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

August 9, 2025

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 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 respective goals of academic and corporate science to document how this distinction sheds light on scientists' research outcomes after moving to industry employment. There is a pronounced within-scientist shift in research strategies after transitions to industry: Scientists persistently explore more new-to-the-scientist concepts in their projects while simultaneously re-using known research skills at an increased rate. We also provide evidence that industry employers select scientists with greater and more diverse experience using a wider range of research tools and techniques. Our findings are consistent with industry prioritizing the ability to apply the research skills of scientists hired from academia across a wider range of projects in their field.

1. Introduction

Scientists with research training and experience in academia have human capital that is specialized on multiple dimensions. On the one hand, academic scientists have deep knowledge in narrow niches in which they seek to advance the frontier of knowledge in their field (Jones, 2009; Leahey, Beckman & Stanko, 2017). By hiring or collaborating with these scientists, corporate R&D programs can benefit from access to ideas at the forefront of knowledge in academic science (Zucker, Darby & Armstrong, 2002; Fabrizio, 2009). On the other hand, scientists in academia also develop extensive advanced skills—often rooted in tacit knowledge—to design, implement, and evaluate experimental research projects (Senker, 1995; Mokyr, 2005). While these research skills represent a type of knowledge, they have the unique characteristic that they provide the tools that enable scientists to test the properties of, and relationships between, objects of investigation in experiments. They are distinct from other types of scientific knowledge insofar as they provide the means through which scientists can manipulate objects (about which they may have additional knowledge) when carrying out research projects. A scientist with specialized expertise on a narrow range of scientific topics may, nonetheless, use a wide range of different research skills to analyze these topics. Conversely, a different scientist may use the same set of core skills to carry out projects on a diverse range of scientific topics.

Our interest in this paper is to examine whether the extent of scientists' research skills may be an important antecedent of employment transitions from academia to industry and influence how scientists' human capital is applied in industry research. We highlight the role of research skills as a discrete component of scientists' human capital that is relevant to understanding firms' scientific human capital strategies and scientists' career outcomes. The existing literature analyzing how scientific human capital contributes to industry research has two primary focuses. First, it considers whether scientists are 'generalists' or 'specialists', defined by the overall diversity of their existing knowledge across scientific topics (Jones, 2009; Agrawal, Goldfarb & Teodoridis, 2016; Teodoridis, Vakili & Bikard, 2019; Nagle & Teodoridis, 2020). Second, it analyzes how scientists' relative prior experience in basic or applied research provides more industry-relevant human capital (Toole & Czarnitzki, 2009; Baba, Shichijo & Sedita, 2009; Subramanian, Lim & Soh, 2013). This paper seeks to add to this line of research by showing how a distinction between knowledge and skills sheds light on the drivers of career transitions between academia and industry and scientists' subsequent research outcomes.

Consider a scientist carrying out research in the broad field of endocrinology, for example. Her research may cover a set of scientific objects in this field that may include: a disease (e.g., diabetes or hyperthyroidism); the role of hormones in physiological processes (e.g., insulin secretion or heart rate regulation); the impact of hormones on a particular organ (e.g., the brain or pancreas); and particular chemicals and drugs that influence hormone synthesis and metabolism in the body (e.g., Glucagon-Like Peptide 1 or metformin). She may use a variety of research skills that enable her to manipulate scientific objects in her research projects. She may cultivate a range of different types of cells, use myriad techniques to analyze different types of tissues, carry out genetic analysis or editing, use computational bioinformatics, use advanced imaging techniques, analyze epidemiological data, or design trials, among the range of potential research skills used in this field. The scientist with a larger set of these research skills can carry out a wider range of research tasks within her broader area of expertise than a scientist with fewer skills. Other things being equal, the scientist who can carry out more research tasks can use her skills to contribute to a larger set of possible research projects.

While extensive research has shown that engagement with academic ideas is central to building effective absorptive capacity (Fabrizio, 2009), the wider literature also identifies a divergence in the respective goals of scientific research between academia and industry (Evans, 2010; Sauermann & Stephan, 2013). The reward system of academic science prioritizes major discoveries about fundamental phenomena (Merton, 1957; Latour, 1987; Dasgupta & David, 1994). On the other hand, firms' returns are based on understanding the relationships between scientific or technological objects that can subsequently be leveraged for commercial purposes (Evans, 2010; Sauermann & Stephan, 2013). As the knowledge frontier advances, academic scientists are incentivized to develop deeper human capital in narrower research niches where they can make major contributions at the forefront of their field (Jones, 2009; Agrawal et al., 2016). However, firms benefit from the ability to experiment rapidly with new ideas in a field that have come from beyond their boundaries (Cohen & Levinthal, 1989, 1990; Arora & Gambardella, 1994). The primary goal of corporate research programs is to identify commercially valuable relationships between scientific objects that can be exploited ahead of rivals (Evans, 2010; Sauermann & Stephan, 2013). This leads corporate science to select research projects that are more exploratory of areas that are distant from existing hubs of knowledge in a scientific field, rather than diving deeply into those topics that are at the center

of academic research in that field (Evans, 2010). Consequently, industry employers hiring academic scientists may place relatively higher value on those components of a scientist's human capital that enable them to carry out a wider, more diverse range of research projects in their field (compared to the returns to these components of human capital for scientists in academia where the relative returns to depth of topic knowledge may be greater).

We develop several testable implications that build on this line of argument. First, scientists transitioning from academic to industry employment will shift research strategies to persistently cover new-to-the scientist concepts at an increased rate. Second, that after moving to industry, scientists' research will involve previously used research skills at an increased rate but that they will otherwise be less likely to re-cover the same concepts compared to scientists remaining in academia. Third, that the academic scientists who transition to industry employment will have demonstrated a greater range of research skills in their past projects prior to being hired by an industry employer.

We test these predictions by examining scientists in the stem cell field of regenerative medicine as major scientific breakthroughs occurred in the 1990s and 2000s. Stem cell research projects were motivated by the potential uses of stem cells to address human disease and required basic scientific research into their scientific properties to develop knowledge of the opportunities opening up. These advances—promising both high commercial and high scientific value—led both industry and academia to pay significant attention to the stem cell field. In turn, leading to scientific collaboration and mobility across the different types of employers (Blomfield & Vakili, 2023). We create a panel dataset comprising the career histories of all U.S.-based scientists working in the stem cell field just during the period in which these major scientific advances took place. We link this to rich bibliographic data on scientists' publications in the Scopus and PubMed databases. We use this dataset to examine how the characteristics of scientists' prior knowledge are linked to their propensity to acquire industry affiliations and the nature of changes in the concepts used in scientists' research after employment transitions.

We find strong evidence in support of our predictions. Academic scientists who move to industry employment increase their rate of covering new-to-the-scientist topics while also re-using existing skills at a higher rate. We also find that academic scientists hired by industry have used a larger and more diverse range of prior research skills than scientists staying in academia. To bolster confidence in the causality of our results, we examine a wide range of

alternative explanations and use two instrumental variable strategies to provide plausibly exogenous shifts local industry demand for scientific human capital in the stem cell field. Our results advance the literature on the drivers of scientific labor mobility and firms' scientific human capital strategies by highlighting the role of research skills as an important, distinct component of scientists' human capital. Our findings have wider implications for the literature on how the relative balance of scientific labor between industry and academia affects the accumulation of knowledge in a scientific field.

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; Latour, 1987; 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). One line of research has focused on differences in scientists' human capital according to the extent to which they are generalists or specialists based on the overall breadth or diversity of their past research (Agrawal et al., 2016; Leahey et al., 2017; Teodoridis, 2018; Teodoridis et al., 2019). A second set of studies have examined 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 et al., 2009; Subramanian et al., 2013).

However, academic scientists also develop the specialized research skills needed to carry out advanced research on their chosen research topics. 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, scientists' skills to cultivate and manipulate different types of cells or to use gene editing techniques can be applied to research projects that cover a wide range of potential diseases, organs, or physiological processes. The results of these experiments may then also be evaluated using a range of analytical techniques depending on the needs of a project. A scientist whose research focuses on relatively few topics may still use a wide range of different research skills to analyze

phenomena in their area of interest. Conversely, another scientist may use a narrow set of research skills but apply these in projects covering a wider range of research topics. This may lead to differences between the scientists in the breadth of their human capital both with respect to specific topics and the research skills they utilize in research projects.

