to the main results. The magnitudes of the coefficients are typically slightly larger than those in the main analysis. This may reflect the federal policy changes leading academic scientists to explore relatively more new conceptual space by providing a major shock the existing balance between scientists' research choices and funding provision. As an additional check that the results are not primarily driven by changes in U.S. science policy during the sample period, I also replicate the results using a sample of scientists based in countries outside the United States in Table A.3. More information on that sample is provided in the online appendix.

Differences in Scientists' Characteristics

A further concern is that the results may be primarily driven by scientists' transitions to industry at a particular career stage, by a particular cohort, or by a subset of scientists with atypical academic ability. However, this does not appear to be the case. Tables A.4, A.5, and A.6 in the online appendix replicate the core results for the dependent variables above splitting the sample at the median value on variables that measure scientists', ability, cohorts, or career stage. First, to separate scientists by ability, I calculate the median number of citations received by scientists to papers published during the pre-sample period from 1990 to 1995 inclusive. I do this separately for groups of scientists according to their first publication year to account for differences in citations received to papers during this period due to differences in career lengths or career stage during this period. I then split the sample according to whether a focal scientist an above or below median number of citations during this decade. I exclude scientists who had a corporate affiliation prior to 1996 to ensure scientists' observation on this variable is not driven by changes in citations or publication patterns due to non-academic employment, rather than academic ability. I also exclude those who have no observed publications prior to 1996 because I cannot make a pre-sample calculation for these individuals. The results presented in Table A.4 are very similar for both sub-samples.

Next, to examine whether the result is driven by different cohorts in my sample, I split the sample according to the year in which scientists had their first publication (whether it was strictly before 1990). The results in Table A.5 show that the pattern of results is similar in both groups of scientists regardless of whether they had a first publication before or after 1990. Finally, I split the sample, according to scientists' number of active years prior to an observation (based on the time elapsed since their first publication and with the cut taking place at 12 years). The results in Table A.6 show a consistent pattern of results across the sub-samples. However, in this smaller sample the coefficient on *Percent New Stem*

Terms for the sub-sample containing observations of scientists with fewer than 12 years career length is not significant (although with an identical point estimate to that in the longer career length sub-sample).

Table A.7 adds an additional control variable for the number of co-authors a scientist has in a given year. This is to check whether changes in co-authoring patterns that coincide with changes in affiliation type may be driving the results. These could still reflect genuine changes in the application of scientists' human capital to different types of research and the scientists themselves working on more diverse topics. But it could also reflect scientists' publications having more diverse conceptual content because co-authors bring this knowledge to collaborations and the co-author performs that aspect of the research. In which, case it would be difficult to draw conclusions about whether a focal scientist is herself exploring these new concepts. However, the results show that the relationship between industry affiliation and conceptual exploration is positive and significant across models.

To provide further nuance on the relationship between scientists' characteristics, employment type, and their propensity to explore new conceptual space, I present results without scientist fixed effects in Table A.8. The results show that scientists who work in industry tend to have lower publication rates when scientist fixed effects are not included in the regression models. Conditional on actively publishing, however, scientists have similarly higher rates of conceptual exploration compared to academic scientists in these models. The lower average publication rates of scientists in this sample when working in industry appears to be driven by individual-level characteristics correlated with selection into industry employment. It does not appear to result directly from scientists' type of employer. Nonetheless, as an institutional setting, industry employment appears to promote scientists to explore more new research areas compared to academia – regardless of the differences in time-invariant individual characteristics that drive selection into industry employment.

A further concern about the results is that firms intentionally choose to hire scientists who are already changing their research direction to explore new conceptual space. Alternatively, scientists who want to move to industry may start to change their research direction to cover topics and ideas that will be more valued in industry. In this case, the transition to industry would be a consequence of changes in scientists' research direction that were already under way. This should lead to a decline in the observed rate at which scientists explore new concepts after obtaining industry employment if they are then working on these new topics. However, the elevated rate of exploration may continue after the transition if it is a signal from scientists to potential industry employers about the flexibility of their knowledge and skills across research topics. Figures 1 to 3 show the yearly effects of acquiring a corporate affiliation in a ten-year window before and after the transition to industry employment for those scientists who transition to industry during the sample period for the *Percent New Terms*, *Percent New Stem Terms*, and *Total New Terms* dependent variables.

[INSERT FIGURES 1 TO 3 ABOUT HERE]

Exploration of New Stem Cell Topics

I now move to analyze the changes in scientists' research direction within the stem cell domain after the acquisition of corporate affiliations using the additional type of dependent variable generated by a topic model of scientists' publication abstracts. The results of this analysis are presented in Table 4 and are consistent with the results from models using bibliographic data to measure scientists' exploration in MeSH-defined conceptual space. Models 1 and 2 shows that, conditional on publishing a stem cell article, a scientist who has moved to industry is more likely to be publishing in a new topic than controls who remain in academia. The point estimates for each model respectively suggest that scientists are approximately 2 percentage points more likely to publish on a new topic in the stem cell field after moving to industry, relative to a sample mean of 12 percent. This is approximately equivalent to scientists being 17 percent more likely to publish in a new stem cell topic in a given year when they are employed in industry rather than academia relative to the baseline rate.

[INSERT TABLE 4 ABOUT HERE]

Models 3 and 4 show that these results are not driven by changes in the relative stem cell/non-stem cell composition of publications between scientists moving to industry and those remaining in academia. The point estimates imply that scientists are approximately 5 percentage points more likely to have a new stem cell topic in each stem cell publication after moving to industry. Unreported results (available from the author) show that industry-affiliation is not significantly related to a scientist's propensity to publish in stem cell research relative to other fields. The coefficient is negative but not close to significant. (The share of a scientist's publications that are in a stem cell field is 0.01 lower with an exact p-value of 0.25.) This suggests that the main results are not driven by scientists having a greater tendency to do research in stem cell fields relative to non-stem cell fields based on having academic or industry employment. Overall, these results suggest that firms are utilizing scientists' more widely applicable expertise and scientific skillset across topics, rather than focusing on exploiting their existing knowledge in a highly specialized niche as in academia.

Returnees from Industry to Academia

To probe the robustness of my results, I analyze the research outputs of scientists who transition to industry employment from academia during my sample period and subsequently return to academia. The results presented in Table 5 show that scientists who spend time working in industry are less likely to explore new scientific conceptual space upon returning to academia than they were during industry employment. The omitted category in the regression models is employment in academia before transitioning to industry. This provides the baseline against which the coefficient on other affiliation types can be interpreted. There is a consistent pattern of results across the models with the different measures of conceptual space exploration. Scientists engage in more conceptual exploration upon acquiring an industry affiliation than during their prior academic employment. However, upon returning to academia, they once again exhibit less conceptual exploration than during the period in which they are employed in industry. Indeed, their level of conceptual exploration is similar to that of their pre-industry academic employment across the models (where pre-industry academic employment is the omitted category in the regression models).

[INSERT TABLE 5 ABOUT HERE]

Scientific Labor Market Instruments

There are reasons to worry that scientists' selection into industry employment may bias the results. For example, academic scientists may be more likely to enter industry employment if they have made a specific, potentially commercially valuable finding in their recent research. This would lead be a downward bias in my results as scientists would select into industry employment in order to exploit already identified opportunities for specific findings in recent research and be less likely to explore new topics. The results would then not reflect the true effect of industry employment on the exploitation of topic knowledge relative to wider scientific skills, but rather be driven by selective entry to exploit existing knowledge. Alternatively, scientists may choose to move to industry because they desire to change their research direction. Academia is typically associated with greater research autonomy than industry (Dasgupta and David, 1994; Stern, 2004; Sauermann and Stephan, 2013). However, industry may offer an advantageous setting to change research direction if highly specialized knowledge (and the research outputs demonstrating this to funders) is more important to access resources for research on particular topics in academia. This would lead to an upward bias in my results.

