Hiring for Knowledge or Skills: How Does Corporate Research Exploit the Human Capital of Scientists Hired from Academia?

Michael Blomfield
University of Massachusetts, Amherst
mblomfield@umass.edu

April 1, 2023

Abstract

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 has typically focused 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 components of scientific human capital: topic knowledge and scientific research skills. We build on prior literature on scientific careers and the logic of scientific inquiry to examine how each component is exploited and further developed in corporate research using longitudinal data on scientists in the regenerative medicine field. Scientists moving to industry have more conceptually diverse subsequent research output, but there is little evidence of other changes in their research productivity or the basic-applied focus of their research. We also find evidence that industry employers select scientists with more diverse experience using a wider range of tools, techniques, and interventional drugs in their research. Our findings are consistent with industry placing additional value on the ability to flexibly apply research skills of scientists hired from academia within their wider areas of expertise. The findings also provide insights into how the division of scientific labor between academia and industry may shape the process of knowledge accumulation in "Pasteur's Quadrant," where there is substantial overlap in research between academia and industry.

1. Introduction

An extensive body of research has documented how firms can benefit from building internal scientific capabilities to improve their capacity to absorb and develop new ideas (Cohen and Levinthal, 1989, 1990; Rosenberg, 1990; Gambardella, 1992; Cockburn & Henderson, 1998). Scientific knowledge can facilitate corporate innovation by guiding R&D processes and identifying valuable new information (Fleming & Sorenson, 2004; Arora, Belenzon, & Sheer, 2017). Employing scientists can also help firms develop relationships with external knowledge holders to broker ideas from beyond the organizations' boundaries (Sorenson, Rivkin, & Fleming, 2006; Fleming, Mingo, & Chen 2007; Fabrizio, 2009). Scientists' human capital forms a micro-level foundation of these absorptive and relational capabilities. One way in which firms can acquire this resource is by hiring scientists with extensive specialized training and research experience in academia (Zucker & Darby, 1997; Zucker, Darby & Torero, 2002; Gittelman, 2005; Ejsing et al., 2013).

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 & 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 in their field (Senker, 1995). These skills enable them to understand the properties of, and relationships between, objects of investigation in scientific experiments. A scientist with narrow topic expertise may use a wide range of different research skills to analyze a narrow set of phenomena in their area of interest. Conversely, another scientist may use a narrow set of skills but apply these to a wide range of topics. This leads to significant differences between scientists in the range of research projects to which they can apply their human capital. One has more limited topic knowledge but a more flexible skillset. The other may have knowledge of more topics but lack the ability to carry out research using a wide range of techniques in these areas. Scientists' research skills facilitate the application of other aspects of their human capital to research projects.

The goal of this paper is to examine how this distinction between topic knowledge and research skills can shed light on the ways in which scientists' human capital is used and further developed in academic and industry employment. We analyze changes in the research of scientists transitioning between academic and industry employment in the stem cell field of regenerative medicine, where academic and industry actors share interests in similar topics. We find that scientists moving to industry subsequently explore new scientific concepts at a higher rate and that they have greater pre-transition experience using a wider range of tools and techniques in their research. Our findings are consistent with industry employers placing greater importance on hiring for, and exploiting, scientists' research skills that can be applied across a more diverse portfolio of research projects.

The proposed distinction between topic knowledge and research skills can be illustrated in the social sciences. Consider an economist with advanced research skills in microeconometric analysis. She could specialize in applying advanced microeconometric research skills within one subfield 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-ranging topic knowledge but a more specialized research skill toolkit. She may focus within one subfield of economics, as a topic specialist, and apply many different analytical skills to her research. Alternatively, she may have wider breadth in both her topic knowledge and research skills. Finally, she may focus her research on microeconometric theory or the development of new microeconometric techniques. While topic knowledge and research skills are both specialized forms of human capital developed by academic researchers, there can be important differences in the range of projects to which they can be applied.

Prior research has primarily focused on the first dimension of specialization of scientific human capital—topic knowledge—when examining how scientific human capital contributes to innovation (Jones, 2009; Agrawal, Goldfarb & Teodoridis, 2016; Teodoridis, Vakili & Bikard, 2019). This wider literature is driven by the question of how firms can internalize and exploit advanced topic knowledge by hiring a scientist from academia or collaborating with external experts. Firms' absorptive capacity reflects their ability to identify, evaluate, and reuse external knowledge in ways that are valuable for the firm (Cohen & Levinthal, 1989, 1990; Arora & Gambardella, 1994). Applying external scientific knowledge to new internal uses is an inherently experimental process, using internal scientists' knowledge and skills to decompose and recombine a range of ideas (Henderson, 1994; Senker, 1995; Fabrizio, 2009). Organizations need to possess the internal scientific human capital that facilitates experimentation with a range of ideas and represents the micro-foundation of absorptive capacity at the organization level (Zahra & George, 2002; Lewin, Massin & Peeters, 2011). Firms can achieve organization-level absorptive capacity across different areas of science in multiple ways. First, firms may employ individual scientists as internal experts in a specialized niche. In this case, individual specialization across different scientists provides the necessary range of knowledge and skills at the organization level. Alternatively, firms may apply individual scientists' human capital more flexibly across a wider range of research projects of interest to the firm. Here, the individual scientists' capacity for research breadth facilitates organization-level capabilities across a range of scientific areas of interest to the firm. In this case, the range of projects to which scientists' human capital can be applied is an important part of their human capital strategies (both in hiring and matching employees with tasks).

An additional literature has sought to explain why academia and industry provide appropriate institutional settings for different types of research (Bush, 1945; Nelson, 1959; Arrow, 1962; Aghion, Dewatripont & Stein, 2008; Lacetera, 2009). Academia is seen to offer a research environment suited for basic research because its reward structure lowers scientists' exposure to immediate commercial pressure and emphasizes contributions to fundamental scientific understanding. Conversely, industry is viewed as supporting applied research because firms' and scientists' incentives can be better aligned where there is less ex ante uncertainty about projects' value and less risk that scientists pursue private scientific interests, rather than commercial value, in their research choices. Nonetheless, extensive past research shows that academic and industry scientists have significant overlap in research in many fields (Rosenberg & Nelson, 1994; Sauermann & Stephan, 2013; Bikard, 2018). In "Pasteur's Quadrant," scientific discoveries may contribute both to fundamental scientific understanding of phenomena and have promise for new commercial products (Stokes, 1997; Bikard, 2015). Yet, if 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 in each type of employment.

This paper provides insights into firms' human capital strategies and scientists' knowledge accumulation in different types of employment by analyzing how the research direction of scientists moving from academia to industry changes once they acquire an industry affiliation.

To the extent that firms place particular emphasis on the flexibility of scientists' human capital, we would expect to see their human capital applied across a more diverse range of projects after transitions to industry employment. We would also expect to see that scientists hired by industry have used a more diverse range of research skills in their past research projects—such that they can be seen by a potential employer to possess a more widely applicable skillset for of projects in industry research.

Analyzing the stem cell field of regenerative medicine, we find that when scientists move to industry their research explores new scientific concepts at a higher rate relative to academia. Firms appear to place greater emphasis on exploiting scientists' human capital more flexibly across projects. Conversely, scientists in academic employment continue to focus more narrowly within their existing areas of expertise. Notably, scientists who subsequently return to academic employment from industry exhibit reduced rates of conceptual exploration (similar to that prior to their transition to industry). This is consistent with our results being driven by the scientists' type of employer, rather than being the result of employment transitions more generally. To bolster confidence that this is a causal relationship, we use instrumental variables based on local economic conditions to generate exogenous variation in scientists' employment affiliations.

In line with this interpretation, we find evidence that scientists who transition to industry employment used a wider range of research tools and techniques in their pre-transition academic career. However, they appear similar to other (non-transitioning) scientists in the range of their expertise on specific diseases and anatomical topics. This is consistent with industry employers selecting scientists who have a wider range of research skills to apply across potential projects in their wider areas of expertise—providing a more flexible resource for industry research projects. Our results also shed light on how the division of scientific labor between academia and industry in Pasteur's Quadrant affects knowledge accumulation in an area of science. Institutional differences play an important role in the type of research scientists produce—even when working in similar areas. In particular, if scientists need deep topic expertise to push forward the knowledge frontier and make scientific breakthroughs (Jones, 2009; Kaplan & Vakili, 2015; Agrawal et al., 2016), these may be less likely to occur if the balance of scientific labor shifts more heavily toward industry where scientists' research ranges more widely and appears to be less focussed on deep dives into particular topics.

2. Scientific Human Capital

2.1. Topic Knowledge and Research 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 phenomena that are of interest to a field (Merton, 1957; Dasgupta & David, 1994). As the "burden of knowledge" required to make these types of contributions increases, scientists are increasingly 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 typically analyzed 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). Other research has focused on scientists' relative experience with commercial or patenting activities as distinct parts of their human capital from knowledge developed through academic research (Toole & Czarnitzki, 2009; Baba, Shichijo & Sedita, 2009; Subramanian, Lim & Soh, 2013).

However, academic scientists also develop the specialized research skills needed to carry out advanced research in their field. 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, an economist who does research in applied microeconomics could have skills in one (or more) of microeconometric analysis, game theoretic modeling, structural modeling, general equilibrium modeling, among other research skills. Her human capital and research program could take several forms: (1) She could apply a narrow range of skills to a narrow set of topics in, e.g., labor economics; (2) She could apply a wide range of skills to a narrow set of topics in labor economics; (3) She could apply a narrow range of skills to a wide range of topics in, e.g., labor, development, public, or health economics; and (4) She could apply a wide range of skills to a wide range of topics in labor, development, public, or health economics.

If each of her skills and topics are counted as single constructs for determining whether her human capital makes her a generalist or specialist, a simple specialist-generalist continuum would rank (4) as a generalist; (2) and (3) as an intermediate generalist-specialist; and (1) as a specialist. This approach masks important differences in her human capital. In Scenario (2) she has highly specialized topic expertise with an extensive skillset that could be applied across a wider range of research areas. Her research skills also allow her to carry out a more comprehensive range of projects within her topics of expertise. In Scenario (3), she has wide-ranging topic expertise, but only a narrow skillset that can be used on research projects within these topics. The set of projects in which she could utilize her topic knowledge in future research projects either is limited to those for which she has the necessary research skills, or she must invest in acquiring new skills.

While topic knowledge and scientific research skills are both specialized forms of human capital, there are differences in the range of future research projects to which they can be applied. This has important implications for the ways in which firms can generate value from the human capital of scientists hired from academia. The research skills to design, execute, and analyze complex scientific projects using specific tools and techniques may be more readily transferrable to projects across a wider range of research questions than the highly specialized knowledge an academic scientist has developed in other types of topics. Insofar as research skills reflect knowledge of certain methods, coherent bundles of methodological knowledge can be conceptualized as a type of topic knowledge. For example, microeconometrics or social network analysis could be seen as topics within the economics or sociology disciplines. However, this bundle of knowledge has a distinct character of being applicable as a tool that underpins research projects across a wide range of other topics.

2.2. Goals of Academic and Corporate Science

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). Investing in scientific research can offer firms access to valuable external knowledge through scientific networks (Powell, Koput & 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 & Henderson, 1998; Gittelman & Kogut, 2003) and acquire insights into new knowledge through engagement with specialists in academia (Zucker, Darby & Armstrong, 2002; Fabrizio, 2009). Basic research can also be used to develop valuable new ideas internally that can be applied in firms' products (Arora et al., 2017). A foundational element of a basic science capability is scientific human capital. To absorb frontier scientific knowledge effectively, firms need scientists who can understand and apply external knowledge and who can build collaborative relationships with external experts through publishing, co-authoring, and attending scientific conferences (Liebeskind et al., 1996; Cockburn & Henderson, 1998; Simeth & Raffo, 2013).