Similar patterns of specialization can be seen in the social sciences. 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, or network analysis, 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 the observed topics or skills 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 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 skill set that could be applied across a wider range of research topics, should she desire to do this in her future research. In Scenario (3), the scientist has wide-ranging topic expertise, but only a narrow skill set that can be used on research projects within these topics. The set of future projects in which the scientist in Scenario (3) can utilize her topic knowledge is either 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, quasi-experimental econometric methods, randomized control trials, agent-based modelling, or social network analysis could be seen as topics within the social sciences disciplines. However, this type of knowledge has the distinct character of being applicable as a tool that underpins research projects across a wide range of other topics (Senker, 1995).

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 basic research outputs being difficult for a firm 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 internal scientific capabilities have superior capacity to absorb and apply potential valuable knowledge from beyond the firm's boundaries in academic science (Arora & Gambardella, 1994; Cockburn & Henderson, 1998; Gittelman & Kogut, 2003) and acquire insights into new knowledge through engagement with specialists in academia (Zucker, Darby & Armstrong, 2002; Fabrizio, 2009). 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, for example, through publishing, co-authoring, and attending scientific conferences (Arora & Gambardella, 1994; Liebeskind et al., 1996; Cockburn & Henderson, 1998; Simeth & Raffo, 2013). Scientific research can also be used to develop valuable new ideas internally that can be applied in firms' products (Arora et al., 2017). This organization-level capability is underpinned by firms' scientific human capital resources: the scientists who (at the micro-level) have the knowledge and skills to evaluate scientific ideas and carry out research within the firm.

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 the relationships between scientific or technological objects that can subsequently be leveraged for commercial purposes (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 that may have commercial value may 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, rather than focusing more narrowly on deep dives into specific research topics that are central to academic research in that field. Diverse experimentation across novel areas of science allows firms to uncover potentially valuable relationships before their rivals can do so (Evans, 2010). In a similar vein, Cohen et al. (2002) note that, in many cases, R&D managers highlight research techniques learned from academic science that can be used in their own research programs as being equally—or more important—to their firms as specific research findings articulated by academic scientists. In this case, significant value in corporate R&D comes from internalizing the technical skill of how to apply some technique relevant to the firm, which can be used in experiments in the firm's internal research projects.

Increasingly, larger team collaborations are necessary to provide sufficient breadth of knowledge and skills in scientific research projects (Wuchty, Jones & Uzzi, 2007; Jones, 2011). Nonetheless, coordination costs increase as a greater number of individuals with more specialized human capital are required to execute a project (Becker & Murphy, 1992; Bikard, Murray & Gans, 2015). Employing scientists with the research skills to carry out a more diverse range of projects at the individual level may be a more efficient way of covering the necessary range skills at the organization level compared to relying on a greater number of scientists with narrower skill sets. This would particularly be the case if, as Evans (2010) finds, firms are less interested in developing deep knowledge in a few specific topics through repeated research projects than in exploring a more diverse set of topics in a field. Thus, scientists with wide-ranging research skill sets that enable them to rapidly test a diverse range of potential relationships between objects may be especially attractive to firms. By contrast, it may be less attractive to employ scientists who have experience using fewer of the skills, tools, and

techniques that underpin the range of research projects in their field. Firms may also contract or build relationships between their internal scientists and external researchers who have the relevant highly specialized knowledge to access specific novel ideas as they develop in academic science, rather than having to hire the holders of this knowledge in order to access it (Liebeskind et al., 1996; Cockburn & Henderson, 1998; Zucker, Darby & Armstrong, 2002; Fabrizio, 2009).

2.3. Transitions between Academic and Industry Employment

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, 2015; Sauermann & Roach, 2014) or used survey data to examine scientists' lifecycle earnings in either industry or academic employment (Agarwal & Ohyama, 2013, Ding, Ohyama & Agarwal, 2021). 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; Bercovitz & Feldman, 2008; 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).

An additional literature focuses on how organizations can benefit from the human capital of "star" scientists (Zucker & Darby, 1996; Zucker, Darby & Torero, 2002; Rothaermel & Hess, 2007; Tzabbar, 2009; Kehoe & Tzabbar, 2015; Agrawal, McHale & Oettl, 2017). 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 over the course of their careers. This is particularly pertinent as increasing collaborative activities between industry and academic scientists, changing academic norms, restrictions on grants, and shifts in relative availability of research funds may induce greater mobility between academia and industry (Bozeman & Gaughan, 2007; Vallas & Kleinman, 2008; Roach & Sauermann, 2017).

One notable exception is Kaiser et al. (2018), who show that firms hiring scientists with experience working at universities experience increases in patenting at the firm-level. Similarly, Cassiman et al. (2018) show that hiring scientists with experience at research institutes designed to bridge corporate and academic research can increase patenting and complements a firm's participation in these institutes. However, neither of these studies examine how the new hires' human capital is utilized by firms in corporate research at an individual-level, nor how different aspects of human capital are linked to hiring decisions beyond basic-applied experience. Our goal in this paper is to examine how firms' human capital strategies lead them to hire for and exploit different components of scientists' human capital. We do this by focusing on individual scientist-level outcomes to understand, at a more micro-level, how firms' scientific human capital strategies underpin firm-level outcomes.

2.4. Scientists' Human Capital and Research Direction

In many fields of science, research can be both inspired by potential applications and involve research into fundamental scientific properties of objects (Stokes, 1997; Bikard, 2015). In "Pasteur's Quadrant," where this use-inspired basic research takes place, there can be substantial overlap in the scientific value of contributions to knowledge and the commercial potential of research findings. Nonetheless, scientists in academia and industry may pursue 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). The relevant 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 the existing hubs of knowledge in which academic scientists have focussed more intensively (Evans, 2010). Moreover, coordination and employment costs increase with the number of highly specialized workers needed for a set of projects (Becker & Murphy, 1992; Bikard et al., 2015; Agrawal et al., 2016). This may make it more costly for firms to coordinate research activities across large numbers of specialists with the necessary human capital to carry out a diverse range of projects. Thus, it may 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 fields.

This discussion leads us to three predictions that form the basis of the research questions that we examine in the remainder of the paper:

(1) Scientists will apply their human capital across research projects involving a new-to-the-scientist concepts at a higher rate after transitioning to industry employment (compared to if they had remained in academia)

(2) This increase in the rate of conceptual exploration will be accompanied by repeated use of previously developed research skills but less repeated use of other types of knowledge

(3) Firms will be more likely to hire scientists whose past research projects have utilized a wider range of research skills

Note that Prediction (1) does not imply that scientists will simply change research direction when they transition to industry employment. It suggests that scientists' research strategies will change. Rather than making a one-off shift in research direction which involves covering more new-to-the-scientist concepts, they will instead have more exploratory research strategies in which they cover new-to-the-scientist concepts at a higher rate on an ongoing basis. Although we do not make this one of the formal research questions of the paper, we would also expect that the patterns in Predictions (1) and (2) might be mitigated in the case of star scientists. First, these scientists may be more likely to have been hired precisely because of the value of their existing knowledge linked to their star status from their prior research (Zucker & Darby, 1996; Zucker, Darby & Torero, 2002; Tzabbar, 2009). Second, star scientists may be accorded greater research autonomy by firms because of their cutting-edge human capital in their existing research area and wider value to the firm for external collaborations (Liu & Stuart, 2010; Stephan, 2012). This greater autonomy—by making the industry research environment

more similar to academia—would be expected lower the effect of industry employment on scientists' research strategies.

The remainder of the paper proceeds as follows. Section 3 provides institutional details and describes our empirical approach. Section 4 presents detailed results. Finally, we conclude with a discussion of our results' implication for our understanding of firms' scientific human capital strategies, scientific careers and mobility, and the wider implications of the results for the division of scientific labor between academia and industry.

3. Research Design

3.1. Institutional Details

Our empirical setting for examining 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 rare 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 wide range of new treatments by naming it the "Breakthrough of the Year" (Vogel, 1999). Both academia and industry paid significant attention to the new opportunities for medical treatments opened by advances in stem cell science. It represented an archetype of Pasteur's Quadrant with research questions that involved basic scientific research and inspired by the potential uses for stem cells to underpin new drugs and therapeutic techniques. Large volumes of research were carried out by scientists in both academia and industry, with significant scope for mobility between each setting (Blomfield & Vakili, 2023).