Either of these effects seems to be intuitively plausible. Therefore, I use instrumental variables analysis to evidence of the same mechanism. I use two instrumental variables based on tax credits for research and development expenditures and exogenous changes in local scientific labor market opportunities as instruments for labor market affiliations. My instrumental variables vary both over time within a location and at a given point in time between different locations. The instruments provide exogenous variation in the propensity of scientists to have an industry affiliation. The results of the instrumental variables regressions are presented in Table 6 for the core MeSH-based dependent variable *Percent New Terms* and the topic model-derived dependent variable *New Topic*.

[INSERT TABLE 6 ABOUT HERE]

There is a positive relationship between industry employment and scientists' research covering more new conceptual space for both types of dependent variable. Notably, the point estimates of coefficients on industry employment variable are positive and larger in magnitude than the baseline results. This implies that the first of the two proposed selection mechanisms is more relevant in the OLS models. The core results would then be downward bias from some scientists strategically entering industry employment to exploit valuable existing knowledge. It contrasts with a mechanism in which scientists selectively enter industry employment in order to change their research direction (which would lead to upward bias in the OLS results). This interpretation is also consistent with the trends in Figures 1-3. These figures show that scientists cover new MeSH-defined space at a lower rate during their earlier years of industry (albeit at a higher rate than when in academia). In subsequent years of industry employment, the level of exploration stabilizes at a higher rate.

DISCUSSION

In this paper, I distinguish between scientists' specialized topic knowledge and their advanced scientific research skills. I argue that these represent two important dimensions of scientists' human capital that can be exploited by firms in corporate scientific research. This delineation between aspects of scientists' human capital is conceptually distinct from those typically drawn in the prior literature. Past research has typically divided scientists' human capital according to the specialization or generalization of their topic knowledge (Agrawal et al., 2016; Teodoridis, 2018; Teodoridis et al., 2019) or by the levels of experience in basic and applied research (Toole and Czarnitzki, 2009; Baba et al., 2009; Subramanian et al., 2013). While some scientists may be seen as generalists insofar as they work across a wide range of topics, they may apply a similar set of research skills across these areas. Conversely, other scientists may be more specialized, in terms of the topics they research, but use a variety of advanced methods in their narrower area of specialization. The results in this paper show that scientists working in the stem cell field who acquire industry employment have more conceptually diverse subsequent research output. This is consistent with industry placing greater value on hired academic scientists' advanced research skills that can be applied more widely in their corporate human capital strategies. The findings contrast with an explanation in which firms hire experienced scientists with the primary intention of exploiting their highly specialized, but narrower topic knowledge for commercial purposes.

I argue that these results emerge due to differences in the logic of scientific research in academia and industry. The reward structure of academia privileges fundamental insights into novel phenomena (Dasgupta and David, 1994). For scientists working in industry, their employer's primary rewards come from the role of science in speedily identifying valuable and commercializable relationships (Evans, 2010; Sauermann and Stephan, 2013). Thus, firms may benefit more from applying scientists' knowledge to uncovering these types of relationships across a greater range of conceptual space, rather than focusing on exploiting the highly specialized knowledge developed through their prior academic research. Moreover, prior research suggests that a large part of the value scientific human capital has as a resource to firms is provide absorptive capacity, which involves identifying, interpreting, recombining, and experimenting with new ideas from outside the firm's boundary (Cohen and Levinthal, 1990; Arora and Gambardella, 1994; Fabrizio, 2009). Thus, much of the value of the scientific human capital resource may lie in its ability to be deployed more flexibly over a greater range of topics to buttress the absorptive activities in the firm.

Whereas the literature often focuses on the distinction between basic and applied research when discussing how different institutional features drive the optimal division of labor between academia and industry (Aghion et al., 2008; Lactera, 2009), this paper suggests that these institutional features give rise to another important distinction in the way scientific human capital is used in each setting. Namely, the relative topic breadth and focus of scientists' research. I show that a scientist's human capital tends to be employed in a narrower, more focused area of conceptual space in academia compared to industry – even though the basicness and impact of the scientist's research appears to remain unchanged. Their research cover a greater range of scientific ideas when employed in industry without necessarily shifting to more applied research. Industry employment leads scientists to use their human capital to explore more widely – with greater exploitation of their wider scientific skills relative to their existing specialized topic knowledge. Whereas, academia incentivizes scientists to have more narrow search strategies to build greater topic expertise.

These results also contribute to our understanding of scientific labor markets, where career transitions from academia to industry have been relatively understudied outside of early career and entrepreneurial settings, except for the specific case of star scientists. My results show how the type of institution in which a scientist works affects how they exploit their prior knowledge and skills and hence how their subsequent research adds to their knowledge stock. This helps us understand the role specialized scientific knowledge and skills appear to play in firms' human capital management. However, it also points to important ways in which the division of labor between industry and academia may affect the type of scientific knowledge that is accumulated in a scientific or technical field. Interestingly, transitions to industry do not seem to lead to negative changes in the quality of scientists' output (as reflected by forward citations). Nor are there changes in the basicness of scientists' research (at least as captured by CHI journal classifications). Nonetheless, shifts in the division of innovative labor to industry in a given area appear to lead to scientists covering greater conceptual space in their research. This comes with the implicit trade-off that they may develop less of the deep topic knowledge associated with academic science that may be necessary to push forward the knowledge frontier and make scientific breakthroughs (Jones, 2009; Kaplan and Vakili, 2015). This presents insights to policymakers to understand the consequences for knowledge accumulation when designing policies for funding research or promoting mobility between academia and industry in Pasteur's Quadrant.

An extensive literature, primarily in the economics tradition, has sought to divide human capital into different components based on its specialization with respect to occupations, employers, and tasks (Becker, 1962; Gibbons and Waldman, 2004; Lazear, 2009; Gathmann and Schönberg, 2010). General human capital comprises the knowledge and skills that workers can apply across a range of occupations. Task-specific human capital comprises knowledge and skills that are specialized to the tasks undertaken in a specific occupation and can be applied to other occupations only where these have a similar task environment. Finally, firm-specific human capital comprises knowledge and skills that are more tightly coupled to a specific employer's work environment. These skills contribute more to workers' productivity at that specific firm than they would at other employers. Academic scientists' human capital is composed of both deep topic knowledge and the advanced research skills needed to undertake experimental projects in their field. Both these dimensions of human capital are specialized and developed in the task environment of scientific research. However, to the extent that research skills are more widely applicable to research tasks across topics, than advanced knowledge of a given topic, it is a more general form of human capital applicable to a greater total number of tasks. This does not mean that these skills are necessarily less specialized; developing many advanced research skills may require major investments of time and effort. Indeed, skills in novel methods may be rarer than advanced topic knowledge in some areas of science as they diffuse to become part of scientists' skillsets over time. Future research could elaborate which types of research skills are more transferable across different topics to understand how this facilitates scientific research breadth.

There are a number of limitations to this study. The content of publications may be an imperfect proxy for scientists' knowledge stocks. Although the vocabulary indexed independently by the NLM should accurately reflect the conceptual space in which a scientist has published, this may miss important parts of the conceptual space in which they also have knowledge but have not yet published. Additionally, since papers have multiple co-authors, any given author may not have developed knowledge of all parts of the conceptual space indexed by the NLM. The topic-based measures complement the MeSH term analyses. They use the conceptual content of a research output as described by its authors to communicate this to readers. These results show similar patterns to models using MeSH terms. However, they also have limitations due to not being generated from the entire corpus of stem cell publications and not covering the many publications scientists in both industry and academia have from research projects outside the stem cell field.