However, while engagement with academic ideas is central to absorptive capacity, there is divergence in the respective goals of scientific research between academia and industry (Evans, 2010; Sauermann & Stephan, 2013). The reward system of academic science privileges major discoveries about fundamental phenomena (Merton, 1957; Latour, 1987; Dasgupta & David, 1994). On the other hand, firms' returns are based on understanding relationships between objects that can subsequently be developed for commercial ends (Evans, 2010; Sauermann & Stephan, 2013). Firms benefit from rapidly identifying new applications of knowledge that can be patented and form the basis of new products. These differing goals may also lead to divergence in the value that is derived from topic knowledge and research skills across industry and academic research.

To push forward the frontier of knowledge, scientists need increasingly specialized expertise at the forefront of a narrow field (Jones, 2009; Agrawal et al., 2016). Uncovering relational possibilities between objects may still require substantial subject knowledge, but also relies heavily on applying research skills to investigative projects across a range of scientific objects and phenomena. For example, Evans (2010) shows that academia-industry collaboration tends to produce research that is more distant from existing hubs of knowledge in a scientific field. The purpose of this diverse experimentation is to allow firms to uncover potentially valuable relationships before their rivals can do so. Compared to academia, a wide-ranging set of scientific research skills that enable scientists to test the wide-ranging potential relationships between objects may be especially important to industry employers. Cohen et al. (2002) note that, in many cases, R&D managers highlight research techniques emerging from academic science as being as, or more, important to their firms than specific research findings articulated by academic scientists. In this case, significant value for firms in corporate R&D comes from internalizing the technical skill of how to do something or apply some technique relevant to the firm in its R&D processes.

Firms 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 & Henderson, 1998; Zucker, Darby & Armstrong, 2002; Fabrizio, 2009). Increasingly, larger team collaborations are necessary to provide sufficient breadth of human capital in scientific research projects (Wuchty, Jones & Uzzi, 2007; Jones, 2011). Nonetheless, coordination costs increase as individuals with more specialized knowledge are required to execute a project (Becker & Murphy, 1992). Employing scientists with the ability to execute a more diverse range of projects at the individual level may be a more efficient way of covering the necessary range of knowledge and skills at the organization level than relying on a greater number of scientists with more narrow skills.

2.3. Transitions between Academia and Industry

Most empirical research on scientists' decisions to work in industry or academia, and the impact it has on their research outputs, is primarily focused on three specific types of academia-to-industry transitions. However, these groups may not be representative of many scientists moving to industry. First, research has focused on surveys of graduating doctoral students to understand their career decisions (Stern, 2004; Roach & Sauermann, 2010; Sauermann & Roach, 2014). 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 or for firms where they can participate more actively in the academic scientific community. 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 & Ding, 2006; D'Este & Perkmann, 2011; Roach & Sauermann, 2015; Azoulay, Liu & Stuart, 2016; Fini, Perkmann & Ross, 2022). This literature identifies an important role for both social norms and imprinting in early academic careers in scientists' decisions to become entrepreneurs (Stuart & Ding, 2006; Azoulay et al., 2016). It also shows that scientists becoming entrepreneurs do not appear to experience negative effects on their scientific output (Azoulay, Ding & Stuart, 2009; Hvide & Jones, 2018).

A third literature focuses on how the human capital of "star" scientists can positively impact a firm (Lacetera et al., 2004; Zucker, Darby & Torero, 2002; Rothaermel & Hess, 2007). These scientists are recognized by their peers as highly knowledgeable and influential scientists who can enhance a firm's 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 & 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. Beyond these highly specific contexts, less is known about the impact of transitions on scientists' research during the course of their careers. This is particularly pertinent as increasing collaborative activities between industry and academic scientists, changing academic norms, restrictions on the use of grants, and shifts in relative availability of research funds may induce greater labor mobility between academia and industry (Bozeman & Gaughan, 2007; Vallas & Kleinman, 2008; Roach & Sauermann, 2017).

2.4. Scientists' Human Capital and Research Direction

The prior literature typically approaches differences in industry and academic research in the context of the distinction between basic and applied science (Bush, 1945; Aghion et al., 2008; Lacetera, 2009). However, in Pasteur's Quadrant, where there is close alignment between the basic scientific value and 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 similar themes in basic research. On the one hand, the reward system of academia incentivizes scientists to focus on developing deep topic expertise through their research projects (Jones, 2009). On the other hand, a central way in which scientists add value in corporate research is by rapidly experimenting with the possibilities of scientific objects and phenomena to identify and understand relationships that may be of commercial value to the firm (Evans, 2010). Coordination and employment costs increase with the number of specialized workers needed for a project, which may make it costly to coordinate across large numbers of specialists with the necessary human capital (Becker & Murphy, 1992; Agrawal et al., 2016). Additionally, the potentially commercially valuable relationships may be distributed widely across a field of science, requiring firms to take on a diverse range of projects with respect to existing hubs of knowledge in which academic scientists have specialized (Evans, 2010). Therefore, it will be advantageous for firms to hire scientists with a wider range of research skills and exploit their ability to carry out a more diverse range of research projects within their wider area of expertise. Thus, we predict that scientists will apply their human capital across research projects involving a wider range of scientific concepts after transitioning to industry employment (compared to if they had remained working in academia). We also predict that firms will select for scientists whose past research demonstrates they have a wider range of skills. This background would indicate that their human capital could be used across more diverse projects to understand relationships between scientific objects that are potentially valuable to the firm.

To provide novel insights into the drivers of firms' scientific human capital strategies, we analyze the consequences of scientists' transitions to industry employment in the stem cell research field. We use detailed publication 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. We examine whether scientists' research output indicates that they are applying their human capital across a wider range of scientific concepts after moving to industry employment, or whether they continue to focus on the areas in which they already had expertise. We also analyze whether scientists who move from academia to industry have a wider range of ex ante experience working with different research techniques, tools, and interventional agents in their past research. This would suggest that firms place greater emphasis on these aspects of scientists' human capital in their hiring decisions.

If transitions to industry are associated with scientists using their research skills to apply their human capital more broadly, this suggests that industry employers are placing a greater emphasis on exploiting scientists' research skills more flexibly across projects compared to scientists' research strategies in academia. Alternatively, firms may not seek to utilize research skills across projects differently to academic scientists. In this case, we would expect to see that firms' scientific hires' research remains focused on concepts in their prior knowledge set in a manner similar to scientists remaining in academia. We would not expect to see scientists' research covering new concepts at a higher rate after they transition to industry employment. Additionally, if scientists are being hired by industry to carry out a more diverse range of research projects, firms may choose to hire scientists whose past research demonstrates greater experience working with a wider range of research techniques, tools, and interventional approaches in their areas of expertise. If scientists hired by firms do not have evidence of possessing a wider range of research skills, this would cast doubt on the idea that firms are placing value on the flexible application of these skills when making decisions on recruiting and utilizing scientific human capital.

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

& Toole, 2000; Murray & Stern, 2007; Azoulay et al., 2018, Bikard, Vakili & Teodoridis, 2019; Arora et al., 2019; Fini et al., 2022). Understanding how transitions between academia and industry affect scientists' use and development of their human capital helps us to understand how policy changes that affect the locus of research activity in a scientific field will impact knowledge accumulation within it. If scientists' research in industry leads them to explore scientific concepts 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 & Greve, 2006; Nagle & Teodoridis, 2020), higher (lower) relative industry employment may increase (decrease), the rate at which these occur.

In the next section, we describe our setting and empirical approach. The results in Section Four proceed in two parts. First, we establish how scientists' employment transitions from academia to industry are associated with changes in the range of scientific concepts they cover in their research. We carry out a range of robustness checks, examine whether the results are likely to be driven by wider changes in their research (such as carrying out more applied projects), and use instrumental variables to bolster confidence that this result is causal. We then examine how industry matches with scientific employees on the basis of different components of their human capital. We show that academic scientists are more likely to transition to industry employment if they have experience using a greater range of research tools and techniques. This is consistent with industry selecting for scientists with the ability to apply their human capital more flexibly in corporate research programs.

3. Research Design

3.1. Institutional Details

Our empirical setting for testing the 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 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). Both academia and industry paid significant attention to the new opportunities for medical treatments opened up by advances in stem cell science. It represented an archetype of Pasteur's Quadrant. Large volumes of research were carried out by scientists in both academia and industry, with significant scope for mobility between each setting (Blomfield & Vakili, 2022).

In this paper, we examine how moving to industry changed the research focus of U.S. scientists working in stem cell research. This helps shed light on how transitions to industry affect scientists' individual research outputs, the relative emphasis on exploiting 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, we are interested in the extent to which scientists are more (or less) likely to apply their skillset to a more diverse range of research projects after moving from academia to industry. This finding would suggest that firms' human capital strategies place relatively greater emphasis than academic science on using scientists' research skills to explore a more diverse range of new scientific ideas. Conversely, if scientists' research remains focused on exploiting their existing knowledge in a similar way to academia, this would suggest that industry values the topic knowledge and research skills components of scientists' human capital similarly to academia.

3.2. Data

We use 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 (Blomfield & Vakili, 2022). To identify stem cell publications, a keyword search was performed for the phrase "stem cell" or variants in titles, abstracts, or keywords of the articles contained in Scopus. We create a full career history for each scientist in our sample. For each scientist, author information and affiliations, co-author information, abstracts, and citation statistics were extracted for all papers authored until 2010 inclusive. We limit the sample to the 1996 to 2010 period to mitigate the risk that the results are affected by industry scientists being more likely to switch from active research to managerial tasks later in their industry employment. Scientists whose first publication appeared before 1976 were excluded 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 5,312 scientists with an identifiable affiliation during the sample period.

We extend this data by scraping the keywords associated with the articles from the PubMed database, which includes most journal articles published in the life sciences. PubMed contains precisely indexed keywords from a managed dictionary to describe the content of archived papers. The Medical Subject Headings (MeSH) vocabulary is managed by subject-specific experts at the National Library of Medicine (NLM). There are approximately 30,000 descriptor terms in the MeSH vocabulary, which are used to organize concepts in medicine and life sciences research into a hierarchical tree format. Indexing is independent of article authors. Terms are assigned by indexers at the NLM who select them based on a specific protocol. Each article is placed at a point in the space of scientific concepts based on its content on each dimension measured by index terms. For each author in the sample, we use this data to calculate the number of MeSH terms used each year, their cumulative count of unique terms used in their careers to date, the number of terms used for the first time in each year, and to examine directional changes in scientists' research.

3.3. Main Variables

Our core interest is how the conceptual content of scientists' research changes when they transition from academic to industry employment. The main dependent variable, *% New Descriptor Terms*, is defined as the percentage of MeSH descriptor terms indexed to articles published by a focal scientist i in a given year t that had not been previously indexed to scientist i's articles between t-1 and the year in which i has her first recorded publication in Scopus. It reflects the share of conceptual terms used in a given year that are new to the scientist's revealed antecedent knowledge set containing previously used terms. Thus, this variable represents the extent to which scientists' research covers scientific concepts that are new to their research relative to covering scientific concepts familiar to them.

MeSH indexing also allows descriptor terms to be combined with qualifier terms following a "/" symbol. Qualifiers provide more precision about which specific aspect of the concept associated with a MeSH descriptor term is the focus of the paper. For example, the MeSH descriptor term "induced pluripotent stem cells," is used to index research a specific type of stem cell that was a significant area of stem cell research toward the end of our sample period. Induced pluripotent stem cells are adult stem cells that can be reprogrammed to a pluripotent state from which they can then be developed into a wide range of cell types. The descriptor "induced pluripotent stem cells" may be combined with a number of qualifiers to denote what aspect of induced pluripotent stem cells is the precise topic of the paper. It is commonly combined with qualifiers such as "/cytology," "/metabolism," "/pathology," and "/transplantation" to denote different areas of focus in research on this type of cell. While our main dependent variable is based on the MeSH descriptor terms, we carry out additional analyses where each descriptor-qualifier combination is counted as a unique term for the purpose of measuring conceptual space (*% New Terms incl. Qualifiers*).