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 the drivers of academia-to-industry career transitions, how mobility to industry affects 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 skill set to a more diverse range of research projects after academia-to-industry career transitions and whether the breadth of this skill set is associated with scientists being hired by industry employers.

3.2. Data

We use the Scopus database of scientific publications to identify the set scientists who had a publication in a stem cell field and were based in the United States during the period from 1996 to 2000, just before and during the period in which the initial breakthroughs were made in human embryonic stem cell science that led to increased scientific and corporate attention to this field (Blomfield & Vakili, 2023). 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 then follow these scientists' careers until 2010. This means that we have a minimum of ten years of observations per scientist. 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 in the sample period.

We augment this data with the keywords associated with scientists' articles in the PubMed database, which contains the vast majority of journals published in the life sciences. PubMed keywords are precisely indexed 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, and the number of terms used for the first time in each year to examine directional changes in scientists' research.

3.3. Main Variables

Our first set of analyses examine how the conceptual content of scientists' research changes when they transition from academic to industry employment. The main dependent variable for this analysis is: *% New MeSH Terms*. This 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 scientific concepts—as defined by the MeSH classification system—that are covered in a scientist's research in a given year that are novel with respect to the scientist's revealed antecedent knowledge set (defined by the previously indexed 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. If a scientist has persistently higher values of this variable after moving to industry employment, it would indicate that they have made a shift in their research strategies to explore new-to-the-scientist concepts at a consistently higher rate. Following other studies that analyze changes in scientists' research based on bibliographic data (e.g., Azoulay et al., 2009; Evans, 2010), we leave observations undefined in the panel when a scientist has no publications in a given year. 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. Further information on the construction of these variables and the results of robustness tests are presented in the Online Appendix.

The core independent variable for the first set of analyses 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 new MeSH terms indexed 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. Controlling for the stock of prior terms is intended to mitigate the risk that the results are driven by a mechanical correlation in which more experienced or productive scientists have less scope to expand their research to additional concepts in their field 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. 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 'career age' of the scientist. This is calculated as the difference in years between a focal year t and the year of i's first recorded publication. This fixed effect ensures 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.

If firms place a high value on the extent of scientists' research skills when hiring from academia, there should be evidence of this in the types of research that scientists have carried out before they move to industry. To examine this question, we use the dependent variable *First Corporate Affiliation*. This is an indicator variable equal to one in the year in which scientist i has a first recorded corporate affiliation in the dataset and equal to zero in all years strictly prior to this event. We create an additional set of independent variables to capture the different components of scientists' human capital by separating the MeSH descriptor terms indexed to scientists' papers according to their highest-level thematic categorization (which represents the first branch in the MeSH tree hierarchy).

Five of the sixteen MeSH categories each account for approximately 10 percent or more of the terms indexed to the papers in the sample. These are "Anatomy," "Diseases," "Processes & Phenomena," "Chemicals & Drugs," and "Analytical, Diagnostic, and Therapeutic Techniques & Equipment." The categories collectively account for approximately 75 percent of all terms indexed to papers in the sample. 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.

To illustrate how the MeSH category classification operates, we can consider recent advances in gene editing research based on CRIPSR systems, using 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 subject of gene editing techniques would be classified in Diseases. Finally, the pancreas's function, as an organ targeted by a research project, would be classed in Anatomy. Further information on the most frequently occurring MeSH terms in each category is provided in the Online Appendix. Further information about the most common MeSH terms in the sample is provided in the Online Appendix.

We create independent variables that capture scientists' antecedent knowledge in each of the major components in the sample. First, we measure breadth of knowledge in each category k using the natural logarithm of one plus the number of terms indexed to scientist i's articles in category k between t-1 and the year in which i has her first recorded publication. Second, we measure the diversity of knowledge in each category using a diversification index defined as one minus the square root of the Herfindahl Index of the terms previously indexed to scientist i's articles in that category. We focus this analysis on scientists' first corporate affiliation to ensure that scientists' research during prior industry employment do not influence the contents of scientists' knowledge sets prior to observed industry transitions. Were we to examine all transition events beyond the first, we may conflate scientists' propensity to work in industry with differences in the antecedent knowledge sets of scientists moving to industry (as observed knowledge sets would be influenced by research undertaken during prior industry employment).

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. The Analytical, Diagnostic, and Therapeutic 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 take while carrying out research projects or the tools they use to analyze, diagnose, or otherwise manipulate these objects. If industry places particularly high value 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. In this case, we would expect to find that scientists with broader and more diverse antecedent knowledge sets in this category have greater likelihood of being hired by industry.

Table 1 below provides summary statistics of the main variables used 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]

3.4. Econometric Approach

We first 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.

Second, 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. Specifically, 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 corporate affiliation in that year (and zero otherwise). Our key independent variables in these models capture the breadth and diversity of scientists' antecedent knowledge sets in each of the main MeSH categories. Scientists are dropped from the estimation sample following the first year (if any) in which they have a corporate affiliation. We include an additional cohort fixed effect in these models (defined as the year in which scientist i's first publication appears in Scopus), since the individual fixed effects are inappropriate in this model. 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 can only observe differences in the human capital of scientists who do and do not transition to industry employment over time. Thus, we make the core assumption is that the employment transitions that we observe reflect (at least in part) firms' preferences for who to hire (such that the labor market transitions are not driven only by scientists with greater breadth or diversity of knowledge 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 (relative to other components of knowledge) in their human capital decisions.

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 a research project uncovers a particularly valuable finding for commercial exploitation. We use 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 yields 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 based on the demand for scientific labor in corporate R&D programs. 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 scientists change their rate of conceptual exploration in their research prior to transitioning to industry employment. This suggests that individual scientists are not making changes to their research in anticipation of moving to industry. Nonetheless, the OLS estimates may be biased if some scientists' motivation for moving to industry is a specific desired post-transition research trajectory. For example, scientists may be moving to industry to exploit a commercially valuable finding from their prior research (which would lead to downward bias in the OLS estimates) or to switch to doing a different kind of research (which would lead to upward bias). By focusing on changes in employment linked exogenous variation in the demand for scientific labor, we can identify the effect for those scientists who were the most indifferent to working in industry to academia and only move to industry due to demand shocks.

The first instrument is the state-level R&D tax credit. This variable captures the relative cost to firms in that state of investing in R&D based on the tax credits available to them for these expenditures. Wilson (2009) and Rao (2016) show that that U.S. firms 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. Scientists employed by universities and hospitals are not included in this category (these employers have their own respective NAICS categories). We take a Baritk-style approach based an shocks to an initial variable from pre-sample historical base period receiving exogenous shifts (Bartik,

1991; Blanchard & Katz, 1992; Card, 2001; Autor, Dorn & Hanson, 2013). We defining a state's pre-sample employment in this category as a share of the state's labor force based on its level in 1995. We then multiply each state-year observation 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 industry employment opportunities facing scientists in each state. This variable is used to provide variation in the availability and attractiveness 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 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 underpinning each instrument is 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 in a firm's location, 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 Section 4, we show there is no anticipation effect with respect to pre-transition changes in scientists' research direction. This implies that scientists do not anticipate greater opportunities for transitioning to industry employment in the future and, because of this, change their research trajectory. Nonetheless, we must assume that changes in the national level of scientific R&D employment or in local R&D tax credits are exogenous to any changes in individual scientists' preferences for academic

or industry employment that are based on a scientists' desire for a specific research direction that is made more attractive by moving to industry. 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 estimating the effect of the instruments on scientists' affiliations in the first stage. We include an indicator variable in the model to denote years in which the dependent variable is undefined while arbitrarily setting the value of the dependent value to zero for these years.

4. Results

4.1. Industry Employment and Research Direction

4.1.1. 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' research covers concepts associated with new-to-the-scientist MeSH after transitioning to industry employment. Models 1 and 2 show that the share of MeSH descriptor terms indexed to a scientists' research that had not previously been indexed to their publications is significantly 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 both Models 1 and 2 suggest that scientists' share of new-to-the-scientist MeSH terms increases by approximately 2.1 percentage points following transitions to industry employment. This is equivalent to a 4.4 percent increase in the share of terms that represent new scientific content relative to the mean of the sample (which is 0.49).