This paper has not examined how firms' scientific human capital management decisions relate to other channels through which they can internalize valuable external knowledge. Prior research has examined how firms can take advantage of complementarities between internal and external knowledge resources (Cassiman and Veugelers, 2006; Fabrizio, 2009). These findings are consistent with firms using external knowledge strategies for accessing highly specialized expertise on specific research projects through collaboration but choosing to use internal human capital as a more general scientific resource. Further exploration of the links between hiring and collaboration strategies – particularly beyond the role played by star scientists – could generate valuable insights here. Different types of firm might also diverge in the value placed on the different dimensions of academic scientists' human capital. Future research might valuably identify how different types of firms vary in their use of scientific labor. Pertinent considerations may include the scope of research activities, the extent of internal engagement with basic and applied research, and the diversity of scientific human capital within the firm. It may also be fruitful to examine whether differences in the extent to which firms apply the human capital of scientists across different areas of research affects their ability to extract value from this knowledge directly, through internal spillovers, and through collaborations with external scientists. Such analyses would help us better understand the micro-level mechanisms through which firms can learn by hiring specialized knowledge workers and how human capital strategies operate as a micro-foundation of absorptive capacity.

REFERENCES

Agarwal R., and A. Ohyama

2013 “Industry or academia, basic or applied? Career choices and earnings trajectories of scientists,” *Management Science*, 59(4): 950–970.

Aghion P., M. Dewatripont, and J.C. Stein

2008 “Academic freedom, private-sector focus, and the process of innovation,” *RAND Journal of Economics*, 39(3): 617–635.

Agrawal A., A. Goldfarb, and F. Teodoridis

2016 “Understanding the changing structure of scientific inquiry,” *American Economic Journal: Applied Economics*, 8(1): 100–128.

Almeida P., and B. Kogut

1999 “Localization of knowledge and mobility of engineers in regional networks,” *Management Science*, 45(7): 905-917.

Arora A., S. Belenzon, and A. Patacconi

2018 “The decline of science in corporate R&D,” *Strategic Management Journal*, 39(1): 3–32.

Arora A., S. Belenzon, and A. Patacconi

2019 “A theory of the US innovation ecosystem: evolution and the social value of diversity,” *Industrial and Corporate Change*, 28(2): 289–307.

Arora A., S. Belenzon, A. Patacconi, and J. Suh

2019 “The changing structure of American innovation: Some cautionary remarks for economic growth,” *NBER Working Paper*, No. 25893.

Arora A., S. Belenzon, and L. Sheer

2017 “Back to basics: Why do firms invest in research?” *NBER Working Paper*, No. 23187.

Arora A., and A. Gambardella

1994 The changing technology of technological change: general and abstract knowledge and the division of innovative labor,” *Research Policy*, 23(5): 523–532.

Arrow K.

1962 Economic Welfare and the Allocation of Resources for Invention,” in *The Rate and Direction of Inventive Activity: Economic and Social Factor*, Princeton, NJ: Princeton University Press

Azoulay P., W. Ding, and T.E. Stuart

2009 “The Effect of Academic Patenting on the Rate, Quality, and Direction of, Public, “Research Output,” *Journal of Industrial Economics*, 57(4): 637–676.

Azoulay P., J.S. Graff Zivin, D. Li, and B.N. Sampat

2018 “Public R&D investments and private-sector patenting: Evidence from NIH funding rules,” *Review of Economic Studies*, 86(1): 117–152.

Azoulay P., C. Liu, and T.E. Stuart

2016 “Social influence given, partially, “deliberate matching: Career imprints in the creation of academic entrepreneurs,” *American Journal of Sociology*, 122(4): 1223–1271.

Baba Y., N. Shichijo, and S.R. Sedita

2009 “How do collaborations with universities affect firms’ innovative performance? The role of “Pasteur scientists” in the advanced materials field,” *Research Policy*, 38(5): 756–764.

Becker G.

1962 “Investment in Human Capital: A Theoretical Analysis,” Journal of Political Economy, 70(5): 9–49.

Bikard M.

2015 “The Division of Innovative Labor between Academia and Industry and Pasteur’s Quadrant,” Working Paper.

Bikard M.

2018 “Made in Academia: The Effect of Institutional Origin on Inventors’ Attention to Science,” *Organization Science*, 29(5): 818–836.

Bikard M., K. Vakili, and F. Teodoridis

2019 “When Collaboration Bridges Institutions: The Impact of Industry Collaboration on Academic Productivity,” *Organization Science*, 30(2): 426–445.

Blei D.M.

2012 “Probabilistic topic models,” *Communications of the ACM*, 55(4): 77–84.

Blei D.M., and J. Lafferty

2007 “A correlated topic model of science,” *Annals of Applied Statistics*, 1(2): 634–642.

Blei D.M., A.Y. Ng, and M. Jordan

2003 “Latent Dirichlet allocation,” *Journal of Machine Learning Research*, 3: 993–1022.

Bozeman B., and M. Gaughan

2007 “Impacts of grants and contracts on academic researchers’ interactions with industry,” *Research Policy*, 36(5): 694–707

Bush V.

1945 *Science the Endless Frontier: A Report to the President on a Program for Postwar Scientific Research*. Washington, DC: U.S. Government Printing Office.

Cassiman B., and R. Veugelers

2006 "In Search of Complementarity in Innovation Strategy: Internal R&D and External Knowledge Acquisition," *Management Science*, 52(1): 68–82.

Cockburn I., and R. Henderson

1998 "Absorptive capacity, coauthoring behavior, and the organization of research in drug discovery," *Journal of Industrial Economics*, 46(2): 157–182.

Cohen W.M., and D.A. Levinthal

1990 "Absorptive Capacity: A New Perspective on Learning and Innovation," *Administrative Science Quarterly*, 35(1): 128–152.

Czarnitzki D., C. Grimpe, and A.A. Toole

2015 "Delay and Secrecy: Does Industry Sponsorship Jeopardize Disclosure of Academic Research?" *Industrial and Corporate Change*, 24(1): 251–279.

D'Este P., and M. Perkmann

2011 "Why do academics engage with industry? The entrepreneurial university and individual motivations," *Journal of Technology Transfer*, 36(3): 316–339.

Dasgupta P., and P.A. David

1994 "Towards a new economics of science," *Research Policy*, 23(5): 487–521.

David P.A., B.H. Hall, A.A. Toole

2000 "Is public R&D a complement or substitute for private R&D? A review of the econometric evidence," *Research Policy* 29(4-5): 497–529.

Dyer J.H., and H. Singh

1998 "The Relational View: Cooperative Strategy and Sources of Interorganizational Competitive Advantage," *Academy of Management Review* 23(4): 660–679.

Eisenhardt K.M.

1985 "Control: Organizational and economic approaches," *Management Science*, 31(2): 134–149.

Ejsing A.K., U. Kaiser, H.C. Kongsted, and K. Laursen

2013 "The role of university scientist mobility for industrial innovation," *IZA Discussion Paper*, No. 7470.

Evans J.A.

2010 "Industry Induces Academic Science to Know Less about More," *American Journal of Sociology*, 116(2): 389–452.

Fabrizio K.

2009 "Absorptive capacity and the search for innovation," *Research Policy*, 38(2): 255–267.

Fleming L., and O. Sorenson

2004 “Science as a map in technological search,” *Strategic Management Journal*, 25(8-9): 909–928.

Fleming L., S. Mingo, and D. Chen 2007

2007 “Collaborative Brokerage, Generative Creativity, and Creative Success,” *Administrative Science Quarterly*, 52(3): 443–475.

Foss N.J., and K. Laursen

2005 “Performance pay, delegation and multitasking under uncertainty and innovativeness: An empirical investigation,” *Journal of Economic Behavior and Organization*, 58(2): 246–276

Gambardella A.

1992 “Competitive Advantages from in-House Scientific Research: The US Pharmaceutical Industry in the 1980s,” *Research Policy*, 21(5): 391–407.