To ensure that our results are not simply driven by scientists' research covering a wider range of distinct terms that are highly proximate in conceptual space, we recreate our results aggregating terms at different levels of the MeSH tree hierarchy. We first associate the MeSH descriptor terms indexed to a paper with their digit-based codes in the MeSH tree hierarchy using the NLM's public crosswalk. Different MeSH descriptor terms can be located at different levels in the tree hierarchy. Some MeSH terms are also associated with multiple digit-based codes because they are relevant to multiple higher-level concepts. When creating the variables based on levels in the MeSH hierarchy, we weight each term by one divided by the number of codes to which they are linked. We then aggregate the codes at the 9-digit level, which is the fourth tier in the MeSH hierarchy. Some—but not all—MeSH terms are organized at a lower, more granular, level than the fourth tier in the MeSH hierarchy.

Using this consistent level of analysis helps to ensure that the results are not driven by scientists' research covering a wider range of terms at a fine-level of granularity that are closely linked and classified within a single 9-digit MeSH category. Publications that cover more new terms in the same 9-digit MeSH category may not reflect that a scientist is taking on research projects that cover a wider range of concepts that are new to the scientist when compared to publications covering a similar number of new terms (or fewer new terms) that are distinct at the 9-digit level. We next repeat this process aggregating MeSH terms to the third level of the hierarchy (the 6-digit level) and replicate our analysis. This leads to the two additional dependent variables: *% New Terms, 9-digit Level* and *% New Terms, 6-digit Level*. Where a scientist has no publications in a given year, these are undefined in the panel (following other studies that analyze changes in scientists' research based on bibliographic data, e.g., Azoulay et al., 2009; Evans, 2010).

As a robustness check, we use additional dependent variables that count the total number of new MeSH terms. This is defined as the number of 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. We create this variable at each of four levels described above. This does not account for the rate of conceptual exploration relative to the exploitation of previously covered conceptual space—i.e., the extent to which scientists' new knowledge and existing knowledge sets are distinct—as does the percentage-based variable. 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. Finally, we perform additional robustness checks in which we redefine our dependent variables as only including a new term when this term has not been used from t-1 to t-5 (as opposed to in any year in a scientists' career prior up to t-1 in the baseline models).

To provide further insights into how employment transitions from academia to industry are associated with changes in scientists' research, we analyze three further dependent variables. These are the scientist i's number of publications, the average number of citations received by her publications, and the median CHI journal ranking of her publications in year t. The purpose of the variables is to ensure that scientist-level changes in the number of publications, quality of publications, or changes in their focus away from basic to more applied research are not driving the main results. For example, scientists may explore a greater range of new conceptual space per publication but have fewer or less impactful publications. 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. 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 four-point scale. Where scientist i has one or more publications in year t, we take the CHI scores for the journals in which her articles were published (if ranked in the CHI index) and calculate the median value across all of her publications from that year. This variable provides insights into whether the same scientist performs more applied research when moving to industry. It shows whether scientists are covering more conceptual space, not simply because they are applying their skills more widely over topics, but because they are qualitatively changing the type of research projects they undertake. This may then lead them to appear to cover more conceptual space. However, this may be driven by increased use of new terms that—while distinct—are closely related downstream applications of their prior knowledge.

Our core independent variable of interest is *Corporate Affiliation.* This is an indicator variable equal to one for all scientist-year observations in which scientist i has a corporate affiliation on an article, and zero otherwise. If an individual's affiliation is not directly observed on a publication by scientist i in year t, we use the most proximate directly observed affiliation. When examining whether the NLM has indexed new MeSH terms to a scientist's publications in a given year, we control for the natural logarithm of one plus the count of unique terms that have previously been indexed to that scientists' papers. This control for the stock of prior terms is intended to mitigate the risk that the results are driven by a mechanical correlation in which scientists with more prior publications have less scope to cover additional concepts in proximate areas of science when undertaking new research projects (limiting the potential for new terms to be indexed to their papers).

For similar reasons, we 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. We also control for the natural logarithm of one plus the cumulative number of citations received by the scientists in years prior to t. This is intended to ensure the results are not due to heterogeneity in scientists' quality or the impact of their prior work. Finally, we include a fixed effect for the 'age' of the scientist. This is calculated as the difference in years between the focal year t and the year of her first recorded publication. The purpose of this fixed effect is to ensure that changes in scientists' research linked to their career stages are not driving the results. For example, this may be the case if scientists are more likely to transition to industry employment at certain stages of their career that also corelate with their propensity to explore new concepts in their subsequent projects.

3.4. Econometric Approach

We compare changes in the research output of U.S. scientists acquiring and maintaining corporate affiliation to scientists remaining in academia. We use scientist fixed effects to isolate within-scientist changes in research direction as scientists transition between academic and industry employment. To analyze changes in scientists' research direction, we use the following specification for our core analysis:

$$Y_{it} = \beta_0 + \beta_1 CorporateAffiliation_{it} + \Delta X_{it} + s_i + \tau_t + \varepsilon_{it}$$

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 the vector of control variables as described above. Year fixed effects, τ_t, are included to control for time-varying factors affecting all scientists. Time-invariant individual characteristics are controlled for using scientist fixed effects, s_i. Thus, the analysis can be seen as isolating the within-scientist changes in research direction that are associated with acquiring a corporate affiliation. A core assumption underlying our empirical approach is that changes in scientists' research after acquiring industry affiliations are (at least in part) driven by their employers' decisions regarding how to use their human capital. On the one hand, this could involve employers giving strong directions about which projects to undertake. Alternatively, a firm may make broad decisions about the areas of interest and goals of their research programs, while scientists choose how to explore within these parameters. In either case, the 'treatment' would result from employment in industry research.

3.5. Instrumental Variables

Scientists transitioning to industry may strategically time transitions or systematically differ in unobserved characteristics to other scientists that make them more attractive to industry employers and are linked to changes in their research focus. For example, scientists may switch to industry employment immediately after their research uncovers a particularly valuable result for commercial exploitation. Alternatively, they may switch to industry in order to carry out research in new areas if funding policies make it more difficult to acquire resources for research on topics a scientist has not previously researched. An ideal experiment to show that scientists' research changes when moving to industry would involve the random assignment of scientists to industry. In this paper, we use 2SLS analysis using instrumental variables based on changes in local taxes and national shifts in labor demand to generate plausibly exogenous variation in the assignment of scientists to types of employers. This generates a local average treatment effect for those scientists who on the margin are the most indifferent between industry and academic employment. Specifically, we use variation in taxes and employment rates across U.S.

states over time as instruments that affect the opportunities for scientists to take industry employment. In the 2SLS analysis, we draw on the empirical work of Moretti & Wilson (2017) on how scientists' labor market decisions respond to changes in tax policy. We thank the authors for making their data publicly available.

The instrumental variables focus on exogenous changes in the demand side of scientific labor markets in a state at a particular point in time. In the results section, we show that there is no evidence of individual scientists changing their rate of conceptual exploration in their research prior to transitioning to industry employment. This suggests they are not making changes to their research before moving to industry in anticipation of moving to industry. Nonetheless, the OLS estimates may be biased if many scientists' motivation for moving to industry is either to exploit a commercially valuable finding from their prior research (which would lead to a downward bias on the OLS estimates) or to switch to doing a different kind of research (which would lead to an upward bias on the OLS estimates). This would mean the OLS estimates are partially driven by systematic differences in scientists' desire to move to industry employment based on their anticipated research trajectory. By focusing on changes in employment linked exogenous variation in the demand for scientific labor, we can identify the effects on scientists who were otherwise the most indifferent to working in industry compared to academia and only move to industry due to demand shocks.

The first instrument is the state-level R&D tax credit. This is based on the tax rates in the state in which scientist i worked during the year t-1. For a scientist's first publication year, we do not observe their location at t-1. In this case, we assume that scientist i was located in the same state during the year t-1. This variable captures the relative cost to firms in that state of investing in R&D based on the R&D 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 shifts in demand for scientific labor over time. 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 R&D 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). We begin by defining a state's pre-sample level of employment in this category based on its level in 1990, providing a five-year gap to the first sample year. We then adjust this value for each future yearly observation by multiplying it by the percentage growth in employment in NAICS category 54171 in all other U.S. states since the baseline year (i.e., the national rate of job growth excluding growth in the focal state) to create a yearly predicted value of employment opportunities in the state. We then take the natural logarithm of this value to generate an instrumental variable based on the state's predicted scientific R&D employment. This takes a Baritk-style approach with an initial variable from pre-sample historical base period receiving exogenous shifts (Bartik, 1991; Blanchard & Katz, 1992; Card, 2001; Autor, Dorn & Hanson, 2013). We exclude the scientist's state of residence from the calculation to ensure that changes in the instrument's value are not driven by any potential endogenous changes during the sample period in a scientist's own state that may increase both employment opportunities and potentially affect the research strategies of scientists in that state. We again construct the instrument for the state in which scientist i worked during the previous year and carry out a similar adjustment for a scientist's first observed year.

This variable is intended to provide variation in the number of industry employment opportunities for research scientists in a state based on national shifts in demand for scientific labor. This, in turn, should be based on changes in the returns to firms of employing scientists taking place in other states (and hence not be endogenous to firms or scientists that were previously in the focal state). For example, if wider economic factors lead firms to view employing scientists as more valuable, firms in other states should increase their scientific employment. However, this would not be driven by factors linked to specific scientists or firms in a focal state. Therefore, it captures the relative size of the set of opportunities a scientist would have to transition into industry employment based on changes in labor demand for scientists in other states.

The key identifying assumptions are that the changes in R&D tax credits and predicted labor market opportunities are only associated with changes in the number of opportunities for scientists to work in industry and the attractiveness of industry employment. If R&D tax credits increase, it becomes more profitable to employ a scientist, other things being equal.

Likewise, if there are exogenous shifts in the return to corporate science, this will increase corporate demand for scientific labor. Therefore, if the short run supply of highly trained scientific labor is somewhat constrained, firms may improve wage offers to attract scientists. This would induce some scientists to move to industry who would otherwise have remained in academic employment. Identification requires that these instruments are not correlated with changes in preferences about which topics scientists choose to research within their existing employment type. In particular, we require that scientists do not anticipate greater opportunities for transitioning to industry employment in the future and, because of this, change their research trajectory. In Section 4, we show that there is no evidence that scientists change their research trajectory prior to moving to industry employment. Notably, the rate of conceptual exploration for scientists who switch to industry employment is almost identical in each of the five years prior to this employment transition.

The 2SLS estimates represent an average of the local average treatment effects for those scientists at the margin who are most likely to transition to industry when industry demand for scientific labor exogenously changes through each of these channels. While scientists are still choosing to work for industry, the instruments are capturing the treatment effect of industry employment on scientists who would otherwise be in academic employment (were it not for changes in industry's demand for scientific labor for reasons that are unrelated to the scientists' own research). In addition to using instrumental variables for labor market transitions, we carry out additional tests replicating our results using a similar set of scientists based in countries outside the United States. This is to show that the results are not driven by any unobserved factors that are specific to the U.S. context.

3.6. Components of Human Capital

If firms place additional emphasis on utilizing scientists' more flexible research skills in their human capital strategies, there should be evidence of this in the types of research that scientists have carried out before they are hired by a firm. To examine this question, we separate MeSH descriptor terms in the sample according to their highest-level thematic categorization, which represents the first branch in the MeSH tree. Five of the sixteen categories account for approximately 10 percent or more of the terms indexed to the papers in the sample. They collectively account for approximately 75 percent of all terms. These are "Anatomy," "Diseases," "Processes & Phenomena," "Chemicals & Drugs," and "Analytical, Diagnostic, and Therapeutic Techniques & Equipment."