[INSERT TABLE 2 HERE]

Figure 1 provides an event study scientists' the rate at which scientists' research covers new-to-the-scientist concepts for a ten-year window from five years before to five years after scientists' transition to industry employment. 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 exploring new-to-the-scientist concepts in their research projects increases only after transitions to industry. Figure 1 also shows that the increase in exploration of new concepts increase does not simply represent a one-off shift in research direction after transitions to industry employment. Rather, scientists moving to

industry experience a sustained change in their research strategies to increase exploration of new-to-the-scientist concepts on an ongoing basis.

[INSERT FIGURE 1 HERE]

The core results from Table 1 are similar when new-to-the-scientist concepts are defined at different levels of the MeSH hierarchy. The results in Table A.2 of the Online Appendix show results from models in which the dependent variable is defined at alternative levels of granularity: the descriptor term-qualifier pair level; the 9-digit level; and the 6-digit level. All results are positive with p-values below the 0.01 level. However, the increase in the rate of conceptual exploration relative to the sample mean is greater when we define exploration as being over greater distance in MeSH-defined conceptual space. The results imply that scientists acquiring affiliations are not simply exploring more new concepts that are very close to their existing knowledge. Instead, industry affiliations are associated with scientists not only exploring more new-to-the-scientist concepts but also going further in conceptual space to find them. Figure A.1 replicates the results from Figure 1 at the different levels of granularity in the MeSH hierarchy and yields very similar patterns with no evidence of anticipatory effects. Figure A.2 replicates the event studies of Figures 1 and A.1 for scientists' transitions between academic employers. There is no evidence of a sustained change in research strategies toward increased rate of exploring new-to-the-scientist concepts. There is a single one-year increase in conceptual exploration that does not persist and has a peak magnitude that is 33 to 50 percent smaller than that associated with academia-to-industry transitions across models.

As an additional robustness check, Table A.3 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 covers new-to-the-scientist terms increases by approximately 7 percent following transitions to industry. The percentage increase in exploration implied by the point estimates range from approximately 6 percent (at the most granular level) to 9 percent (at the least granular level). The models in Table A.4 limit 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.

4.1.2. Returnees to Academia

If the relevant mechanism is industry employment causing scientists to change their research approach, the rate at which scientists' research covers new-to-the-scientist concepts after returning to academic employment should be similar to their pre-industry rates. 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. Scientists who either are working in industry in the first year of the sample or who move to industry but never return to academia are excluded from the sample. However, including all scientists has negligible impact on the magnitude or significance of the coefficients. The results of this analysis are presented in Column 3 of Table 2. 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.

We find that scientists who return to academia after working in industry have similar rates of exploring new concepts as they had prior to their industry affiliation. Scientists engage in research that involves new-to-the-scientist concepts at a higher rate upon acquiring an industry affiliation than during their prior academic employment. However, upon returning to academia, they once again exhibit a similar rate of conceptual exploration as during their prior academic employment. Importantly, 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 only transitions to industry employment that are associated with a change in scientists' rate of exploring new concepts. Career transitions back to academic employers do not have similar results. Instead, scientists' rate of conceptual exploration returns to very close to that prior to their industry employment. This is consistent with an industry-specific effect on scientists' exploitation and further development of their human capital. In Table A.5, we replicate this result for each of the dependent variables defined at different levels of the MeSH hierarchy.

4.2. Alternative Explanations

4.2.1. Other Differences 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 and 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 CHI journal ranking reflects how applied or basic is the scientific research underlying the articles in a focal journal. More information on the construction of these variables is provided alongside the results in Table A.6 the Online Appendix.

There are no clear changes in the number of publications per year or the number of citations scientists receive per year after acquiring industry employment. Nor is there any evidence of a significant change in the median CHI ranking of the journals in which a scientist published. Overall, it does not appear that scientists are 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, they are applying their scientific human capital across a wider range of concepts in MeSH space.

4.2.2. Changes in Stem Cell Funding

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, 2023). 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.7 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 equilibrium between scientists' research choices and funding provision. It would also 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, 2023). Both of these would cause downward bias to our main results. 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.8. These control scientists are drawn from countries with policies permitting human embryonic stem cell research during the sample period. Again, the point estimates are significant and comparable to those in the core analysis, which suggests the main results are not driven by U.S.-specific changes during this period.

4.2.3. Career Stage, Cohort and Team Composition

A further concern is that our core findings may primarily result from scientists' who transition to industry at a particular career stage, such as after a PhD or Post-Doc. Alternatively, they may be driven by a particular generational cohort of scientist that is not representative of the wider scientific community. However, this does not appear to be the case. In Table A.9, we separately replicate the core results splitting the sample according to scientists' career stage and their year of first publication. The results are consistent across sub-samples.

As an additional robustness check, we control for both 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 Panel A of Table A.10). 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 a more diverse set of topics in industry employment (just in larger teams). Alternatively, 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. However, the results show that the relationship between industry affiliation and rate of exploration of new-to-the-scientist concepts is positive and significant across models (and uncorrelated with the number of co-authors on a project.

This is also an important robustness check for the risk that the results may be linked to misclassification of scientists in industry employment. In some instances, affiliations may be imperfectly recorded by bibliographic databases based on the affiliations information on the original publication. An academic scientist may incorrectly appear to be affiliated with an industry actor when, in fact, it was one of their co-authors. If scientists' type of employer is randomly misclassified based on imperfect recordings in Scopus, this would lead to measurement error attenuating our results. However, this would be different if the following three conditions were met: (1) Academic scientists' affiliation were more likely to be recorded with error when they have more co-authors; (2) Academic scientists were more likely to have a corporate co-author when they have more co-authors; and, (3) Academic scientists' research were to typically cover more new-to-the-scientist research when they have more co-authors. In this case, the probability of incorrect classification of a scientist to industry would be increasing in the number of co-authors. If an academic scientist were more likely to be misclassified as working in industry when they have more co-authors and their research were to typically cover more new-to-the-scientist research when they have more co-authors, this would bias our estimates upwards. Nonetheless, the results in Panel B of Table A.10 show that the main effect when adding co-authorship controls is similar to the core results in Table 2, suggesting that the number of co-authors does not appear to be driving the main effect.

There may be a similar concern that affiliations are more likely to be miscategorized when a scientist has more publications in a given year and those with more publications per year are more likely to have research cover new-to-the-scientist concepts. Alternatively, scientists' increased propensity to explore new concepts after moving to industry may be driven by a subset of highly prolific scientists who have many publications in industry. To check whether this may be affecting our results, we follow a similar process, but this time control for the number of publications a scientist has in a focal year. We then additionally interact this with

the corporate affiliation indicator variable. There is no evidence in the results that the main effect of is driven by the number of publications scientists have in a given year.

4.2.4. Using Instruments for Employment Transitions

While the robustness tests help rule out specific alternative explanations, there are still concerns that scientists' selection into industry employment may bias the results. For example, one key risk is that academic scientists may be more likely to enter industry employment if they have made a commercially valuable finding in their recent research. This would bias OLS estimates downward as scientists would be selecting into industry employment to focus on exploiting ideas from their prior research. Thus, they would be less likely to explore new topics after moving to industry compared to those scientists moving to industry for other reasons. Nonetheless, academic scientists may wish to move to industry employment for a range of reasons that are independent of their preferences over specific research topics—for example, because of salary differences between types of employers.

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 within a state over time 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 effect (LATE) in 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—increasing industry demand for scientific labor—but who otherwise would remain working in academia. The models also include a fixed effect denoting each year since a scientist moved state. This blocks a potential path in which changes in the values of the instruments being driven by changes in scientists' location (suggesting transitions between employers) may be independently associated with changes in scientists' research direction. (Although, as shown in Table 2, we find no evidence that employment transitions, other than to industry are associated with increases in the dependent variables).