Gathmann C., and U. Schönberg

2010 “How general is human capital? A task-based approach,” *Journal of Labor Economics*, 28(1): 1–49.

Gibbons R., and M. Waldman

2004 “Task-specific human capital,” *American Economic Review*, 21(3): 203–207.

Gittelman M.

2005 “What makes research socially useful? Complementarities between in–house research and firm-university collaboration in biotechnology,” *Revue d'économie industrielle*, 110(1): 57–73.

Gittelman M., and B. Kogut

2003 “Does Good Science Lead to Valuable Knowledge? Biotechnology Firms and the Evolutionary Logic of Citation Patterns,” *Management Science*, 49(4): 366–382.

Henderson R.

1994 “The evolution of integrative capability: Innovation in cardiovascular drug discovery,” *Industrial and Corporate Change*, 3(3): 607–630.

Henderson R., and I. Cockburn

1994 “Measuring competence? Exploring firm effects in pharmaceutical research,” *Strategic Management Journal*, 15(S1): 63–84.

Hicks D.

1995 “Published Papers, Tacit Competencies and Corporate Management of the Public/Private Character of Knowledge,” *Industrial and Corporate Change*, 4(2): 401–424.

Hvide H.K., and B.F. Jones

2018 “University innovation and the professor’s privilege,” *American Economic Review*, 108(7): 1860–1898

Jones B.F.

2009 “The Burden of Knowledge and the ‘Death of the Renaissance Man’: Is Innovation Getting Harder? *Review of Economic Studies*, 76(1): 283–317.

Kaplan S., and K. Vakili

2015 “The double-edged sword of recombination in breakthrough innovation,” *Strategic Management Journal*, 36(10): 1435–1458.

Lacetera N.

2009 “Different missions and commitment power in R&D organizations: Theory and evidence on industry-university alliances,” *Organization Science*, 20(3): 565–582.

Lacetera N., I. Cockburn, and R. Henderson

2004 “Do firms change capabilities by hiring new people? A study of the adoption of science-based drug discovery,” *Advances in Strategic Management*, 21: 133–159.

Lane P., and M. Lubatkin

1998 “Relative absorptive capacity and interorganizational learning,” *Strategic Management Journal*, 19(5): 461–477

Latour B.

1987 “*Science in action: How to follow scientists and engineers through society*. Cambridge, MA: Harvard University Press.

Lazear E.

2009 “Firm-Specific Human Capital: A Skill-Weights Approach,” *Journal of Political Economy*, 117(5): 914–940.

Leahey E.

2007 “Not by productivity alone: How visibility and specialization contribute to academic earnings,” *American Sociological Review*, 72(4): 533–561.

Leahey E., C.M. Beckman, and T.L. Stanko

2017 “Prominent but less productive: The impact of interdisciplinarity on scientists’ research,” *Administrative Science Quarterly*, 62(1): 105–139.

Lewin A.Y., S. Massini, and C. Peeters

2011 “Microfoundations of internal and external absorptive capacity routines,” *Organization Science*, 22(1): 81–98.

Liebeskind J., A.L. Oliver, L. Zucker, and M. Brewer

1996 ‘Social networks, learning, and flexibility: Sourcing scientific knowledge in new biotechnology firms,” *Organization Science*, 7(4): 428–443.

Liu C., and T.E. Stuart

2010 “Boundary spanning in a for-profit research lab: an exploration of the interface between commerce and academe,” *Working Paper*.

Merton R.K.

1957 “Priorities in scientific discovery: A chapter in the sociology of science,” *American Sociological Review*, 22(6): 635–59.

Moretti E., and D. Wilson

2017 “The effect of state taxes on the geographical location of top earners: evidence from star scientists,” *American Economic Review*, 107(7): 1858–1903.

Murray F., and S. Stern

2007 “Do formal intellectual property rights hinder the free flow of scientific knowledge? An empirical test of the anti-commons hypothesis,” *Journal of Economic Behavior and Organization*, 63(4): 648–687.

Nelson R.R.

1959 “The Simple Economics of Basic Scientific Research,” *Journal of Political Economy*, 67(3): 297–306.

Nagle F., and F. Teodoridis

2020 “Jack of all trades and master of knowledge: The role of diversification in new distant knowledge integration,” *Strategic Management Journal*, 41(1): 55–85

National Science Foundation

2017 *Business R&D and Innovation: 2015*. Arlington, VA: National Science Foundation.

Powell W.W., K. Koput, and L. Smith-Doerr

1996 “Interorganizational collaboration and the locus of innovation: networks of learning in biotechnology,” *Administrative Science Quarterly*, 41(1): 116–145.

Rao N.

2016 “Do tax credits stimulate R&D spending? The effect of the R&D tax credit in its first decade,” *Journal of Public Economics*, 140: 1–12.

Rappert B., A. Webster, and D. Charles

1999 “Making sense of diversity and reluctance: Academic-industrial relations and intellectual property,” *Research Policy*, 28(3): 873–890.

Roach M., and H. Sauermann, 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., and H. Sauermann

2015 "Founder or Joiner? The role of preferences and context in shaping different entrepreneurial interests," *Management Science*, 61(9): 2160–2184.

Roach M., and H. Sauermann

2017 "The declining interest in an academic career," *PLoS ONE*, 12(9): e0184130

Rosenberg N.

1990 "Why do firms do basic research (with their own money)?" *Research Policy*, 19(2): 165–174.

Rosenberg N., and R.R. Nelson

1994 "American universities and technical advance in industry," *Research Policy*, 22(3): 323–348.

Rosenkopf L., and P. Almeida

2003 "Overcoming local search through alliances and mobility," *Management Science*, 49(6): 751–766.

Rothaermel F.T., and A.M. Hess

2007 "Building Dynamic Capabilities: Innovation Driven by Individual-, Firm-, and Network-Level Effects," *Organization Science*, 18(6): 898–921.

Sauermann H., and W.M. Cohen

2010 "What makes them tick? Employee motives and firm innovation," *Management Science*, 56(12): 2134–2153.

Sauermann H., and M. Roach

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

Sauermann H., and P. Stephan

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

Senker J.

1995 "Tacit Knowledge and Models of Innovation," *Industrial and Corporate Change*, 4(2): 425–447.

Simeth M., and J. Raffo

2013 "What makes companies pursue an Open Science strategy? *Research Policy*, 42(9): 1531–1543.

Singh J., and A. Agrawal

2011 “Recruiting for ideas: How firms exploit the prior knowledge of new hires,” *Management Science*, 57(1): 129–150.

Singh J., and L. Fleming

2010 “Lone Inventors as Sources of Breakthroughs: Myth or Reality?” Management Science 56(1): 41–56.

Song J., P. Almeida, and G. Wu

2003 “Learning-by-hiring: When is mobility more likely to facilitate interfirm knowledge transfer,” *Management Science*, 49(4): 351–365.

Sorenson O., J. Rivkin, and L. Fleming

2006 “Complexity, Networks and Knowledge Flow,” *Research Policy* 35(7): 994–1017.

Stephan P.

2012 *How Economics Shapes Science*. Cambridge, MA: Harvard University Press.

Stern S.

2004 “Do scientists pay to be scientists? *Management Science*, 50(6): 835–853.

Stokes D.E.

1997 *Pasteur's Quadrant—Basic Science and Technological Innovation*. Washington, DC: Brookings Institution Press.

Stuart T.E., and W. Ding

2006 “When Do Scientists Become Entrepreneurs? The Social Structural Antecedents of Commercial Activity in the Academic Life Sciences,” *American Journal of Sociology*, 112(1): 97–144,”

Subramanian A.M., K. Lim, P.H. Soh

2013 “When birds of a feather don’t flock together: Different scientists and the roles they play in biotech R&D alliances,” *Research Policy*, 42(3): 595–612.