The Anatomy category classifies the system in the body or type of tissue with which a paper is concerned. The Processes & Phenomena category comprises the physiological processes involved in the research. The Chemicals & Drugs category contains the chemicals that can be used as interventional agents to affect biological processes or disease. The Techniques & Equipment category contains different tools and techniques that can be used to manipulate, diagnose, or treat health conditions, in addition to a range of analytical approaches in research. The Disease category comprises the diseases, disorders, and injuries affecting health. The Organisms category is excluded as a primary category of analysis despite containing approximately 10 percent of MeSH descriptor terms in the sample. This is because it has three very common terms "Animals," "Humans," and "Mice" that collectively account for more than 65 percent of observations in that category. These three terms are respectively between 5 and 10 times more common than the next most frequently occurring term in the other five categories.

We can provide an intuitive illustration of how the MeSH category classification works using the example of recent advances in gene editing research based on CRIPSR systems with diabetes as an illustrative disease. CRISPR/Cas systems as an antiviral defence mechanism are classified in the Processes & Phenomena category. Gene editing as a clinical or research tool/technique is classified under Techniques & Equipment. The CRISPR-Associated Proteins that can be used to edit genes are in the Chemicals & Drugs category. Diabetes as a target of gene editing techniques would be classified under Diseases. Finally, the pancreas or pancreatic cells being edited would be classed under Anatomy. Further information on the most frequently occurring MeSH terms in each category in this sample is provided in the Online Appendix.

While the MeSH vocabulary is not designed primarily to distinguish between scientific concepts that reflect research skills and topic knowledge, different categories imply different types of expertise. In particular, the Techniques & Equipment category contains terms that denote the tools and techniques scientists use in their research for analytical, diagnostic, or therapeutic purposes. These terms represent the actions the scientists are taking in their research underlying a publication. In particular, this category indexes: the research tools and investigative approaches that scientists use to analyze a particular phenomenon; the different techniques scientists use that form the basis of potential treatment approaches and, the analytical methods for identifying diseases and analyzing the feature of cells and tissues. While MeSH terms in other categories imply expertise about an object or process, these terms reflect expertise in the specific actions that scientists can take in the process of their research, or the tools they can use to analyze, diagnose, or otherwise affect these objects. If industry places more emphasis on scientists' ability to experiment across a range of scientific relationships to uncover those that are potentially commercially valuable, scientists' ability to use a wide range of tools and techniques in experiments may be especially valuable to firms.

We analyze the role of the different components of scientists' human capital in transitions from academic to industry employment using two approaches. First, we measure scientists' knowledge set in each of these five MeSH categories as the natural logarithm of one plus the total number of terms indexed to scientist i 's research in years strictly prior to the focal year t that are in MeSH category m. This variable represents the volume of knowledge a scientist has of concepts in a given category. Second, we calculate a diversification index for each of the five MeSH categories. This is equal to one minus the square root of the Herfindahl index for scientist i in MeSH category m based on the distribution of terms that they had used in category m in years strictly prior to the focal year t. Models using the count-based variable show the link between the size of a scientists' knowledge set in a given MeSH category and transitions to industry employment. Models using the diversification index variable provide more nuance by also incorporating how far scientists specialize on concepts within their knowledge set or use a greater range of the concepts it contains with more frequency. For example, using this variable a scientist whose research papers have covered ten distinct MeSH terms in a category three times each is less specialized than a scientist whose papers have covered one term twenty-one times and a further nine terms each on one occasion.

We use discrete time hazard models to analyze the relationship between scientists' knowledge and skills in different MeSH categories and the likelihood that that they transition to industry employment for the first time. We use complementary log-log regression models in which the dependent variable is an indicator variable that is equal to one in year t if scientist i has her first recorded corporate affiliation in that year (and zero otherwise). Scientists are dropped from the estimation sample following the first year (if any) in which they have a corporate affiliation. Since scientist fixed effects cannot be included in this type of model, we include an additional cohort fixed effect (defined as the year in which scientist *i*'s first publication appears in Scopus). This is to ensure that the results are not driven by scientists' who began their careers in different time periods having systematically different research trajectories and rates of transitioning to industry employment. We again include year and age fixed effects and publication and citations controls.

An ideal experiment would randomly assign human capital in different categories to scientists and have them apply to industry jobs to identify how differences in ex ante human capital are causally linked to scientists being offered jobs. In this analysis, we observe differences in the human capital of scientists who do and do not transition to industry employment over time. The core assumption in this analysis is that the matches we observe reflect firms' preferences for who to hire and that they are not caused only by scientists with wider ranging knowledge or skills in a category also having greater preferences for industry employment. The primary purpose of the analysis is to provide supportive evidence that firms place particular value on the breadth of scientists' skills in their human capital decisions. Table 1 below provides summary statistics of the main variables in the paper. Table A.1 in the Online Appendix provides summary statistics for the additional variables that are defined separately for each MeSH category.

[INSERT TABLE 1 HERE]

4. Results

4.1. Scientists' Exploration of New Conceptual Space

We begin by examining how the acquisition of an industry affiliation is associated with changes in scientists' rate of exploring new areas of MeSH space in their research. The analysis in Table 2 shows how the rate at which scientists use new MeSH terms relative to reusing existing terms changes after transitioning to industry employment. Models 1 and 2 use our preferred dependent variable *% New Descriptor Terms*. They show that the percentage of descriptor 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. Model 1 includes scientist and year fixed effects, but no other control variables. Model 2 includes the full set of controls. The point estimates in Model 2 suggest that scientists' share of new MeSH terms in their publications increases by approximately 2.1 percentage points following transitions to industry. This is equivalent to a 4.4 percent increase in the share of terms that represent new scientific content relative to the baseline rate of conceptual exploration.

[INSERT TABLE 2 HERE]

The results are similar when concepts are measured at different levels of aggregation in the MeSH hierarchy. Models 3 to 5 report results from the full model for the dependent variables in which MeSH terms: include qualifier modification; are aggregated at the 9-digit level; and are aggregated to the 6-digit level. The point estimates in Model 5 suggest that (at the highest level of aggregation) scientists' share of new terms in publications increases by approximately 1.4 percentage points following transitions to industry. This is equivalent to a 5.2 percent increase on the baseline rate of exploring new areas of conceptual space defined at the 6-digit level. Panels A to D of Figure 1 plot scientists' conceptual exploration for a ten-year window from five years before to five years after scientists' transition to industry employment for each of the dependent variables. There is little evidence of a pre-trend in which firms are selecting scientists who are already increasing their rate of exploration of new conceptual space. Instead, scientists' rate of conceptual exploration increases only after transitions to industry.

[INSERT FIGURE 1 HERE]

As a robustness check, Table A.2 replicates the results using the count of new terms as a dependent variable in Poisson regression models. This is to ensure that the results are not being driven by scientists' publications covering less total conceptual space after transitions to industry. The point estimates across models imply that the rate at which scientists' research is indexed to new terms increases by approximately 7 percent following transitions to industry. Table A.3 limits the analysis to MeSH terms that a scientist has not used in the previous five years. This it to ensure that the results are not driven by concepts that a scientist last covered in their research many years prior and have not been part of their recent research focus. The point estimates and standard errors of these coefficients are again highly consistent with those presented in the main analysis. Overall, the results show that, in our sample, scientists' research covers more concepts that are new to the scientist after transitioning to industry employment.

4.2. Other Changes in Scientists' Research

The MeSH term-based dependent variables that measure a scientists' research direction are based on observed research outputs. However, there may be changes in the rate at which involvement with industry leads scientists to be subject to secrecy or publication delays to protect valuable ideas from leaking to competitors (Czarnitzki, Grimpe & Toole, 2015). For example, we could be under-counting the number of new MeSH terms covered in scientists' research when working in industry (since we would only be observing part of their overall research portfolio). To examine whether changes in publication habits or wider changes in scientists' research may be driving the main results, we replicate the core analysis with a scientist's number of publications, number of citations received by their publications, and the median CHI ranking of the journals in which they publish in year t as 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 a similar number of citations to those in academia conditional on their number of publications in a year. This suggests the results are not driven by changes in the quality of a scientist's papers after moving to industry. Finally, Models 5 and 6 show that the median CHI ranking of the journals in which a scientist published (which measures the "basicness" of the typical research paper a focal journal publishes) appears to be unchanged after scientists move to industry. Overall, the 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 ideas of more marginal scientific value. Nonetheless, the results in Table 2 show that they are applying their scientific human capital across a wider range of conceptual space.

[INSERT TABLE 3 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 (Gottweis, 2010; Furman, Murray & Stern, 2012; Blomfield & Vakili, 2022). 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 subfield 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 the results, we replicate the analysis including only the observations from the period before President Bush took office in 2001. The results presented in Table A.4 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 to the existing balance between scientists' research choices and funding provision. It would be consistent with some scientists moving to industry to pursue their existing research interests in the more uncertain federal funding climate immediately following President Bush's election (Blomfield & Vakili, 2022). As an additional check that the results are not primarily driven by changes in U.S. science policy during the sample period, we also replicate the results using a sample of scientists based in countries outside the United States in Table A.5. More information on this sample is provided in the notes to this table in the Online Appendix. Again, the point estimates are significant and slightly larger than in the core analysis suggesting the main results are not driven by U.S.-specific factors.

4.3. 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. In Table A.6, we replicate the core results for each dependent variable splitting the sample according to scientists' ability, cohort, and career stage. In Panel A, to separate scientists by ability, we calculate the median number of citations received by scientists to papers published during the pre-sample period from 1990 to 1995 inclusive. We 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. We then split the sample according to whether a focal scientist has an above or below median number of citations during this decade. We 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. We also exclude those who have no observed publications prior to 1996 because we cannot make a pre-sample calculation for these individuals. In Panel B, we split the sample according to the year in which scientists had their first publication (whether it was strictly before 1990) to examine whether the result is driven by different cohorts in the sample. In Panel C, we split the sample according to scientists' career length prior to an observation (based on the time elapsed since their first publication and with the cut taking place at 12 years). The results are consistent across subsamples in each of the ability, cohort, and age split-sample analyses.

Next, we include a control variable for the number of co-authors a scientist has in a given year and an interaction between the number of co-authors and the industry affiliation indicator variable (the results are presented in Table A.7). This is to check whether changes in co-authoring patterns that coincide with changes in affiliation type may be driving the results. The results 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 they 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. If this were the case, it would then 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 rate of conceptual exploration is positive and significant across models.

4.4. Returnees from Industry to Academia

If the relevant mechanism is industry employment, scientists returning to academic employment should return to their pre-industry rates of conceptual exploration. To test this, we examine the research outputs of scientists who transition to industry employment from academia during the sample period and then subsequently return to academia. The results in Table 4 show that scientists who work in industry have lower rates of exploring new scientific conceptual space upon returning to academia than they had during industry employment. The omitted category in the regression models is scientists' academic employment before transitioning to industry. The coefficients on the variables of interest can be interpreted as reflecting changes relative to this baseline.

[INSERT TABLE 4 HERE]

There is a consistent pattern of results across models. Scientists engage in conceptual exploration at a higher rate upon acquiring an industry affiliation than during their prior academic employment. However, upon returning to academia, they once again exhibit lower rates of conceptual exploration than during the period in which they are employed in industry. Scientists' rates of conceptual exploration upon returning to academia is very similar to that of their pre-industry academic employment across models. This also suggests that increases in scientists' rates of conceptual exploration are not driven by any type of transition to a new employer. It is specifically transitions to industry that led to this change in scientists' research direction. Transitions back to academia do not have similar results. Instead, scientists' rates of conceptual exploration return to very close to their pre-industry employment levels. This is consistent with an industry-specific effect on scientists' exploitation and further development of their human capital.

4.5. Using Instruments for Employment Transitions

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 bias OLS estimates downward as scientists would be selecting into industry employment in order to exploit already identified commercially valuable opportunities and thus be less likely to explore new topics. The results would then not reflect the true effect of industry employment, but rather be driven by selective entry into industry employment to exploit existing knowledge. Since labor market outcomes are a matching process, the agency leading to selection may be on either or both sides of the market: scientists or firms. 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 & David, 1994; Stern, 2004; Sauermann & Stephan, 2013). However, industry may offer an advantageous setting to change research direction if highly specialized knowledge (and the research output demonstrating this to funders) is more important to access resources for research in academia. This would cause upward bias in the OLS results.