Table 3 presents the results of the instrumental variables regressions. There is a positive relationship between industry employment and scientists' research covering more new-to-the-scientist conceptual space in 2SLS models using one or both instrumental variables. For all

models, we report the Kleibergen-Paap F statistic and p-values for both the Hansen J statistic and the Anderson-Rubin test (Andrews, Stock & Sun, 2019). As expected, the estimates of the coefficients on the industry employment variable are positive and larger in magnitude than in the baseline OLS models. This suggests that core results could have been biased downward by some scientists selecting into industry employment to exploit commercially valuable research findings. It is also plausible that the scientists who would choose to transition to industry employment when its relative attractiveness increase (and hence are relevant to estimating the LATE), are more willing than other scientists to change their research direction when faced with increased pecuniary incentives.

[INSERT TABLE 3 HERE]

Notably, the two instruments yield very similar estimates of the LATE. Since the underlying casual path is similar in both approaches, the similarity of the LATE estimates enhance confidence in the results. In the respective approaches, changes in the availability of R&D tax credits or exogenous changes in national R&D environment increases industry labor demand. In turn, this leads to increased transitions to industry employment among scientists who are more susceptible to transition to industry when these jobs are more available and attractive. Thus, it should be the same types of scientists with moving to industry for the same type of reasons in both 2SLS models.

Overall, the 2SLS results suggests that there may be two important 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 industry employment. In the second, for the mass of scientists who form the backbone of the corporate scientific labor force—and who choose industry jobs when they are a relative more attractive type of employment, such that their transitions drive the LATE—transitions lead to greater exploration of new-to-the-scientist concepts. Additional results are presented in Tables A.11 and A.12 of the Online Appendix show that the estimates are of similar magnitude when including only one of the two instrumental variables in separate regressions. We also show that the results are not sensitive to whether standard errors are clustered at the scientist or state level.

As an additional robustness check, in Tables A.13 and A.14, we exclude all scientists who transition between academic and industry employment while simultaneously changing their

state of employment. This means that the LATE estimate is based on scientists who switch to industry employment in their state of residence when that state's industry demand for scientific labor exogenously shifts. The results are consistent with those presented in the main analysis albeit marginally larger in magnitude. This would be consistent with scientists transitioning to an industry employer in another state having slightly lower rates of exploring new-to-the-scientist concepts than scientists transitioning to industry employers in their original state. Plausibly those who move state when acquiring industry employment require a greater compensating differential for making this move, which they may be more likely to receive when they have commercially valuable findings from recent research projects that a firm wishes to exploit.

4.3. Heterogeneity in Scientists' Research Changes

4.3.1. Type of Knowledge Re-use

The results so far have shown that scientists explore new-to-the-scientist concepts at a higher rate after acquiring industry affiliations. We now analyze changes in the indexation of MeSH terms to scientists' research according to whether they are in the tools and techniques MeSH category or in other branches of the MeSH tree. If firms' hiring decisions are motivated by scientists' ability to apply their research skills to a wide range of topics relevant to the firm—but which may be new to the scientist—this should be apparent in the types of concepts indexed to their research when working in industry. Specifically, we would expect to find that industry scientists' research projects have a higher rate of re-use of previously known tools and techniques while being less likely to cover previously known concepts at a higher rate in other MeSH categories. We also examine whether there are differences in the results according to whether scientists have had greater scientific impact prior to their transitions to industry.

In Table 4, we replicate the core model from Column 2 of Table 2 and then breakdown whether there are changes in the number of MeSH terms indexed to publications that have previously indexed to the scientists' research in the tools and techniques and other MeSH categories respectively. The results in Panel A show that industry scientists' research output is more likely to involve repeat tools and techniques, but less likely to contain repeated terms in other MeSH categories. We report the incidence rate ratios from Poisson regressions and find an 8 percent increase in the rate of repeat indexation of MeSH terms from the tools and techniques category to scientists' publications. However, when we carry out the same analysis

across MeSH categories, we find a 4 percent decrease in the rate of repeat indexation of MeSH terms to scientists' publications.

[INSERT TABLE 4 HERE]

4.3.2. Scientists' Prior Impact

Next, we examine: (1) Whether scientists with greater prior scientific impact (as captured by their peers' citations to their past research publications) exhibit a lower propensity to explore new-to-the-scientist concepts when in industry employment compared to scientists with lower prior impact; and, (2) Whether the research of industry-affiliated scientists with greater prior scientific impact is more likely to include concepts that they had previously been indexed to their research by the NLM. As discussed in Section 2.4, this would be consistent with an interpretation in which scientists who have received greater scientific recognition linked to their existing human capital are more likely to continue to work in these areas in industry employment—potentially serving as repositories of specialized expertise or brokers to the wider scientific community. Conversely, scientists with less recognition by their peers would be more likely to have their human capital applied to a wider range of projects by an industry employer.

We use scientists' citation impact to measure the recognition of a scientists' ex ante human capital. We first calculate the median number of citations received by scientists to papers published during the pre-sample period from 1990 to 1995 inclusive relative to other scientists with the same first publication year. 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 also exclude those who have no observed publications prior to 1996 because we cannot make a pre-sample impact calculation for these individuals. We then calculate the ratio of citations received by a focal scientist during this period relative to the median for the cohort of scientists sharing their first publication year. We interact this variable with the industry employment indicator variable to allow for heterogeneity in response to industry employment according to scientists' recognition for their prior research.

The results are presented in Panel B of Table 4. Model 1 replicates the core model from Model 2 of Table 2 including this interaction term. The results show that as scientists' prior citation impact increases relative to their peers, they are relatively less likely to explore new-to-the-scientist concepts when in industry employment. Model 2 indicates that the research of

scientists with greater prior impact has a similar propensity to contain repeat tool and technique terms to other scientists transitioning to industry. However, the results from Model 3 show that where scientists' prior impact is greater, it becomes more likely that their research contains repeated terms from other MeSH categories, compared to other scientists in industry employment. This is consistent with scientists with higher quality or reputations being more likely to be hired for, and continuing to exploit, existing areas of knowledge in industry employment (compared to other scientists who acquire industry affiliations). In Table A.15, we find a similar pattern of results (albeit quantitatively larger and more precisely estimated) using a binary prior impact variable.

4.4. Human Capital of Scientists Moving to Industry

If firms place particular value on scientists' research skills in their human capital strategies, there should be evidence of this in hiring. Firms should be more likely to hire scientists from academia who have used a wider range of research skills, tools, and techniques in their past research compared to scientists with narrower skill sets, other things being equal. The results in Table 4 are from use complementary log-log 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.

[INSERT TABLE 5 HERE]

In Model 1 of Table 5, 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 of increases in the number of terms in scientists' prior knowledge set on transitioning to industry employment for each MeSH category. 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. The point estimates of the average marginal effects suggest that a one standard deviation increase in the variable that measures the breadth of scientists' previously used techniques is associated with approximately a 0.30 percentage point increase in the probability a scientist acquires a first industry affiliation. Relative to the baseline probability that a scientist transitions to industry employment in a given year (1.33 percentage points), these are equivalent to increases in transition rates of approximately 22 percent respectively.

The results in Model 2 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 implies that a one standard deviation increase in diversification in this MeSH category is associated with a 0.23 percentage point increase in the rate of transition to industry employment on average. This is equivalent to approximately an 18 percent increase in the rate of transitions to industry relative to the baseline. The results in Table 5 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 more diverse range of them (rather than predominantly specializing in some of these techniques or tools).

The results are inconsistent with respect to scientists' prior use chemicals and drugs and phenomena and processes involved in scientists' prior research. While scientists moving to industry appear to have previously used a greater range of chemicals and drugs—which are typically interventional agents (see Online Appendix for more information)—there is not clear evidence that they are more generalists than other scientists in this domain. Conversely, while scientists transitioning to industry employment are less likely to specialized with respect to the physiological processes in their research, there is not clear evidence that they have a greater range of knowledge in this domain. Neither type of model suggests that scientists transitioning to industry employment have a greater range of prior knowledge or are more or less specialized than other scientists on disease or anatomical topics.

Next, we seek to further show the value of considering knowledge and skills as important components of scientists' human capital—in the context of transitions from academia to industry—by contrasting our results in Model 2 with those that are obtained using a single diversification index defined by terms indexed to research across all MeSH space (not separately by MeSH categories). The results indicate that there is no relationship between overall prior research diversity and scientists moving to industry. Neither generalists nor specialists appear more likely to transition to industry in this setting. Yet, as the results in Model 2 show, we find that those with more diverse past use of tools and techniques are more likely to transition to industry in our sample. This finding emphasizes the value of considering knowledge and skills as distinct parts of scientists' human capital when examining scientists' career outcomes.