Taylor A., and H. Greve

2006 “Superman or the Fantastic Four? Knowledge Combination and Experience in Innovative Teams,” *Academy of Management Journal*, 49(4): 723–740.

Teodoridis F.

2018 “Understanding team knowledge production: The interrelated roles of technology and expertise,” *Management Science*, 64(8): 3625–3648.

Teodoridis F., M. Bikard, and K. Vakili

2019 “Creativity at the knowledge frontier: The impact of specialization in fast-and slow-paced domains,” *Administrative Science Quarterly*, 64(4): 894–927.

Toole A., and D. Czarnitzki

2009 “Exploring the Relationship Between Scientist Human Capital and Firm Performance: The Case of Biomedical Academic Entrepreneurs in the SBIR Program,” *Management Science* 55(1): 101–114.

Vallas S.P., and D.L. Kleinman

2008 “Contradiction, convergence and the knowledge economy: The confluence of academic and commercial biotechnology,” *Socio-Economic Review*, 6(2): 283–311.

Vogel G.

1999 “Breakthrough of the year: Capturing the promise of youth,” *Science*, 286(5448): 2238–2239.

Wilson. D.

2009 “Beggar thy neighbor? The in-state, out-of-state, and aggregate effects of R&D tax credits,” *Review of Economics and Statistics*, 91(2): 431–436.

Zahra S.A.

1996 “Governance, Ownership, and Corporate Entrepreneurship: The Moderating Impact of Industry Technological Opportunities,” *Academy of Management Journal*, 39(6): 1713–1735.

Zahra S.A., and G. George

2002 “Absorptive capacity: A review, reconceptualization, and extension,” *Academy of Management Review*, 27(2): 132–142.

Zucker L.G., and M.R. Darby

1997 “Present at the biotechnological revolution: transformation of technological identity for a large incumbent pharmaceutical firm *Research Policy*, 26(4-5), 426–429.

Zucker L.G., M.R. Darby, and J. Armstrong

2002 “Commercializing knowledge: university science, knowledge capture, and firm performance in biotechnology,” *Management Science*, 48(1): 138–153.

Zucker L.G., M.R. Darby, and M. Torero

2002 “Labor mobility from academe to commerce,” *Journal of Labor Economics*, 20(3): 629–660.

TABLES & FIGURES

Table 1: Summary statistics

Variable	Obs	Mean	S.D.	Min	Max
Corporate Affiliation	66,782	0.135	0.342	0	1
Percent New Terms	38,277	0.554	0.201	0	1
Percent New Stem Terms	38,277	0.330	0.172	0	1
New Topic	66,782	0.131	0.337	0	1
Total New Terms	38,277	3.291	0.790	0	6.163
Percent New Terms (Prior 8 Years)	38,277	0.574	0.197	0	1
ln (1+Lag Cum Publications)	66,782	2.548	1.157	0	6.084
ln (1+Lag Cum MeSH Terms)	66,227	4.619	1.699	0	7.941
ln (1+Lag Cum MeSH Stem Terms)	66,227	4.213	1.544	0	7.149
ln (1+Lag Cum MeSH Terms) (Prior 8 Years)	66,227	4.384	1.627	0	7.657
ln (1+Lag Cum SC Publications)	66,782	0.901	0.647	0	4.419
ln (1+Lag Cum SC Topics)	66,782	0.776	0.568	0	3.664
Age (Years Since First Publication)	66,782	12.09	7.235	0	32
ln (1+Number of Publications in a Year)	66,782	0.749	0.747	0	4.094
ln (1+Citations per Publication in a Year)	39,918	3.488	1.134	0	9.021
Median CHI Ranking	31,919	3.010	0.798	1	4
N SC Publications in a Year	66,782	0.259	0.688	0	14
First Publication Year	66,782	1990.1	6.518	1976	2000
First SC Publication Year	66,782	1996.9	3.194	1980	2000
State R&D Tax Credit Rate	66,458	0.057	0.052	0	0.200
State Predicted R&D Employment	66,458	9.557	1.302	0	11.46

Notes: The difference in the cumulative term measure and the number of observations on other yearly variables occurs because some papers do not have indexed MeSH terms. I treat these observations as missing. Therefore, until I observe a first paper with indexed MesH terms for scientist *i*, I am unable to generate a cumulative measure of their prior MeSH term use. Results of robustness tests (available from the author) show no meaningful differences to the main results in the paper if scientists with missing terms in any year are excluded from the sample entirely. I also exclude scientists whose first publication is from before 1994 from the sample in (and only in) the regressions using the topic model-based dependent variable for reasons explained in the paper. However, I include their 'New Topic' and 'First SC Publication' values in the summary statistics calculations because their values for other variables are included.

Table 2: Changes in the rate at which scientists use new MeSH terms in research outputs when acquiring corporate affiliations

	Percent New Terms		Percent New Stem Terms	
	(1)	(2)	(3)	(4)
Corporate Affiliation	0.013***	0.014***	0.009**	0.011***
	(0.005)	(0.005)	(0.004)	(0.004)
Controls	No	Yes	No	Yes
Scientist FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	38,277	38,277	38,277	38,277
Scientists	5,323	5,323	5,323	5,323
R-squared	0.076	0.253	0.119	0.338

Notes: All estimates are based on panel OLS regression with year and scientist fixed effects. The control variables are ln(1+Cum MeSH Terms) and ln(1+Cum Pubs) along with age controls and a constant term. In Models 3 and 4, ln(1+Cum MeSH Terms) only counts the stem terms counted in the dependent variable. Standard errors (in parentheses) are clustered at the scientist level. *** p<0.01, ** p<0.05, * p<0.1.

Table 3: Changes in the impact and basicness of scientists' research output when acquiring a corporate affiliation

	Number of Pubs		Citations per Pub		Med Pub CHI	
	(1)	(2)	(3)	(4)	(5)	(6)
Corporate Affiliation	0.015	0.016	0.066*	0.068*	-0.018	-0.015
	(0.017)	(0.017)	(0.035)	(0.035)	(0.020)	(0.020)
Controls	No	Yes	No	Yes	No	Yes
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	66,782	66,782	39,918	39,918	31,919	31,919
Scientists	5,348	5,348	5,348	5,348	5,152	5,152
R-squared	0.041	0.041	0.086	0.086	0.005	0.005

Notes: All estimates are based on panel OLS regression with year and scientist fixed effects. The number of publications is measured for each scientist-year observation as ln(1+N Pubs) and the number of citations per publication is measured as ln(1+(N Cites/N Pubs)) for each yearly observation. The median CHI of scientists' publications is not observed for all journals in the dataset. If a scientist-year observation lacks a CHI value, despite having a publication, the variable is treated as missing (if it is observed for some publications, but not all, it is included in the regressions based on the value for scientists' CHI-covered journal publications in that year). The control variables are ln(1+Cum Pubs), age controls and a constant. Standard errors (in parentheses) are clustered at the scientist-level to allow for serial correlation and are robust to arbitrary heteroskadasticity. *** p<0.01, ** p<0.05, * p<0.1

Table 4: Changes in scientists' research into new topics in the stem cell field when acquiring a corporate affiliation

	New Topic			
	(1)	(2)	(3)	(4)
Corporate Affiliation	0.021**	0.017**	-0.007	-0.006
	(0.008)	(0.008)	(0.006)	(0.006)
N SC Pubs			0.407***	0.394***
			(0.013)	(0.012)
Corporate Affiliation*			0.062***	0.047***
N SC Pubs			(0.021)	(0.019)
Controls	No	Yes	No	Yes
Scientist FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	59,723	59,723	59,723	59,723
Scientists	4,805	4,805	4,805	4,805
R-squared	0.081	0.172	0.516	0.572