Either of these selection effects are intuitively plausible. Therefore, we use instrumental variables analysis to provide evidence that our main effects reflect a causal relationship. We use two instrumental variables based on state R&D tax credit rates and changes in the demand for scientific labor market opportunities as instruments for scientists' labor market affiliations. Each instrument varies both over time within a state and between different states in a given year. The instruments provide exogenous variation in the likelihood scientists will be employed by industry. The local average treatment effects from the 2SLS models provide evidence of the causal effect of industry employment among those scientists who would transition to industry when exogenous changes make it more attractive to firms to hire scientists, but who otherwise would remain working in academia. Table 5 presents the results of the instrumental variables regressions.

[INSERT TABLE 5 HERE]

There is a positive relationship between industry employment and scientists' research covering more new conceptual space for both types of dependent variable. The estimates of the coefficients on the industry employment variable are positive and larger in magnitude than the baseline results. This implies that the first of the two proposed selection mechanisms has a greater impact on the results in the OLS models. The core results would be biased downward by some scientists selecting into industry employment to exploit commercially valuable existing knowledge. The instrumental variables results contrast with a selection mechanism in which scientists choose to enter industry employment in order to change their research direction (which would have led to upward bias in the OLS results). This also suggests there are two important and distinct channels through which entry into industry employment is linked to changes in scientists' research direction. In the first, there is selection into industry employment of scientists to exploit specific valuable knowledge, which is then a focus of future exploitation. In the second, for the mass of scientists who form the backbone of the corporate scientific labor force and who choose industry jobs if they are a more attractive type of employment, these transitions lead to more conceptually diverse future research output. Additional results are presented in the Online Appendix showing that the results are insensitive to whether standard errors are clustered at the scientist or state level and similar when only including one of the two instrumental variables in the model.

4.6. Human Capital of Scientists Moving to Industry

If firms are placing additional emphasis on scientists' more flexible research skills in their human capital strategies, there should be evidence of this in the types of research that scientists have carried out before they are hired by a firm. In Table 6, we use discrete time hazard models to examine how differences in scientists' ex ante human capital across MeSH categories are associated with the probability that they transition to industry employment for the first time in their career. In Models 1 and 2, we measure the scientist's knowledge set in each of the five MeSH categories above as the count of the total terms used by scientist i in years strictly prior to the focal year t. The columns report estimates of the average marginal effect for each MeSH category (Table A.12 reports the marginal effects at means to indicate that the effects are relatively consistent at this point in the distributions). The results show that scientists transitioning to industry employment for the first time have used a significantly greater range of research techniques and tools in their past research than scientists remaining in academia. Transitioning scientists have also worked with a wider range of chemicals and drugs. The most frequently occurring terms that comprise this group of terms are "messenger RNA," "monoclonal antibodies," and "DNA-binding proteins." This reflects that the scientists transitioning to industry employment have experience using a wider range of potential interventional agents in their prior research.

[INSERT TABLE 6 HERE]

These results are consistent with industry selecting scientists who have experience using a wider range of techniques, tools, and interventional agents in their past research. However, more extensive knowledge of anatomical and disease topics is not positively associated with these transitions, suggesting this latter type of expertise is valued similarly in industry and academia. The estimates of the average marginal effects suggest that a one standard deviation increase in the techniques variable is associated with an approximately 0.3 percentage point increase in the rate at which scientists transition to industry employment on average. For chemicals and drugs, this is approximately 0.4 percentage points. Relative to the baseline probability that a scientist transitions to industry employment in a given year of 1.3 percentage points, these are equivalent to increases in transition rates of approximately 20 and 30 percent respectively.

The results in Models 3 and 4 of Table 6 use diversification index-based independent variables. Notably, experience using a more diverse range of techniques and tools is again associated with a higher likelihood of transitioning to industry employment. The point estimate of the average marginal effects in Model 4 implies that a one standard deviation increase in diversification in this MeSH category is associated with a 0.2 percentage point increase in the rate of transition to industry employment on average. This is equivalent to approximately a 15 percent increase in the rate of transitions to industry relative to the baseline. Scientists whose past research has covered a more diverse range of processes in the body are more likely to transition to industry. In our sample, these are most commonly processes taking place in cells that scientists are using skills to manipulate in their research project. Our most frequent terms are "cell differentiation," "signal transduction," "mutation," "apoptosis," "transfection," and "gene expression regulation."

The results in Table 6 indicate that firms are selecting scientists' who, not only have a larger set of technique and tool knowledge, but also who more frequently use a wider range of them (rather than predominantly specializing in some of these techniques or tools). Again, there is no evidence that scientists transitioning to industry are more or less specialized than other scientists on disease or anatomical topics. Overall, it does not appear to be the case that firms simply focus on hiring generalists or specialists in terms of scientists' overall scientific human capital. Rather, industry appears to value scientists who have demonstrated the skills to use a wider range of techniques and tools within their areas of expertise. It is important to note that each category of the MeSH vocabulary contains a diverse set of terms. Further research is needed to identify to what extent the relationship between industry employment and types of knowledge and skills may vary across research fields.

5. Discussion

Past research has typically analyzed scientists' human capital according to specialization or generalization of topic knowledge (Agrawal et al., 2016; Teodoridis, 2018; Teodoridis et al., 2019) or by the levels of experience in basic and applied research (Toole & Czarnitzki, 2009; Baba et al., 2009; Subramanian et al., 2013). In this paper, we argue that it is also important to understand scientists' specialized topic knowledge and advanced scientific research skills as distinct components of scientists' human capital. In our analysis we examine how these components are utilized by firms in corporate scientific research and the role they play in employment transitions. This delineation between aspects of scientists' human capital is conceptually distinct from those typically drawn in the prior literature. 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 transition to industry employment have more conceptually diverse subsequent research output. This is consistent with industry placing greater value on hired scientists' advanced research skills that can be applied more widely as part of their human capital strategies. In line with this interpretation, we find evidence that firms select scientists with experience using a greater range of tools, techniques, and types of interventional agents in their research prior to entering industry employment. Scientists with more wide-ranging knowledge of diseases or parts of the human body are not more likely to be hired by firms than other scientists in academia. However, scientists who have used a greater range of research techniques, tools, and interventional modalities do appear to be more attractive to firms. Returning to our earlier example, an economist with narrow topic knowledge but more wide-ranging skills may appear to be similar on a one-dimensional specialist-generalist continuum as an economist with more wide-ranging topic knowledge but a narrower skillset. However, in the context of firms hiring scientists from academia, our results suggest that researchers with these two human capital configurations are quite different. The individual with the wider skillset appears to be more attractive to firms and is more likely to transition to industry employment.

We argue that these results emerge due to differences in the logic and goals of scientific research in academia and industry. The reward structure of academia privileges fundamental insights into novel phenomena (Merton, 1957; Dasgupta & David, 1994). For scientists working in industry, their employer's primary rewards come from the role of science in speedily identifying potentially valuable relationships for commercialization (Evans, 2010; Sauermann & Stephan, 2013). Firms may benefit more from applying scientists' research skills to uncovering these types of relationships across a greater range of conceptual space. Evans (2010) shows that academic scientists tend to explore ideas further away from the core areas of academic scientific focus when collaborating on research projects with industry. In this paper, we complement these findings by showing that academic scientists transitioning to industry employment subsequently have a higher rate of exploring conceptual space that is new to the scientist in their research. We also show that industry appears to select scientists on the basis of having more widely extensive research experience using a variety of techniques, tools, and interventional agents. This is consistent with industry selecting for scientists with the ability to apply their human capital more flexibly on these dimensions.

Whereas prior literature has focussed on the distinction between basic and applied research when examining how 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: the rate of exploring new conceptual space. We show that a scientist's human capital tends to be employed in a more focused area of conceptual space in academia compared to industry—even though the basicness and impact of their research may be similar. In our setting, we find that industry employment leads scientists' research projects to explore new scientific concepts at a higher rate. Conversely, academic employment leads scientists to have narrower search strategies that are associated with building more focused deep topic expertise.

Our results also contribute to our understanding of scientific labor markets. Career transitions from academia to industry have been relatively understudied outside of entrepreneurial and early career settings or the specific case of star scientists. Our results show how transitions affect how scientists exploit their human capital and how their subsequent research adds to their knowledge. This helps us to understand the role that specialized scientific knowledge and skills appear to play in firms' human capital strategies. More widely, it also points to important ways in which the division of labor between industry and academia may affect the type of scientific knowledge accumulated in a scientific field. In particular, industry employment leads scientists to cover more conceptual space that is new to the scientist in their research.

Where the balance of innovative labor in Pasteur's Quadrant shifts to industry, scientists' different research strategies to exploit and further develop their human capital may affect the type of knowledge produced in their field. Scientists may develop less of the deep topic knowledge associated with academic science that prior research suggests is necessary to make major scientific breakthroughs (Jones, 2009; Kaplan & Vakili, 2015). Policymakers may need to account for these consequences for aggregate knowledge accumulation when designing policies that promote mobility between academia and industry. Future research could helpfully identify how depth of knowledge on different dimensions of scientific human capital is linked to scientists' propensity to make major scientific breakthroughs.

An extensive literature, primarily in economics, has sought to divide human capital into different components based on its specialization with respect to occupations, employers, and tasks (Becker, 1962; Gibbons & Waldman, 2004; Lazear, 2009; Gathmann & Schönberg, 2010). To the extent that research skills are more widely applicable to research tasks than advanced knowledge of a given topic, it may be a more general form of human capital applicable to a larger number of tasks. These skills are not necessarily less specialized; developing advanced research skills may require major time and effort investments. Skills in novel methods may be rarer than advanced topic knowledge in some areas of science before they diffuse to become a more common part of scientists' skillsets over time. Future research could elaborate which types of 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 is an imperfect proxy for scientists' knowledge stocks. Although the vocabulary indexed independently by the NLM allows us to define the conceptual space in which a scientist has published, this may miss important parts of the space in which they have knowledge but have not published. The broad MeSH categories also represent imprecise proxies for skills using particular tools and techniques. Future research that provides a more comprehensive matching between granular MeSH terms and dimensions of human capital would make an important contribution to this literature.

Additionally, we do not examine how firms' scientific human capital decisions relate to other mechanisms through which firms can internalize valuable external knowledge. Prior research has examined how firms can take advantage of complementarities between internal and external knowledge resources (Cassiman & Veugelers, 2006; Fabrizio, 2009). Further exploration of the links between hiring and collaboration strategies—particularly beyond the role played by star scientists—could generate valuable insights here. Future research might valuably identify how different types of firms vary in their use, and hiring, of scientific labor. It may be fruitful to examine whether differences in the extent to which firms apply the human capital of scientists across different research areas 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 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.

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*, 25893.

Arora A., S. Belenzon, and L. Sheer (2017) “Back to basics: Why do firms invest in research?” *NBER Working Paper*, 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.

Autor, D., D. Dorn, & G. Hanson. (2013). The China syndrome: Local labor market effects of import competition in the United States. *American Economic Review,* 103(6), 2121–2168.

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.

Bartik, T.J. (1991). *Who benefits from state and local economic development policies?* Kalamazoo, MI: WE Upjohn Institute for Employment Research.

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

Becker G.S., and K.M. Murphy (1992) “The Division of Labor, Coordination Costs, and Knowledge,” *Quarterly Journal of Economics*, 107(4): 1137–1160.

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.

Blanchard, O., and L. Katz. (1992). Regional Evolutions. Brookings Papers on Economic Activity, 1992(1): 1–75.

Blomfield M., and K. Vakili (2022) “Incentivizing Effort Allocation Through Resource Allocation: Evidence from Scientists' Response to Changes in Funding Policy,” *Organization Science*, forthcoming.

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.

Card, D. (2001). Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics*, 19(1): 22–64.