As a robustness check, in Table A.16, we replicate these results using linear probability models to estimate the relationship between scientists' ex ante knowledge and transitions to industry employment. The results follow the same pattern as Table 5. The breadth and diversity of the tools and techniques terms previously indexed to a scientist's research is consistently positive and significant in all models (with p-values ranging from 0.014 to 0.096). This includes robustness tests using highly restrictive models with scientist fixed effects. These yield conservative estimates based on within-scientist variation in the independent variables. If a scientist's research strategy has involved using a wider range of techniques across projects consistently in their career, this will be differenced out by the fixed effect. Notably, none of the independent variables capturing scientists' ex ante knowledge characteristics in the other categories of MeSH terms have a consistent positive relationship with transitions to industry employment that is significant at any conventional levels (at most there is a significant (p-value$<$0.1) relationship in only one of the four models for each of the chemicals and drugs and phenomena and processes category variables).

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 wide range of terms. Further research, at a more granular level, 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: (1) Whether the breadth of their topic knowledge makes them specialists or generalists (Agrawal et al., 2016; Teodoridis, 2018; Teodoridis et al., 2019); and (2) By the extent of their relative 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 appreciate how scientists' research skills represent an important component of scientific human capital which is not fully captured by either the existing generalist-specialist or basic-applied categorizations. This delineation between 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 move to industry employment have more conceptually diverse subsequent research output. It is important to note that this does not simply represent a shift in research direction immediately after transitioning a new employer; it is a persistent increase in the rate at which their research projects cover concepts that are new to the scientist. This finding is consistent with firms' scientific human capital strategies preferring to hire academic scientists whose research skills that can be applied more widely across research projects in their field. In line with this interpretation, we find evidence that firms select scientists with experience using a greater, more diverse range of tools and techniques in their research prior to entering industry employment. Returning to the earlier example from the social sciences, an economist with narrow topic knowledge but more wide-ranging research 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 skill set. However, in the context of firms hiring scientists from academia, our results suggest that researchers with these two human capital configurations will be viewed quite differently. The individual with the wider skill set appears to be more attractive to firms and is more likely to transition to industry employment.

Our findings contribute to the literature on scientific labor markets (Stern, 2004; Roach & Sauerman, 2010, 2017; Agarwal & Ohyama, 2013; Conti & Visentin, 2015; Hoisl & Mariani, 2017; Ding et al., 2021). They indicate that there is a link between the breadth and diversity of scientists' research skills and career transitions to industry employment. We also delineate how transitions affect the ways in which scientists exploit and further develop their human capital when working in industry. This helps adds to our understanding the role that specialized scientific knowledge and skills appear to play in firms' human capital strategies and scientists' career outcomes. A plurality of those graduating from PhD and Post-Doctoral programs in life sciences are hired by industry employers within ten years of graduation (Kahn & Ginther, 2017). Our results have implications for the organization of training programs with respect to ensuring graduates have the skills that will make them more attractive to industry employers.

An additional stream of research in the science of science has focussed on distinctions between basic and applied research to analyze how institutional features drive the optimal division of scientific labor between academia and industry (Bush, 1945; Nelson, 1959; Rosenberg & Nelson, 1994; Aghion et al., 2008; Lacetera, 2009). This paper finds 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-to-the-scientist 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. This has implications for more policy-focused concerns about 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; Sohn, 2021; Fini et al., 2022). Understanding how transitions between academia and industry affect scientists' use and development of their human capital helps us to identify 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; Kaplan & Vakili, 2015), a greater (lesser) share of science taking place in industry may decrease (increase) the rate at which these occur. Conversely, in circumstances in which individual-level knowledge diversity enables scientists to make 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.

We focus on the role of skills in scientists' career transitions between academic and industry employment. In other contexts, certain types of knowledge may also be relevant considerations when analyzing scientists' research choices. For example, recent research has examined which scientists pivot as diseases emerge or increase in salience (Fry, 2023, Hill et al, 2025). Using similar approaches to analyze the role of different components of scientists' knowledge may be valuable for future research in this vein. As new technologies, such as CRISPR or mRNA vaccines, emerge and lead to new research opportunities, examining different components of knowledge—beyond experience with techniques—may shed more light on the nuances of which

scientists pivot to take advantage of these opportunities. For example, the extent to which scientists are specialists or generalists in particular diseases or physiological process may influence the extent to which they pivot into new areas and impact their subsequent research outcomes.

Nonetheless, there are a number of limitations to this study. First, as is typical for studies using this type of bibliographic data, the keywords indexed to 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 or may mistakenly attribute team member's knowledge to a focal scientist. The broad MeSH categories also represent imprecise proxies for skills using highly specific tools and techniques. Future research that provides a more comprehensive matching between granular MeSH terms and specific types of knowledge 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. 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. Future research might also valuably identify how different types of firms vary in their hiring and use of scientific labor. 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.

Andrews I., J.H. Stock, and L. Sun (2019) "Weak instruments in instrumental variables regression: Theory and practice," *Annual Review of Economics*, 11: 727–753.

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) "Evaluating technological information and utilizing it: Scientific knowledge, technological capability, and external linkages in biotechnology," *Journal of Economic Behavior & Organization*, 24(1): 91–114.

Autor D., D. Dorn, and 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 impact 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., 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., F. Murray, and J. Gans (2015) "Exploring trade-offs in the organization of scientific work: Collaboration and scientific award," *Management Science* 61(7): 1473–1495.

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 (2023) "Incentivizing effort allocation through resource allocation: Evidence from scientists' response to changes in funding policy," *Organization Science*, 34(1): 100–128.

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.

Cassiman B., R. Veugelers and S. Arts (2018) "Mind the gap: Capturing value from basic research through combining mobile inventors and partnerships," *Research Policy*, 47(9): 1811–1824.

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.

Conti A., and F. Visentin (2015) "A revealed preference analysis of PhD students' choices over employment outcomes," *Research Policy*, 44(10): 1931–1947.

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.

Ding W., A. Ohyama, and. R. Agarwal (2021) "Trends in gender pay gaps of scientists and engineers in academia and industry," *Nature Biotechnology*, 39(8): 1019–1024.

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.

Fry C. (2023) "Crisis and the trajectory of science: Evidence from the 2014 Ebola outbreak," *Review of Economics and Statistics*, 105(4): 1028–1038.

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.

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.

Hess, A.M., and F.T. Rothaermel (2011) "When are assets complementary? Star scientists, strategic alliances, and innovation in the pharmaceutical industry," *Strategic Management Journal*, 32(8): 895–909.

Hill, R., Y. Yin, C. Stein, X. Wang, D. Wang, and B.F. Jones. "The pivot penalty in research." *Nature*, 642: 999–1006.

Hoisl K., and M. Mariani (2017) "It's a man's job: Income and the gender gap in industrial research," *Management Science*, 63(3): 766–790.

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.

Jain, A. (2016) "Learning by hiring and change to organizational knowledge: Countering obsolescence as organizations age." *Strategic Management Journal*, 37(8): 1667–1687.

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): 1435–1457.

Kahn S., and D.K. Ginther (2017) "The impact of postdoctoral training on early careers in biomedicine," *Nature Biotechnology*, *35*(1), 90–94.

Kaiser U., H.C. Kongsted, K. Laursen, and A.K. Ejsing (2018) "Experience matters: The role of academic scientist mobility for industrial innovation," *Strategic Management Journal*, 39(7): 1935–1958.

Kehoe R.R., and D. Tzabbar (2015) "Lighting the way or stealing the shine? An examination of the duality in star scientists' effects on firm innovative performance," *Strategic Management Journal*, 36(5): 709–727.

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.

Latour B. (1987) *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, MA: Harvard University Press.

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.

Mokyr, J. (2005) "Long-term economic growth and the history of technology," in P. Aghion and S.N. Durlauf (Eds.) *Handbook of Economic Growth*, Vol. 1, Ch 17. Amsterdam, NL: Elsevier.

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.

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

Roach M., and H. Sauermann (2024) "Can technology startups hire talented early employees? Ability, preferences, and employee first job choice," *Management Science* 70(6): 3619–3644.