Notes: All estimates are based on panel OLS regression with year and scientist fixed effects. The control variables are ln(1+Cum Topics) and ln(1+Cum SC Pubs) along with age controls and a constant term. Scientists with first publications in stem cell research prior to 1994 are excluded from the sample for the reasons explained in the main text. Standard errors (in parentheses) are clustered at the scientist level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 5: Changes in scientists' research after returning to academia from industry (relative to prior academic employment)

	Percent New Terms	Percent New Stem Terms	Total New Terms	New Topic
	(1)	(2)	(3)	(4)
Corporate Affiliation	0.028***	0.015**	0.183***	0.050*
	(0.010)	(0.008)	(0.050)	(0.028)
Post-Corporate Affiliation	0.003	-0.005	-0.084	0.010
	(0.011)	(0.008)	(0.066)	(0.030)
Controls	Yes	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	3,191	3,191	3,191	3,355
Scientists	308	308	308	291
R-squared	0.238	0.324	0.038	0.098

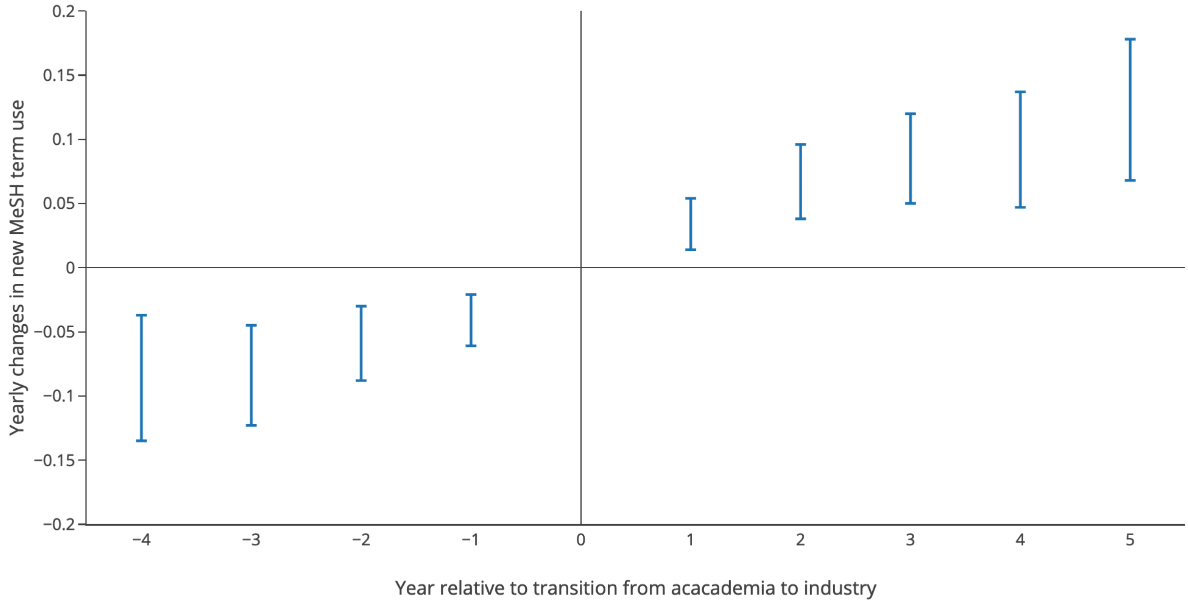
Notes: All estimates are based on panel OLS regression with year and scientist fixed effects. Scientists must return to academia for more than two years and have at least two years of academic employment prior to transitioning to industry to be included in the analysis. The period before a scientist acquires a corporate affiliation provides the baseline against which to interpret the coefficients in the models. All models include age controls and a constant term. In Models 1 and 3, ln(1+Cum MeSH Terms) and ln(1+Cum Pubs) are the additional controls. Model 2 incorporates the ln(1+Cum MeSH Stem Terms) variable as a control in place of ln(1+Cum MeSH Terms). Model 4 includes the additional control variables ln(1+Cum Topics) and ln(1+Cum SC Pubs) in place of ln(1+Cum MeSH Terms) and ln(1+Cum Pubs). Scientists with first publications in stem cell research prior to 1994 are excluded from the analysis in Model 4 for the reasons explained in the main text. Standard errors (in parentheses) are clustered at the scientist level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 6: Changes in scientists' research when acquiring a corporate affiliation (instrumental variables results)

	Corporate Affiliation	Percent New Terms	Corporate Affiliation	New Topic
	(1)	(2)	(3)	(4)
Corporate Affiliation		0.203**		0.262**
		(0.097)		(0.127)
R&D Tax Credit	0.208***		0.265***	
	(0.061)		(0.051)	
Predicted R&D Employment	0.011***		0.010***	
	(0.003)		(0.003)	
Controls	Yes	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	37,761	37,761	59,436	59,436
Scientists	4,938	4,938	4,791	4,791
Cragg-Donald F-stat		47.23		108.95
Hansen J-stat p-value		0.992		0.313
R-squared		0.203		0.157

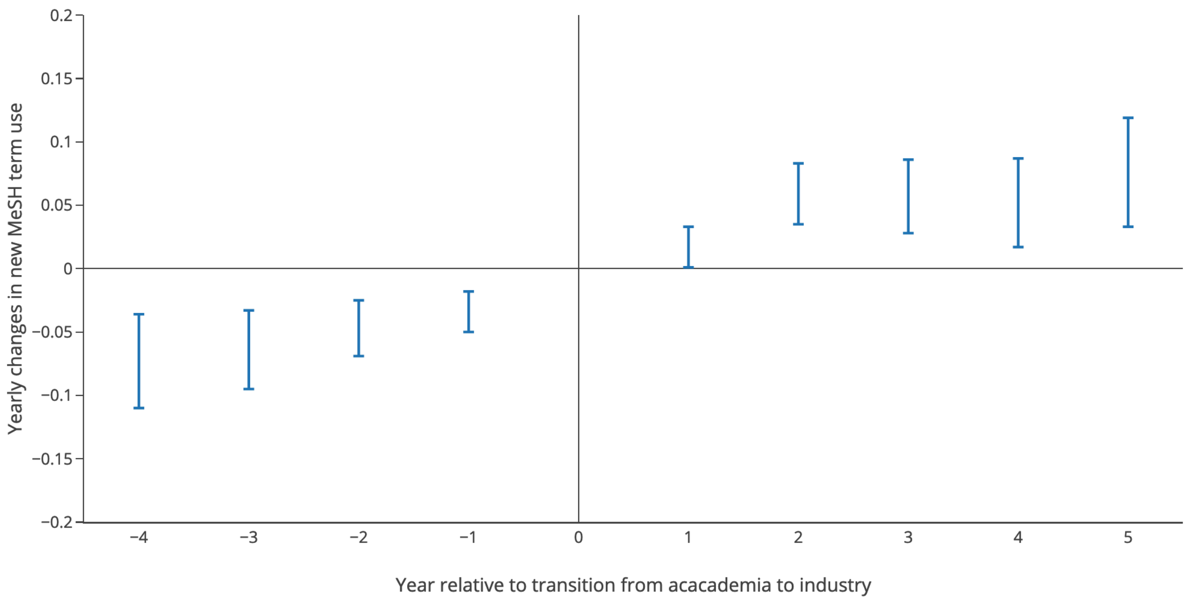
Notes: All estimates are based on panel OLS regression with year and scientist fixed effects. Models 1 and 2 include th ln(1+Cum MeSH Terms) and ln(1+Cum Pubs) control variables. Models 3 and 4 include ln(1+Cum Topics) and ln(1+Cum SC Pubs). Both sets of models include age controls and a quadratic control for the number of years since the scientist was last observed in a new state. This is to mitigate the risk that the results are linked to transitions between states. Observations where scientists leave the United States are dropped from the analysis (as the instruments are U.S.-specific) which leaves fewer observations than in the core analysis. Scientists with first publications in stem cell research prior to 1994 are excluded from the analysis in Models 3 and 4 for the reasons explained in the main text. Standard errors (in parentheses) are clustered at the state level since this is the unit of variation in the instrumental variables. Models with standard errors clustered at the scientist level (available from the author) have smaller standard errors. *** p<0.01, ** p<0.05, * p<0.1.