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 (1989) “Innovation and learning: the two faces of R & D,” *Economic Journal,* 99(397), 569–596.

Cohen W.M., and D.A. Levinthal (1990) “Absorptive Capacity: A New Perspective on Learning and Innovation,” *Administrative Science Quarterly*, 35(1): 128–152.

Cohen W.M., R.R. Nelson, and J.P. Walsh (2002) Links and impacts: the influence of public research on industrial R&D. *Management Science*, 48(1), 1–23.

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, and 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.

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.

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

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

Furman J., F. Murray, and S. Stern (2012) “Growing stem cells: The impact of U.S. policy on the organization of scientific research. *Journal of Policy Analysis and Management* 31(3): 661–705.

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.

Gottweis H. (2010) “The endless hESC controversy in the United States: History, context, and prospects,” *Cell Stem Cell*, 7(5): 555–558.

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.

Jones B.F. (2011) “As Science Evolves, How Can Science Policy?” *Innovation Policy and the Economy*, 11: 103–131.

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.

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.

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.

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.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.

Wuchty S., B.F. Jones, and B. Uzzi (2007) "The increasing Dominance of Teams in Production of Knowledge," *Science*, 316(5827): 1036–1039.

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	Med	SD	p10	p90
Corporate Affiliation$_{it}$	76,984	0.13	0	0.34	0.0	1
% New MeSH Descriptor Terms$_{it}$	42,911	0.49	0.46	0.22	0.23	0.79
% New Terms including Qualifiers$_{it}$	42,911	0.61	0.60	0.19	0.36	0.86
% New Terms at 9-digit Level$_{it}$	42,911	0.37	0.32	0.23	0.12	0.69
% New Terms at 6-digit Level$_{it}$	42,911	0.27	0.21	0.23	0.05	0.58
ln(1+Cumulative Publications)$_{it-1}$	76,984	2.53	2.48	1.17	1.10	4.08
ln(1+Cumulative MeSH Terms Used)$_{it-1}$	76,984	4.47	4.62	1.16	3.14	5.75
ln(1+Cumulative Citations Received)$_{it-1}$	76,984	6.14	6.40	1.83	3.78	8.19
Age (Years Since First Publication)$_{it}$	76,984	14.1	13	7.55	4	25
Number of Publications$_{it}$	76,984	1.88	1	2.84	0	5
Number of Citations to Yearly Pubs$_{it}$	76,984	99.2	12	244.7	0	278
Median Pub CHI Journal Ranking$_{it}$	35,615	3.00	3	0.80	2	4
Post-Corporate Academic Affiliation$_{it}$	65,211	0.39	0	0.49	0	1
First Publication Year$_{i}$	76,984	1990.1	1991	6.53	1980	1998
State R&D Tax Credit Rate$_{it-1}$	67,531	0.06	0.05	0.05	0	0.15
State Predicted R&D Employment$_{it-1}$	67,531	9.58	9.89	1.39	7.37	11.49

Notes: Instrumental variables are only defined for scientists who remain in the United States throughout the sample. Post-corporate affiliation variable is only defined or scientists who are not employed in industry in the first year of the sample, then move to industry, and subsequently return to academia during the sample period or is equal to zero for scientists who remain employed exclusively in academia during the sample period.

Table 2: Changes in the rate at which scientists use new MeSH terms in research outputs when having a corporate affiliation

	% New MeSH Descriptor Terms		% New incl. Qualifiers	% New at 9-Digit Level	% New at 6-Digit Level
	(1)	(2)	(3)	(4)	(5)
Corporate Affiliation	0.021***	0.021***	0.017***	0.017***	0.014***
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
Controls	No	Yes	Yes	Yes	Yes
Age FE	No	Yes	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Scientists	5,312	5,312	5,312	5,312	5,312
Observations	42,911	42,911	42,911	42,911	42,911

Notes: All estimates are based on panel OLS regression with year and scientist fixed effects. Control variables are for lagged citations, publications, and cumulative terms used (in logs). Standard errors (in parentheses) are clustered at the scientist level. ***p<0.001, **p<0.01, *p<0.5, $^{+}$p<0.1.

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

	Number of Publications		Number of Citations		Median Publication CHI Journal Ranking	
	(1)	(2)	(3)	(4)	(5)	6)
Corporate Affiliation	-0.038	-0.006	0.063	0.069	-0.026	-0.025
	(0.036)	(0.035)	(0.072)	(0.073)	(0.021)	(0.021)
Controls	No	Yes	No	Yes	No	Yes
Age FE	No	Yes	No	Yes	No	Yes
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Scientists	5,312	5,312	5,312	5,312	5,134	5,134
Observations	76,984	76,984	76,984	76,984	35,615	35,615

Notes: Estimates in Models 1 to 4 are from Poisson fixed effects models. Estimates in Models 5 to 6 are from OLS models. All models include year and scientist fixed effects. Control variables are for lagged citations, publications, and cumulative terms used (in logs). Models 3 and 4 include control for a scientist's number of publications in a given year to ensure observed impact is not driven by volume of publications instead of publication quality. Not all articles are in journals covered by the CHI ranking, leading to fewer non-missing observations than observed publications at the scientist-year level. Standard errors (in parentheses) are clustered at the scientist level. ***p<0.001, **p<0.01, *p<0.5, $^{+}$p<0.1.

Table 4: Changes in scientists' research after returning to academia from industry (relative to prior academic employment)

	% New Descriptors	% New incl. Qualifiers	% New at 9-Digit Level	% New at 6-Digit Level
	(1)	(2)	(3)	(4)
Corporate Affiliation	0.027***	0.023***	0.023**	0.017**
	(0.008)	(0.007)	(0.007)	(0.006)
Post-Corporate Affiliation	0.005	0.000	0.005	0.003
	(0.007)	(0.006)	(0.006)	(0.005)
Controls	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Scientists	4,494	4,494	4,494	4,494
Observations	37,387	37,387	37,387	37,387

Notes: All estimates are based on panel OLS regression with year and scientist fixed effects. Control variables are for lagged citations, publications, and cumulative terms used (in logs). Scientists must not have been working in industry in the first year of the sample period and must return to academic employment before the last year of the sample period to be included or be employed only in academia during the sample period. Excluding never switchers does not meaningfully affect the results. Standard errors (in parentheses) are clustered at the scientist level. ***p<0.001, **p<0.01, *p<0.5, $^{+}$p<0.1.

Table 5: Changes in scientists' research when having a corporate affiliation (2SLS results)

	Corporate Affiliation	% New Descriptors	% New incl. Qualifiers	% New 9-Digit Level	% New 6-Digit Level
	(1)	(2)	(3)	(4)	(5)
Corporate Affiliation		0.167*	0.126*	0.146*	0.142*
		(0.067)	(0.059)	(0.068)	(0.065)
R&D Tax Credit Rate	0.233***				
	(0.055)				
Predicted Scientific Labor Demand	0.011***				
	(0.003)				
Controls	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Scientists	4,663	4,663	4,663	4,663	4,663
Observations	67,531	67,531	67,531	67,531	67,531
K.-P. F Statistic	17.11				
Hansen J Statistic		0.98	0.98	0.83	0.80

Notes: All estimates are based on panel OLS regression with year and scientist fixed effects. We include all observations for a scientist in the models (including for years they do not publish) since each year's affiliation observation is relevant for the effect of the instruments on scientists' affiliations in the first stage. We then include an indicator variable in the model to denote years in which the dependent variable is undefined and impute an arbitrary value for the dependent variable in these years. Additional models in the appendix show that the results are similar if these observations are not included. We also include a binary control variable denoting whether a scientist has moved state in the previous five years. This is intended to mitigate the risk that changes in values of the instruments are driven by changes in state location and these recent state transitions (suggesting transitions between employers) are associated with changes in scientists' research direction. Other control variables are for lagged citations, publications, and cumulative terms used (in logs). Standard errors (in parentheses) are clustered at the scientist level. Observations for scientists who leave the U.S. during the sample period are dropped from the analysis (as the instruments are U.S.-specific) which leaves fewer observations than in the core analysis. Standard errors (in parentheses) are clustered at the scientist level. Clustering standard errors by state leads to a larger F statistic and smaller p-values across models (see Online Appendix). ***$p<0.001$, **$p<0.01$, *$p<0.5$, $^{+}p<0.1$.

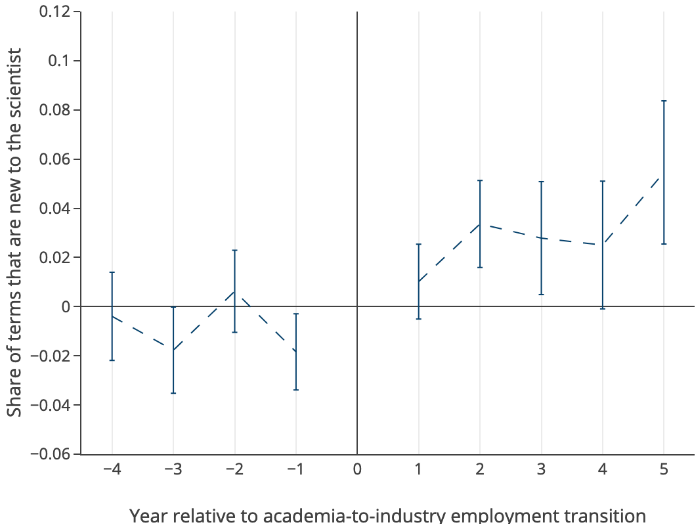
Table 6: Ex ante human capital characteristics of scientists moving to industry employment

	Acquisition of First Industry Affiliation			
	Count-based Independent Variables		Diversification Index-based Independent Variables	
	(1)	(2)	(3)	(4)
Techniques & Equipment	0.003*	0.003*	0.014**	0.010*
	(0.001)	(0.001)	(0.005)	(0.005)
Chemicals & Drugs	0.004***	0.004**	0.012^{+}	0.008
	(0.001)	(0.001)	(0.006)	(0.006)
Phenomena & Processes	0.002*	0.002	0.018***	0.011*
	(0.001)	(0.001)	(0.005)	(0.005)
Diseases	-0.001	-0.001	-0.000	-0.003
	(0.001)	(0.001)	(0.002)	(0.002)
Anatomy	-0.001	-0.001	-0.001	-0.003
	(0.001)	(0.001)	(0.004)	(0.004)
Other Categories	0.000	-0.000	-0.002	-0.005
	(0.001)	(0.002)	(0.007)	(0.006)
Controls	No	Yes	No	Yes
Age FE	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Scientists	4,901	4,901	4,901	4,901
Observations	62,916	62,916	62,916	62,916

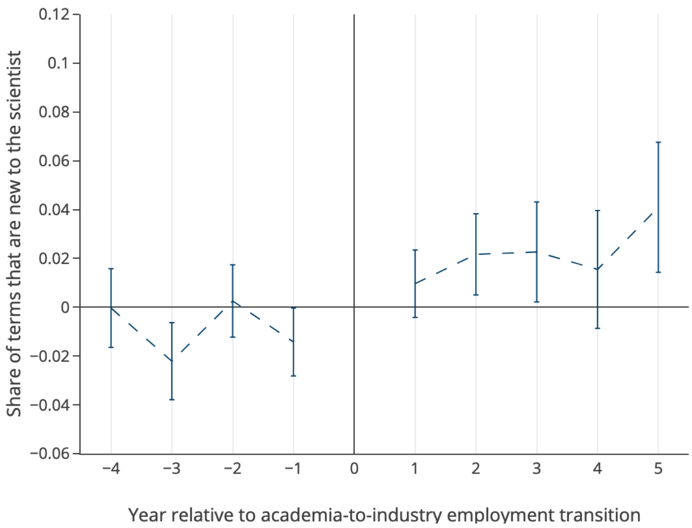
Notes: All models are discrete time proportional hazard models using complementary log-log regressions. Reported results are the average marginal effects with standard errors adjusted by the delta method. Models 1 and 2 use the count of terms in each category used by a scientist in their career strictly prior to the focal year as independent variables, taking the form: $\ln(1+\text{Category Terms}_{it-1})$. Models 3 and 4 use a diversification index to measure a scientist's diversification over terms in each MeSH category. This is defined as one minus the square root of the Herfindahl Index in each MeSH category. Term shares are measured as the number of times a scientist's research has a been indexed to a focal term divided by the number of terms in that category that have been indexed to the scientist's research. It is measured cumulatively for a scientist's career strictly prior to the focal year. In each model the dependent variable is a dummy variable equal to one if a scientist acquires an industry affiliation for the first time in year *t*. Scientists drop out of the sample after acquiring the affiliation. The mean probability a scientist in this sample acquires a first industry affiliation in a given year is 1.3 percentage points. Cohort fixed effects are defined by the first year in which a scientist had a recorded publication. Scientists are excluded if they have already transitioned to industry by the start of the sample. Control variables are for lagged citations and publications. Standard errors (in parentheses) are clustered at the scientist level. ***$p<0.001$, **$p<0.01$, *$p<0.5$, $^{+}p<0.1$.