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.

Sohn E. (2021) "How local industry r&d shapes academic research: Evidence from the agricultural biotechnology revolution," *Organization Science*, 32(3): 675–707.

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.

Tzabbar D. (2009) "When does scientist recruitment affect technological repositioning?" *Academy of Management Journal*, 52(5): 873–896.

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 (1996) "Star scientists and institutional transformation: Patterns of Innovation in the Formation of the Biotechnology Industry," *Proceedings of the National Academy of Sciences*, 93: 12709–12716.

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	76,984	0.13	0	0.34	0	1
% New MeSH Terms	42,911	0.49	0.46	0.22	0.23	0.79
% New Terms including Qualifiers	42,911	0.61	0.60	0.19	0.36	0.86
% New Terms at 9-digit Level	42,911	0.37	0.32	0.23	0.12	0.69
% New Terms at 6-digit Level	42,911	0.27	0.21	0.23	0.05	0.58
ln(1+Cumulative Publications)	76,984	2.53	2.48	1.17	1.10	4.08
ln(1+Cumulative MeSH Terms Used)	76,984	4.47	4.62	1.16	3.14	5.75
ln(1+Cumulative Citations Received)	76,984	6.14	6.40	1.83	3.78	8.19
Career Age (Years Since First Pub)	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	76,984	99.2	12	244.7	0	278
Median Pub CHI Journal Ranking	35,615	3.00	3	0.80	2	4
Post-Corporate Academic Affiliation	65,211	0.05	0	0.21	0	0
First Industry Employment Year = 1	62,049	0.01	0	0.11	0	0
First Publication Year	76,984	1990.1	1991	6.53	1980	1998
State R&D Tax Credit Rate	72,517	0.06	0.06	0.05	0	0.15
State Predicted Scientific Labor Demand	72,517	0.08	0.06	0.08	0.00	0.24
N. Repeated Tech Terms Per Pub	42,911	0.85	0.67	0.93	0	2
N. Repeated Non-Tech Terms Per Pub	42,911	7.97	8	3.76	3	12.6

Notes: Instrumental variables are only defined for scientists who have not left the United States by the year of observation. 'Post-Corporate Academic Affiliation' variable is only defined for 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. 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. It is undefined for years after a scientist has a first observed industry affiliation and for those with industry affiliations prior to the start off the sample period. The indicator is set to zero if the scientist has not yet had (or never has) an observed industry affiliation. State Predicted Scientific Labor Demand is reported with 1,00x the value of the variable for clarity.

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

	% New MeSH Terms		
	(1)	(2)	(3)
Corporate Affiliation	0.021	0.021	0.027
	(0.005)	(0.005)	(0.008)
	[0.000]	[0.000]	[0.000]
Post-Corporate Academic Affiliation			0.005
			(0.007)
			[0.422]
Controls	No	Yes	Yes
Age FE	No	Yes	Yes
Scientist FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Scientists	5,312	5,312	4,494
Observations	42,911	42,911	37,387

Notes: All estimates in Columns (1) to (3) are based on panel OLS regression with year and scientist fixed effects. For inclusion in the estimation sample for Model 3, scientists must not have been working in industry in the first year of the sample period and either: (1) transitioned to industry and returned to academic employment before the end of the sample period; or (2) be employed only in academia during the 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 with p-values in brackets.

Table 3: Changes in scientists' research when having a corporate affiliation (2SLS results)

	Only Tax Credit Instrument		Only Employment Instrument		Both Instruments	
	Corp Aff	% New Terms	Corp Aff	% New Terms	Corp Aff	% New Terms
	(1)	(2)	(3)	(4)	(5)	(6)
Corporate Affiliation		0.137 (0.065) [0.041]		0.145 (0.050) [0.006]		0.141 (0.048) [0.005]
R&D Tax Credit Rate	0.292 (0.070) [0.000]				0.210 (0.056) [0.000]	
Predicted Labor Demand			24.453 (3.461) [0.000]		17.512 (4.269) [0.000]	
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
K.-P. F Statistic	17.7		49.9		29.8	
Hansen J Statistic		--			--	0.89
A.-R. Wald Test		0.030		0.000		0.003
Scientists	5,299	5,299	5,299	5,299	5,299	5,299
Observations	72,517	72,517	72,517	72,517	72,517	72,517

Notes: All estimates are based on panel OLS regression with year and scientist fixed effects. Additional models in the appendix show that the results are similar if these observations are not included. We also include a fixed effect for the number of years since a scientist moved state. This is intended to mitigate the risk that changes in values of the instruments are driven by changes in state location associated with changes in scientists' research direction. Other control variables are for lagged citations, publications, and cumulative terms used (in logs). P-values are reported for Hansen J statistic and Anderson-Rubin Wald test for weak-instrument robust inference. Observations for scientists who leave the U.S. during the sample period are dropped from the analysis (as the instruments are U.S.-specific) from the first year in which a scientist does not have a U.S.-based affiliation. Results (available from the author) are similar if all scientists who ever leave the U.S. are dropped from the estimation sample. Standard errors (in parentheses) are clustered at the scientist level with p-values in brackets.

Table 4: Heterogeneity in the changes in exploration of new-to-the-scientist concepts by scientists' prior impact

	% New MeSH Terms	N Repeated Tech MeSH Terms	N Repeated Other MeSH Terms
	(1)	(2)	(3)
Panel A: Main results			
[N Scientists=5,312, Obs=42,911]			
Corporate Affiliation	0.021	1.080	0.963
	(0.005)	(0.030)	(0.011)
	[0.000]	[0.006]	[0.001]
Panel B: Results by scientists' pre-sample impact			
[N Scientists=3,806, Obs=35,314]			
Corporate Affiliation	0.022	1.089	0.964
	(0.005)	(0.033)	(0.012)
	[0.000]	[0.005]	[0.003]
Corporate Affiliation	-0.009	1.005	1.013
*Pre-Sample Impact	(0.003)	(0.022)	(0.007)
	[0.005]	[0.823]	[0.059]
Controls	Yes	Yes	Yes
Age FE	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes

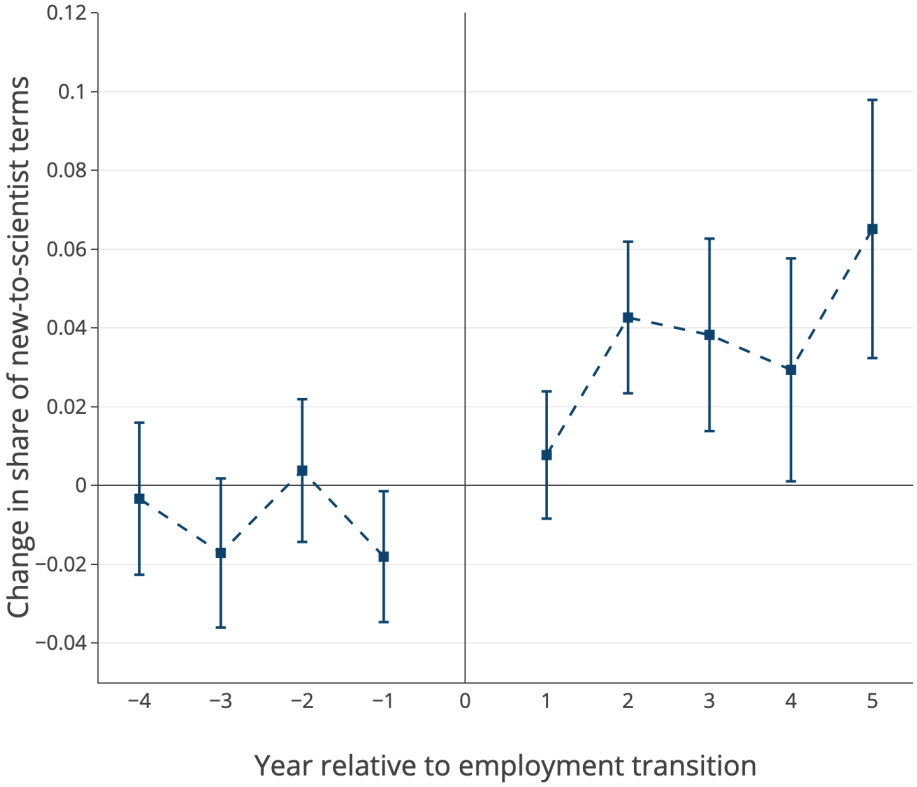
Notes: All estimates are based on Poisson regression with year and scientist fixed effects with incidence rate ratios reported, except Model (1) which reports results from OLS regressions. The dependent variable in Model (1) is the same core dependent variable as Models (1) and (2) in Table 2. The dependent variable in Model (2) is the mean number of MeSH terms indexed to scientist i's publications in year t that had previously been indexed to scientist i's research and are in the Tools and Techniques MeSH category. The dependent variable in Model (3) is the mean number of MeSH terms indexed to scientist i's publications in year t that had previously been indexed to scientist i's research and are in all other MeSH categories. Only scientists with a first publication before 1996 are included in the Models in Panel B to allow for the calculation of the pre-sample impact variable. Pre-sample impact is measured in terms of ratio of the total number of citations to scientist i's papers from before 1996 to the median total number of citations to papers from before 1996 received by scientists with the same first publication year as scientist i. The variable is standardized to provide an equivalent interpretation of main effect across models and with prior tables. The main effect of the pre-sample impact variable is absorbed by the 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 with p-values in brackets.