Figure 1: Changes in the share of MeSH terms in a paper that are new to the scientist following acquisition of corporate affiliation



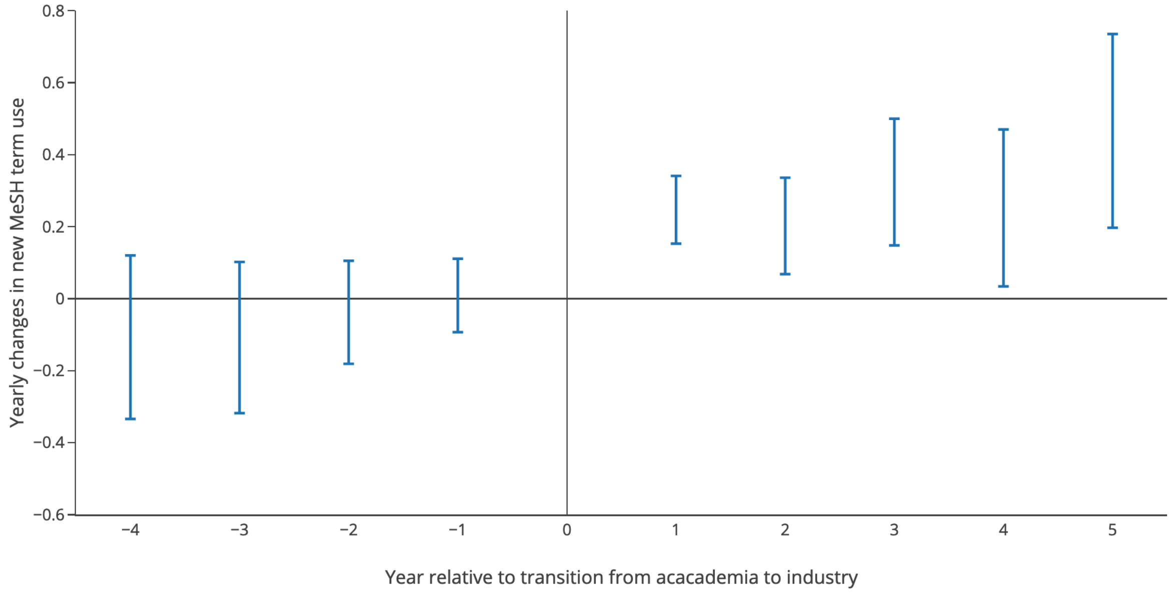
Notes: Results of regression with indicators for each year shown in the graph. The regression sample is limited to observations for the years shown relative to the transition event. Scientists drop from the sample if they return to having exclusively academic affiliations.

Figure 2: Changes in the share of new stem MeSH terms in a paper that are new to the scientist following acquisition of corporate affiliation



Notes: Results of regression with indicators for each year shown in the graph. The regression sample is limited to observations for the years shown relative to the transition event. Scientists drop from the sample if they return to having exclusively academic affiliations.

Figure 3: Changes in the number of MeSH terms in a paper that are new to the scientist following acquisition of corporate affiliation



Notes: Results of regression with indicators for each year shown in the graph. The regression sample is limited to observations for the years shown relative to the transition event. Scientists drop from the sample if they return to having exclusively academic affiliations.

ONLINE APPENDIX

Examples of Terms Defining Topics in the Model Output
Examples of the ten most frequently occurring terms in a topic in descending order of word prominence for five randomly selected topics in stem cell research generated by the topic model algorithm.

cardiac	dna	growth	cell	intestin
heart	damag	factor	pancreat	crypt
function	chromosom	tgfbeta	stem	colon
imag	repair	fgf	islet	cell
cardiomyocyt	recombin	fibroblast	insulin	stem
infarct	genom	prolifer	diabet	epitheli
myocardi	induc	receptor	hormon	epithelium
inject	frequenc	transform	develop	gastric
label	radiat	egf	pancrea	gut
myocyt	exposur	stimul	endocrin	gastrointestin

Table A.1: Changes in the rate at which scientists use new MeSH terms in research outputs after acquiring corporate affiliations (alternative dependent variables)

	Total New Terms		Percent New Terms (prior 8 years)	
	(1)	(2)	(3)	(4)
Corporate Affiliation	0.133***	0.125***	0.012**	0.012***
	(0.022)	(0.023)	(0.005)	(0.005)
Controls	No	Yes	No	Yes
Scientist FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	38,277	38,277	38,277	38,277
Scientists	5,323	5,323	5,323	5,323
R-squared	0.012	0.019	0.043	0.262

Notes: All estimates are based on panel OLS regression with year and scientist fixed effects. The control variables are ln(1+Cum MeSH Terms) and ln(1+Cum Pubs) along with age controls and a constant term. In Models 3 and 4, ln(1+Cum MeSH Terms) only counts the terms used within the previous eight years. Standard errors (in parentheses) are clustered at the scientist-level to allow for serial correlation and are robust to arbitrary heteroskadasticity. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A.2: Changes in the rate at which scientists use new MeSH terms after acquiring corporate affiliations (pre-2001 only)

	Percent New Terms	Percent New Stem Terms	Total New Terms
	(1)	(2)	(3)
Corporate Affiliation	0.015**	0.015**	0.192***
	(0.007)	(0.006)	(0.038)
Controls	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	16,877	16,877	16,877
Scientists	5,267	5,267	5,267
R-squared	0.265	0.335	0.033

Notes: All estimates are based on panel OLS regression with year and scientist fixed effects. Only observations during the period 1996-2000 (inclusive) are included in the sample. All models include age controls and a constant term. In Models 1 and 3, ln(1+Cum MeSH Terms) and ln(1+Cum Pubs) are the additional controls. Model 2 incorporates the ln(1+Cum MeSH Stem Terms) variable as a control in place of ln(1+Cum MeSH Terms). Standard errors (in parentheses) are clustered at the scientist level. *** p<0.01, ** p<0.05, * p<0.1.

Table A.3: Changes in the rate at which scientists use new MeSH terms after acquiring corporate affiliations (only scientists based outside the U.S.)

	Percent New Terms	Percent New Stem Terms	Total New Terms
	(1)	(2)	(3)
Corporate Affiliation	0.029***	0.021***	0.160***
	(0.008)	(0.007)	(0.037)
Controls	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	18,794	18,794	18,794
Scientists	2,527	2,527	2,527
R-squared	0.253	0.320	0.021

Notes: All estimates are based on panel OLS regression with year and scientist fixed effects. The sample is drawn from scientists who had a stem cell publication in the 1996-2000 period and were based exclusively in a non-U.S. country with permissive public funding policies during the entire sample period. All models include age controls and a constant term. In Models 1 and 3, ln(1+Cum MeSH Terms) and ln(1+Cum Pubs) are the additional controls. Model 2 incorporates the ln(1+Cum MeSH Stem Terms) variable as a control in place of ln(1+Cum MeSH Terms). Standard errors (in parentheses) are clustered at the scientist level. *** p<0.01, ** p<0.05, * p<0.1.