Figure 1: Changes in the share of new terms used by scientists

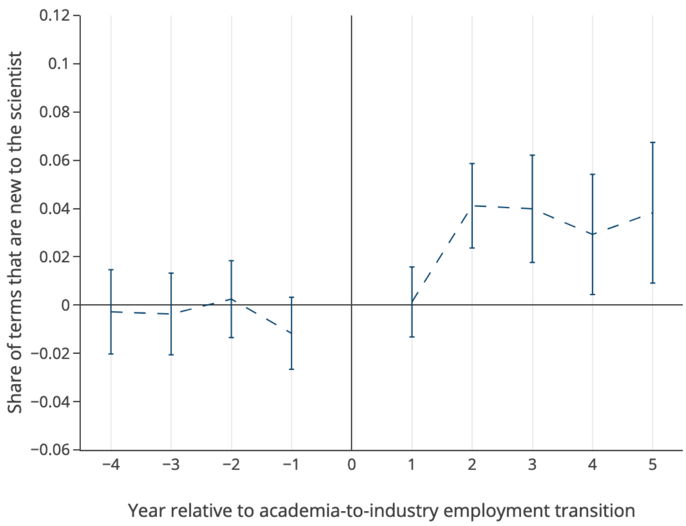
Panel A: Percent new MeSH descriptor terms



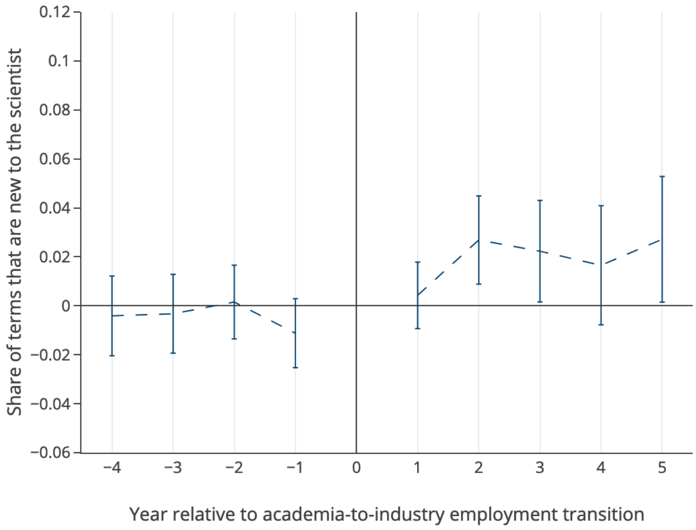
Panel B: Percent new MeSH terms incl. qualifiers



Panel C: Percent new terms at 9-digit MeSH level



Panel D: Percent new terms at 6-digit MeSH level



Notes: The figures represent results from the regression models with full set of controls with variables equivalent to Models 2, 3, 4, and 5 of Table 2. Year 0 is the year prior to first acquiring an industry affiliation. Scientists transitioning to industry are observed for up to five years before and five years after acquiring an industry affiliation. Scientists who never transition to industry are also included as controls. Those who transition out of industry employment within the five years of industry employment are subsequently excluded from the sample. Error bars represent 90% confidence intervals.

Online Appendix: Additional Tables & Figures

Table A.1: Additional summary statistics

Variable	Obs	Mean	Med	SD	p10	p90
ln(1+Cum Anatomy Terms)$_{it-1}$	62,916	2.35	2.48	1.06	0.69	3.61
ln(1+Cum Chem/Drugs Terms)$_{it-1}$	62,916	3.14	3.30	1.26	1.39	4.62
ln(1+Cum Diseases Terms)$_{it-1}$	62,916	1.95	1.95	1.31	0	3.76
ln(1+Cum Phenom/Process Terms)$_{it-1}$	62,916	2.64	2.83	1.17	1.10	4.03
ln(1+Cum Tech/Equipment Terms)$_{it-1}$	62,916	2.46	2.56	1.07	1.10	3.76
ln(1+Cum Other Categories Terms)$_{it-1}$	62,916	2.74	2.77	0.92	1.61	3.85
Div Index Anatomy Terms$_{it-1}$	62,916	0.56	0.65	0.23	0.	0.77
Div Index Chemicals/Drugs Terms$_{it-1}$	62,916	0.69	0.77	0.22	0.42	0.86
Div Index Diseases Terms$_{it-1}$	62,916	0.45	0.55	0.30	0	0.79
Div Index Phenom/Process Terms$_{it-1}$	62,916	0.62	0.71	0.24	0.25	0.81
Div Index Tech/Equipment Terms$_{it-1}$	62,916	0.59	0.67	0.22	0.29	0.78
Div Index Other Categories Terms$_{it-1}$	62,916	0.61	0.65	0.15	0.48	0.73
First Career Industry Transition Year$_{it}$	62,916	0.01	0	0.11	0	0

Notes: The variables measuring the range of scientists' knowledge and knowledge diversification in each MeSH category are only defined for observations until the year in which the scientist first transitions to industry employment (if ever). It remains defined for all years if the scientist never moves to industry employment. Scientist observations are dropped from regressions in years strictly after the year in which they had the first observed industry affiliation. The variable 'First Career Industry Transition Year' is an indicator variable equal to one for scientist i if they had their first observed career industry affiliation in year t (affiliations with industry prior to the start off the sample in 1996 would lead to a scientist being dropped from the sample). The indicator is set to zero if the scientist has not yet had (or never has) an observed industry affiliation.

Table A.2: Using total new terms as a dependent variable

	New MeSH Descriptor Terms		New incl. Qualifiers	New at 9-Digit Level	New at 6-Digit Level
	(1)	(2)	(3)	(4)	(5)
Corporate Affiliation	1.071***	1.068***	1.060***	1.070***	1.087***
	(0.017)	(0.016)	(0.014)	(0.019)	(0.023)
Controls	No	Yes	Yes	Yes	Yes
Age FE	No	Yes	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Scientists	5,312	5,312	5,312	5,312	5,312
Observations	42,911	42,911	42,911	42,911	42,911

Notes: All estimates are based on panel Poisson fixed effects regressions with year and scientist fixed effects. Incidence rate ratios are reported. Dependent variable is the count of new terms per publication in year t. Results are almost identical if count of new terms in year t is the dependent variable. Control variables are for lagged citations, publications, and cumulative terms used (in logs). Standard errors (in parentheses) are clustered at the scientist level. ***p<0.001, **p<0.01, *p<0.5, $^{+}$p<0.1.

Table A.3: New terms dependent variables defined relative to the prior five years

	% New MeSH Descriptor Terms		% New incl. Qualifiers	% New at 9-Digit Level	% New at 6-Digit Level
	(1)	(2)	(3)	(4)	(5)
Corporate Affiliation	0.028***	0.024***	0.017***	0.025***	0.025***
	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)
Controls	No	Yes	Yes	Yes	Yes
Age FE	No	Yes	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Scientists	5,312	5,312	5,312	5,312	5,312
Observations	42,911	42,911	42,911	42,911	42,911

Notes: All estimates are based on panel OLS regression with year and scientist fixed effects. Terms are considered new if they have not been used by a scientist in the prior five years. Control variables are for lagged citations, publications, and cumulative terms used (in logs). Standard errors (in parentheses) are clustered at the scientist level. ***p<0.001, **p<0.01, *p<0.5, $^{+}$p<0.1.

Table A.4: Limiting the sample to the pre-2001 period

	% New MeSH Descriptor Terms		% New incl. Qualifiers	% New at 9-Digit Level	% New at 6-Digit Level
	(1)	(2)	(3)	(4)	(5)
Corporate Affiliation	0.036***	0.034***	0.029***	0.023*	0.018*
	(0.010)	(0.009)	(0.008)	(0.009)	(0.008)
Controls	No	Yes	Yes	Yes	Yes
Age FE	No	Yes	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Scientists	5,268	5,268	5,268	5,268	5,268
Observations	19,586	19,586	19,586	19,586	19,586

Notes: All estimates are based on panel OLS regression with year and scientist fixed effects. Sample limited to observations in the 1996-2000 period (inclusive). Control variables are for lagged citations, publications, and cumulative terms used (in logs). There are fewer authors than in the full sample because some scientists who have a publication in Scopus between 1996 and 2000 have a publication that is not indexed in the PubMed database. This means that the dependent variables are undefined during this period. However, the 44 dropped scientists have a publication that is indexed in both Scopus and PubMed outside this interval. Standard errors (in parentheses) are clustered at the scientist level. ***$p<0.001$, **$p<0.01$, *$p<0.5$, $^{+}p<0.1$.

Table A.5: Replicating the results among non-U.S. scientists

	% New MeSH Descriptor Terms		% New incl. Qualifiers	% New at 9-Digit Level	% New at 6-Digit Level
	(1)	(2)	(3)	(4)	(5)
Corporate Affiliation	0.029*	0.042***	0.038***	0.040***	0.036***
	(0.013)	(0.011)	(0.011)	(0.011)	(0.009)
Controls	No	Yes	Yes	Yes	Yes
Age FE	No	Yes	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Scientists	2,184	2,184	2,184	2,184	2,184
Observations	17,308	17,308	17,308	17,308	17,308

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. Control variables are for lagged citations, publications, and cumulative terms used (in logs). Standard errors (in parentheses) are clustered at the scientist level. ***$p<0.001$, **$p<0.01$, *$p<0.5$, $^{+}p<0.1$.

Table A.6: Changes in new term use by scientists' characteristics (split sample analysis)

	% New MeSH Descriptor Terms		% New incl. Qualifiers		% New at 9-Digit Level		% New at 6-Digit Level	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Quality	*≤Med Cites*	*>Med Cites*	*≤Med Cites*	*>Med Cites*	*≤Med Cites*	*>Med Cites*	*≤Med Cites*	*>Med Cites*
Corporate Affiliation	0.031***	0.021**	0.021**	0.023**	0.027**	0.021**	0.024**	0.012*
	(0.009)	(0.008)	(0.008)	(0.008)	(0.009)	(0.007)	(0.008)	(0.006)
Scientists	1,782	1,765	1,782	1,765	1,782	1,765	1,782	1,765
Obs.	14,323	19,016	14,323	19,016	14,323	19,016	14,323	19,016
Panel B: Cohort	*Pre-1990*	*Post-1990*	*Pre-1990*	*Post-1990*	*Pre-1990*	*Post-1990*	*Pre-1990*	*Post-1990*
Corporate Affiliation	0.020**	0.024***	0.017**	0.019**	0.017**	0.018**	0.016**	0.014*
	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.007)	(0.005)	(0.006)
Scientists	2,163	3,149	2,163	3,149	2,163	3,149	2,163	3,149
Obs.	22,274	20,637	22,274	20,637	22,274	20,637	22,274	20,637
Panel C: Age	*<12 years*	*≥12 years*	*<12 years*	*≥12 years*	*<12 years*	*≥12 years*	*<12 years*	*≥12 years*
Corporate Affiliation	0.030***	0.017**	0.024**	0.015*	0.022*	0.016**	0.018*	0.014**
	(0.008)	(0.006)	(0.007)	(0.006)	(0.008)	(0.005)	(0.008)	(0.005)
Scientists	3,894	3,722	3,894	3,722	3,894	3,722	3,894	3,722
Obs.	15,790	27,121	15,790	27,121	15,790	27,121	15,790	27,121
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: All estimates are based on panel OLS regression with year and scientist fixed effects. In Panel A, 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 in Scopus. Scientists with a corporate affiliation and scientists for who only appear in Scopus from 1996 onwards are excluded. In Panel B, the sample is split according to whether a scientist has a first publication recorded in Scopus strictly before 1990. In Panel C, 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. Standard errors (in parentheses) are clustered at the scientist level. ***p<0.001, **p<0.01, *p<0.5, $^{+}$p<0.1.

Table A.7: Controlling for number of co-authors and publications

	% New MeSH Descriptor Terms		% New incl. Qualifiers		% New at 9-Digit Level		% New at 6-Digit Level	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Co-authors								
Corporate Affiliation	0.018***	0.023***	0.014**	0.018**	0.013**	0.018**	0.011**	0.016**
	(0.005)	(0.006)	(0.004)	(0.006)	(0.004)	(0.006)	(0.004)	(0.005)
Corporate Affiliation		-0.000		-0.000		-0.000		-0.000
*Number of Co-authors		(0.000)		(0.000)		(0.000)		(0.000)
Panel B: Publications								
Corporate Affiliation	0.020***	0.033***	0.016***	0.028***	0.016***	0.025***	0.013***	0.021***
	(0.005)	(0.006)	(0.004)	(0.006)	(0.004)	(0.006)	(0.004)	(0.005)
Corporate Affiliation		-0.003**		-0.003**		-0.002**		-0.002**
*Number of Publications		(0.001)		(0.001)		(0.001)		(0.001)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientists	5,312	5,312	5,312	5,312	5,312	5,312	5,312	5,312
Observations	42,911	42,911	42,911	42,911	42,911	42,911	42,911	42,911

Notes: All estimates are based on panel OLS regression with year and scientist fixed effects. All models in Panel A control for the number of co-authors on a scientist's publications in a given year. All models in Panel B control for a scientist's number of publications in a given year. Control variables are for lagged citations, publications, and cumulative terms used (in logs). Standard errors (in parentheses) are clustered at the scientist level. ***$p<0.001$, **$p<0.01$, *$p<0.5$, $^{+}p<0.1$.

Table A.8: 2SLS results using alternative standard errors with clusters defined at the state (not scientist) level

	Corporate Affiliation	% New Descriptors	% New incl. Qualifiers	% New 9-Digit Level	% New 6-Digit Level
	(1)	(2)	(3)	(4)	(5)
Corporate Affiliation		0.167**	0.126*	0.146*	0.142*
		(0.058)	(0.055)	(0.061)	(0.064)
R&D Tax Credit Rate	0.233***				
	(0.053)				
Predicted Scientific Labor Demand	0.011***				
	(0.003)				
Controls	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Scientists	4,663	4,663	4,663	4,663	4,663
Observations	67,531	67,531	67,531	67,531	67,531
K.-P. F Statistic	20.26				
Hansen J Statistic		0.98	0.98	0.86	0.85

Notes: Models are the same as in Table 5 of the main paper, but with standard errors (in parentheses) clustered at the state (not scientist) level. ***p<0.001, **p<0.01, *p<0.5, $^{+}$p<0.1.

Table A.9: 2SLS results using only the R&D tax credit instrument

	Corporate Affiliation	% New Descriptors	% New incl. Qualifiers	% New 9-Digit Level	% New 6-Digit Level
	(1)	(2)	(3)	(4)	(5)
Corporate Affiliation		0.166*	0.125^{+}	0.135^{+}	0.129^{+}
		(0.081)	(0.072)	(0.081)	(0.076)
R&D Tax Credit Rate	0.278***				
	(0.055)				
Controls	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Scientists	4,663	4,663	4,663	4,663	4,663
Observations	67,531	67,531	67,531	67,531	67,531
K.-P. F Statistic	25.32				

Notes: Models are the same as in Table 5 of the main paper, but with only the R&D tax credit variable used as an instrument. Standard errors (in parentheses) are clustered at the scientist level. ***p<0.001, **p<0.01, *p<0.5, $^{+}$p<0.1.

Table A.10: 2SLS results using only the predicted scientific labor demand instrument

	Corporate Affiliation	% New Descriptors	% New incl. Qualifiers	% New 9-Digit Level	% New 6-Digit Level
	(1)	(2)	(3)	(4)	(5)
Corporate Affiliation		0.168^{+}	0.127^{+}	0.157^{+}	0.154^{+}
		(0.087)	(0.075)	(0.088)	(0.086)
Predicted Scientific Labor Demand	0.013***				
	(0.003)				
Controls	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Scientists	4,663	4,663	4,663	4,663	4,663
Observations	67,531	67,531	67,531	67,531	67,531
K.-P. F Statistic	18.75				

Notes: Models are the same as in Table 5 of the main paper, but with only the predicted scientific labor demand variable used as an instrument. Standard errors (in parentheses) are clustered at the scientist level. ***p<0.001, **p<0.01, *p<0.5, $^{+}$p<0.1.

Table A.11: 2SLS results only including observations with defined publication information

	Corporate Affiliation	% New Descriptors	% New incl. Qualifiers	% New 9-Digit Level	% New 6-Digit Level
	(1)	(2)	(3)	(4)	(5)
Corporate Affiliation		0.207*	0.154^{+}	0.171^{+}	0.158^{+}
		(0.100)	(0.091)	(0.095)	(0.086)
R&D Tax Credit Rate	0.171**				
	(0.060)				
Predicted Scientific Labor Demand	0.011***				
	(0.003)				
Controls	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Scientists	4,663	4,663	4,663	4,663	4,663
Observations	36,079	36,079	36,079	36,079	36,079
K.-P. F Statistic	12.27				
Hansen J Statistic		0.91	0.80	0.56	0.36

Notes: Models are the same as in Table 5 of the main paper, but only including observations where the dependent variable is defined (as per Section 3). Standard errors (in parentheses) clustered at the scientist level. Clustering at the state-level provides less conservative estimates with a larger F Statistic and smaller p-values. ***p<0.001, **p<0.01, *p<0.5, $^{+}$p<0.1.

Table A.12: Ex ante human capital characteristics of scientists transitioning to industry employment (marginal effects at means)

	Acquisition of First Industry Affiliation			
	Count-based Variables		Diversification Index Variables	
	(1)	(2)	(3)	(4)
Techniques & Equipment	0.002*	0.002*	0.009*	0.007*
	(0.001)	(0.001)	(0.003)	(0.003)
Chemicals & Drugs	0.003***	0.003**	0.008^{+}	0.005
	(0.001)	(0.001)	(0.004)	(0.004)
Phenomena & Processes	0.002*	0.001	0.012***	0.008*
	(0.001)	(0.001)	(0.003)	(0.003)
Diseases	-0.001	-0.001	-0.000	-0.002
	(0.000)	(0.000)	(0.002)	(0.002)
Anatomy	-0.001	-0.001	-0.000	-0.002
	(0.001)	(0.001)	(0.003)	(0.003)
Other Categories	0.000	-0.000	-0.001	-0.003
	(0.001)	(0.001)	(0.005)	(0.004)
Controls	No	Yes	No	Yes
Age FE	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Scientists	4,901	4,901	4,901	4,901
Observations	62,916	62,916	62,916	62,916

Notes: All models are discrete time proportional hazard models using complementary log-log regressions. Reported results are the marginal effects at means with standard errors adjusted by the delta method. Models 1 and 2 use the count of terms in each category used by a scientist in their career strictly prior to the focal year as independent variables, taking the form: ln(1+Category $\text{Terms}_{\text{it-1}}$). Models 3 and 4 use a diversification index to measure a scientist's diversification over terms in each MeSH category. This is defined as one minus the square root of the Herfindahl Index in each MeSH category. Term shares are measured as the number of times a scientist's research has a been indexed to a focal term divided by the number of terms in that category that have been indexed to the scientist's research. It is measured cumulatively for a scientist's career strictly prior to the focal year. In each model the dependent variable is a dummy variable equal to one if a scientist acquires an industry affiliation for the first time in year *t*. Scientists drop out of the sample after acquiring the affiliation. The mean probability a scientist in this sample acquires a first industry affiliation in a given year is 1.3 percentage points. Cohort fixed effects are defined by the first year in which a scientist had a recorded publication. Scientists are excluded if they have already transitioned to industry by the start of the sample. Control variables are for lagged citations and publications. Standard errors (in parentheses) are clustered at the scientist level. ***p<0.001, **p<0.01, *p<0.5, $^{+}$p<0.1.

More information on the MeSH categories in the sample

To examine where exploration increases, we separate the MeSH descriptor terms in the sample according to their highest-level thematic categorization. These represent first-level of branching in the MeSH hierarchical tree. We then recreate the core dependent variable to count the share of terms that are new to a scientist and in a focal category. Five of the sixteen categories account for approximately 10 percent or more of the terms indexed to the papers in the sample and collectively account for approximately 75 percent of all terms. These are 'Anatomy,' 'Processes & Phenomena,' 'Chemicals & Drugs,' 'Diseases,' and 'Analytical, Diagnostic, and Therapeutic Techniques & Equipment.'

The Anatomy category classifies the aspect of a body with which a paper is concerned. In our data, many MeSH terms in this category often refer to the specific types of cells that are the focus of the research. For example, the most common terms in the data are "cells, cultured," "cell line," "tumor cells, cultured," "hematopoietic stem cells," "T-lymphocytes," "stem cells," "cell line, tumor," "neurons," "bone marrow cells," and "liver.

The Processes & Phenomena category comprises a range of physiological processes that take place at the level of a gene, cell, organ, or system. In our data, the most common terms are "cell differentiation," "signal transduction," "base sequence," "mutation," "apoptosis," "gene expression regulation," "cell division," "amino acid sequence," "transfection," and "pregnancy."

The Diseases category includes conditions affecting health. In this sample, these are most frequently types of cancer. The most common in the data are "breast neoplasms," "graft vs host disease," "neoplasms," "recurrence," "disease progression," "prostatic neoplasms," "lung neoplasms," "multiple myeloma," "acute disease," and "leukemia, myeloid, acute."

The Chemicals & Drugs category contains the chemicals that can be used as interventional agents to affect the biological processes in an organism. The most common in the data are "RNA, messenger," "antibodies, monoclonal," "DNA-binding proteins," "recombinant proteins," "antineoplastic agents," "transcription factors," "Antigens, CD," 'proto-oncogene proteins," "membrane proteins," and "cytokines."

The Analytical, Diagnostic, and Therapeutic Techniques & Equipment category includes a range of terms that represent different tools for diagnosing or treating health conditions. In our data, the most frequent terms include: "antineoplastic combined chemotherapy protocols," "hematopoietic stem cell transplantation," "bone marrow transplantation," "flow cytometry," "polymerase chain reaction," "transplantation, homologous," "lymphocyte activation," "cloning, molecular," "blotting, western," and "disease models, animals."

The Organisms category is excluded as a primary category of analysis despite containing approximately 10 percent of MeSH descriptor terms in the sample. This is because it has three very common terms "Animals," "Humans," and "Mice" that collectively account for more than 65 percent of observations in that category. These three terms are respectively between 5 and 10 times more common than the next most frequently occurring term in the other five categories that comprise approximately 10 percent or more of the sample each.

There is a wide range of terms within each MeSH category (both overall and in our sample). Further research is needed to establish a more precise mapping between given terms and skills, and how this may vary across research fields. The empirical setting in this paper is the stem cell field. In other settings, the most valuable skills may be linked to concepts in other MesH categories. Thus, we caution against wider generalizations based on the specific MeSH terms from this setting.