Table A.4: Changes in the rate at which scientists use new MeSH terms in research outputs after acquiring corporate affiliations (sample split by pre-sample citations relative to cohort median)

	% New Terms		% New Stem Terms		Total New Terms	
	Low Cites	*High Cites*	*Low Cites*	*High Cites*	*Low Cites*	*High Cites*
	(1)	(2)	(3)	(4)	(5)	(6)
Corporate Affiliation	0.023***	0.022**	0.022***	0.018***	0.086**	0.080*
	(0.009)	(0.009)	(0.008)	(0.006)	(0.044)	(0.043)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11,639	13,662	11,639	13,662	11,639	13,662
Scientists	1,774	1,764	1,774	1,764	1,774	1,764
R-squared	0.198	0.114	0.252	0.166	0.022	0.018

Notes: All estimates are based on panel OLS regression with year and scientist fixed effects. The sample is split according to whether a scientist had strictly more citations in the 1986-1995 period than the median scientist with the same year of first publication recorded in Scopus. Scientists with a corporate affiliation before 1996 are excluded as are scientist for whom the pre-sample citation count is undefined due to only appearing in Scopus from 1996 onwards. The control variables in Models 1, 2, 5, and 6 are ln(1+Cum MeSH Terms) and ln(1+Cum Pubs) along with age controls and a constant term. In Models 3 and 4, ln(1+Cum MeSH Terms) only counts the stem terms counted in the dependent variable. Standard errors (in parentheses) are clustered at the scientist level. *** p<0.01, ** p<0.05, * p<0.1.

Table A.5: Changes in the rate at which scientists use new MeSH terms in research outputs after acquiring corporate affiliations (sample split by first publication year)

	% New Terms		% New Stem Terms		Total New Terms	
	1st Pub Pre-90	*1st Pub Post-90*	*1st Pub Pre-90*	*1st Pub Post-90*	*1st Pub Pre-90*	*1st Pub Post-90*
	(1)	(2)	(3)	(4)	(5)	(6)
Corporate Affiliation	0.012*	0.015**	0.011**	0.011*	0.154***	0.091***
	(0.006)	(0.007)	(0.005)	(0.006)	(0.032)	(0.032)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	19,785	18,492	19,785	18,492	19,785	18,492
Scientists	2,163	3,160	2,163	3,160	2,163	3,160
R-squared	0.113	0.370	0.155	0.449	0.021	0.020

Notes: All estimates are based on panel OLS regression with year and scientist fixed effects. The sample is split according to whether a scientist has a first publication recorded in Scopus strictly before 1990. The control variables in Models 1, 2, 5, and 6 are ln(1+Cum MeSH Terms) and ln(1+Cum Pubs) along with age controls and a constant term. In Models 3 and 4, ln(1+Cum MeSH Terms) only counts the stem terms counted in the dependent variable. Standard errors (in parentheses) are clustered at the scientist level. *** p<0.01, ** p<0.05, * p<0.1.

Table A.6: Changes in the rate at which scientists use new MeSH terms in research outputs after acquiring corporate affiliations (sample split by scientists' career age at observation)

	% New Terms		% New Stem Terms		Total New Terms	
	≤12 years	*>12 years*	*≤12 years*	*>12 years*	*≤12 years*	*>12 years*
	(1)	(2)	(3)	(4)	(5)	(6)
Corporate Affiliation	0.012*	0.014**	0.008	0.008*	0.131***	0.174***
	(0.007)	(0.007)	(0.006)	(0.005)	(0.031)	(0.035)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	19,122	19,155	19,122	19,155	19,122	19,155
Scientists	4,276	3,058	4,276	3,058	4,276	3,058
R-squared	0.377	0.082	0.451	0.110	0.024	0.029

Notes: All estimates are based on panel OLS regression with year and scientist fixed effects. The sample is split according to whether a scientist has a first publication recorded in Scopus strictly more than 12 years prior to the focal observation. This means that scientists can appear in both sub-samples at different stages of their career. The control variables in Models 1, 2, 5, and 6 are ln(1+Cum MeSH Terms) and ln(1+Cum Pubs) along with age controls and a constant term. In Models 3 and 4, ln(1+Cum MeSH Terms) only counts the stem terms counted in the dependent variable. Standard errors (in parentheses) are clustered at the scientist level. *** p<0.01, ** p<0.05, * p<0.1.

Table A.7: Changes in the rate at which scientists use new MeSH terms in research outputs after acquiring corporate affiliations (controlling for number of coauthors)

Panel A: Controlling for number of observed coauthors

	Percent New Terms	Percent New Stem Terms	Total New Terms
	(1)	(2)	(3)
Corporate Affiliation	0.014***	0.011***	0.040*
	(0.005)	(0.004)	(0.021)
N Coauthors	0.000**	0.000	0.016***
	(0.000)	(0.000)	(0.001)
Controls	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	38,268	38,268	38,268
Scientists	5,323	5,323	5,323
R-squared	0.253	0.339	0.256

Panel B: Interacting coauthor control variable with corporate affiliation variable

	Percent New Terms	Percent New Stem Terms	Total New Terms
	(1)	(2)	(3)
Corporate Affiliation	0.019***	0.016***	0.104**
	(0.006)	(0.004)	(0.048)
N Coauthors	0.000***	0.000**	0.016***
	(0.000)	(0.000)	(0.001)
Corporate Affiliation *	-0.000**	-0.000***	-0.002
N Coauthors	(0.000)	(0.000)	(0.002)
Controls	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	38,268	38,268	38,268
Scientists	5,323	5,323	5,323
R-squared	0.253	0.339	0.256

Notes: All estimates are based on panel OLS regression with year and scientist fixed effects. The models in Panel A include a control variable for the scientists' number of observed coauthors in a year. The models in Panel B include this control variable and additionally interact it with the 'Corporate Affiliation' variable. All models include age controls and a constant term. In Models 1 and 3, ln(1+Cum MeSH Terms) and ln(1+Cum Pubs) are the additional controls. Model 2 incorporates the ln(1+Cum MeSH Stem Terms) variable as a control in place of ln(1+Cum MeSH Terms). Standard errors (in parentheses) are clustered at the scientist level. *** p<0.01, ** p<0.05, * p<0.1.

Table A.8: Differences the rate at which scientists use new MeSH terms in research outputs according to whether they have a corporate or academic affiliation (no scientist fixed effects)

Panel A: No additional fixed effects

	Number of Publications	Percent New Terms	Percent New Stem Terms	Total New Terms
	(1)	(2)	(3)	(4)
Corporate Affiliation	-0.149***	0.026***	0.015***	0.058***
	(0.011)	(0.003)	(0.003)	(0.017)
Controls	Yes	Yes	Yes	Yes
Scientist FE	No	No	No	No
Cohort FE	No	No	No	No
Age FE	No	No	No	No
Year FE	Yes	Yes	Yes	Yes
Observations	66,782	38,277	38,277	38,277
Scientists	5,348	5,323	5,323	5,323
R-squared	0.402	0.480	0.536	0.061

Panel B: Incorporating age and cohort fixed effects

	Number of Publications	Percent New Terms	Percent New Stem Terms	Total New Terms
	(1)	(2)	(3)	(4)
Corporate Affiliation	-0.148***	0.026***	0.015***	0.049***
	(0.011)	(0.003)	(0.003)	(0.016)
Controls	Yes	Yes	Yes	Yes
Scientist FE	No	No	No	No
Cohort FE	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	66,782	38,277	38,277	38,277
Scientists	5,348	5,323	5,323	5,323
R-squared	0.423	0.480	0.537	0.077

Notes: All estimates are based on panel OLS regression with year fixed effects. Models in Panel A exclude scientist fixed effects from the regression models. Panel B incorporates fixed effects for scientists' age (measured as the number of years since their year of their first publications) and cohort fixed effects (measured by year of scientists' first observed publications). The number of publications is measured for each scientist-year observation as ln(1+N Pubs). Models include age controls and a constant term. In Model 1, ln(1+Cum Pubs) is the additional control. In Models 2 and 4, ln(1+Cum MeSH Terms) and ln(1+Cum Pubs) are the additional controls. Model 3 incorporates the ln(1+Cum MeSH Stem Terms) variable as a control in place of ln(1+Cum MeSH Terms). Standard errors (in parentheses) are clustered at the scientist level. *** p<0.01, ** p<0.05, * p<0.1.