Table 5: Hazard models of transitions to industry based on ex ante human capital

	Acquisition of First Industry Affiliation		
	Count-based Variables	Diversification Index Variables	
	(1)	(2)	(3)
Techniques & Equipment	0.003	0.010	
	(0.001)	(0.005)	
	[0.051]	[0.030]	
Chemicals & Drugs	0.004	0.008	
	(0.001)	(0.006)	
	[0.002]	[0.216]	
Phenomena & Processes	0.002	0.011	
	(0.001)	(0.005)	
	[0.133]	[0.018]	
Diseases	-0.001	-0.003	
	(0.001)	(0.002)	
	[0.138]	[0.231]	
Anatomy	-0.001	-0.003	
	(0.001)	(0.004)	
	[0.227]	[0.370]	
Other Categories	-0.000	-0.005	
	(0.002)	(0.006)	
	[0.859]	[0.421]	
All MeSH Categories			-0.000
			(0.016)
			[0.998]
Controls	Yes	Yes	Yes
Age FE	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Scientists	4,830	4,830	4,830
Observations	62,049	62,049	62,049

Notes: Models are discrete time hazard models with complementary log-log regressions with average marginal effects reported. Model 1 uses 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 2 and 3 use diversification indices capturing a scientist's diversification over terms for specified MeSH category, defined as $1–\text{HHI}^{0.5}$ (for a scientist's career strictly prior to a focal year). Term shares are defined as the number of times a scientist's research has been indexed to a focal term divided by the number of terms in that category that have been indexed to the scientist's research. 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 exit the sample after acquiring the affiliation. The mean yearly probability a scientist acquires a first industry affiliation is 1.3 percentage points. Cohort fixed effects are defined by the first year in which a scientist ha a recorded publication. Scientists are excluded if they have transitioned to industry by the start of the sample. Control variables for lagged citations and publications. Standard errors (in parentheses) are clustered at the scientist level with p-values in brackets.

Figure 1: Change in the shares of new-to-the-scientist MeSH terms in scientists' research around employment transitions



Notes: The figures represent event study of *% New terms* around transitions from academic to industry employment with scientist, age, and year fixed effects. 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 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 Summary Statistics

Table A.1: Additional summary statistics

Variable	Obs	Mean	Med	SD	p10	p90
ln(1+Cum Anatomy Terms)	62,049	2.34	2.48	1.05	0.69	3.58
ln(1+Cum Chemical & Drugs Terms)	62,049	3.12	3.30	1.26	1.39	4.60
ln(1+Cum Diseases Terms)	62,049	1.93	1.95	1.31	0	3.73
ln(1+Cum Phenomena & Processes Terms)	62,049	2.63	2.77	1.17	1.10	4.01
ln(1+Cum Techniques & Equipment Terms)	62,049	2.44	2.56	1.07	1.10	3.74
ln(1+Cum Other Categories Terms)	62,049	2.73	2.77	0.92	1.61	3.83
Div Index Anatomy Terms	62,049	0.56	0.65	0.23	0	0.76
Div Index Chemicals & Drugs Terms	62,049	0.69	0.77	0.22	0.42	0.86
Div Index Diseases Terms	62,049	0.45	0.55	0.30	0	0.79
Div Index Phenomena & Processes Terms	62,049	0.61	0.71	0.24	0.25	0.81
Div Index Techniques & Equipment Terms	62,049	0.59	0.67	0.23	0.29	0.78
Div Index Other Categories Terms	62,049	0.61	0.65	0.15	0.48	0.73

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. Variables are lagged by one year (other than the indicator variable denoting a scientist's first year with an industry affiliation). Scientists' observations are dropped from the estimation sample in years strictly after the year in which they had the first observed industry affiliation and hence not reported in this table.

Robustness of Main Results to Alternative Definitions of New Concepts

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. The first level of the MeSH hierarchy is a single letter, which locates a concept in one of sixteen categories. The second level of the hierarchy is a three-place alpha numeric code (incorporating the first letter followed by two digits). The next level of tree branching is a six-digit code (incorporating three additional digits from the higher level), which represents the third level of classification. This is followed by branching according to a nine-digit code that adds a further three digits to the six-digit code. Some—but not all—MeSH descriptor terms are organized at a lower, more granular, level than this fourth tier of the MeSH hierarchy. The terms associated with further branching receive additional digits in the hierarchical classification.

For the purposes of our analysis, 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 publicly available crosswalk. Different MeSH descriptor terms can be located at different levels in the tree hierarchy depending on how finely grained the NLM definition of concepts is in that area of science. 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. We then repeat this process for the third level of the hierarchy (the 6-digit level). This leads to the two additional dependent variables: *% New Terms, 9-digit Level* and *% New Terms, 6-digit Level*.

MeSH indexing 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*).

Using the MeSH tree structure to define variables at these consistent levels of analysis helps ensure that the results are not driven by scientists' research covering more new-to-the-scientist terms that are distinct but highly proximate in MeSH space. Publications that cover more new descriptor terms (or descriptor-qualifier pairs) in the same 6- or 9-digit MeSH category may not reflect a scientist is taking on research projects that cover a wider range of new-to-the-scientist concepts compared to publications covering a similar number of new terms (or fewer new terms) that are distinct at the 9-digit level. Instead, they may only be exploring new concepts in the immediate vicinity of their existing knowledge, or which share underlying knowledge with those terms that have previously been indexed to their research. In the results section, we report results using all four dependent variables to show the results are robust across specifications. This implies that our core findings are not driven simply be an increased tendency to explore new-to-the-scientist concepts that are highly proximate to those covered in their prior research.

To provide a further robustness check, we use additional dependent variables that count the total number of new MeSH terms in Poisson regression models (rather than using a percentage-based dependent variable with OLS as in the main resulst). 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 the percentage-based dependent variable 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).

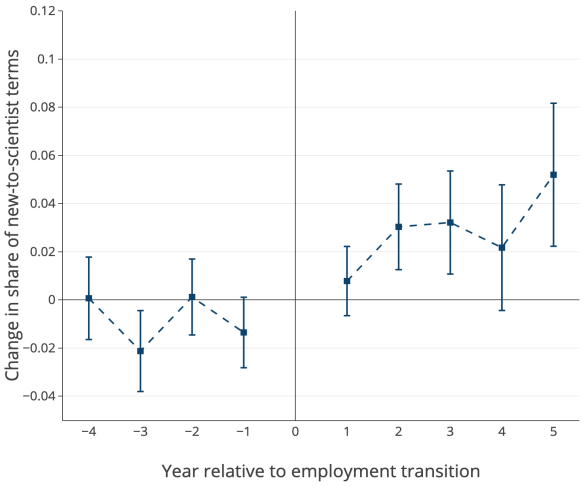
Table A.2: Alternative definitions of new-to-the-scientist terms

	% 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
Mean of DV	0.49	0.49	0.61	0.37	0.27
Scientists	5,312	5,312	5,312	5,312	5,312
Observations	42,911	42,911	42,911	42,911	42,911

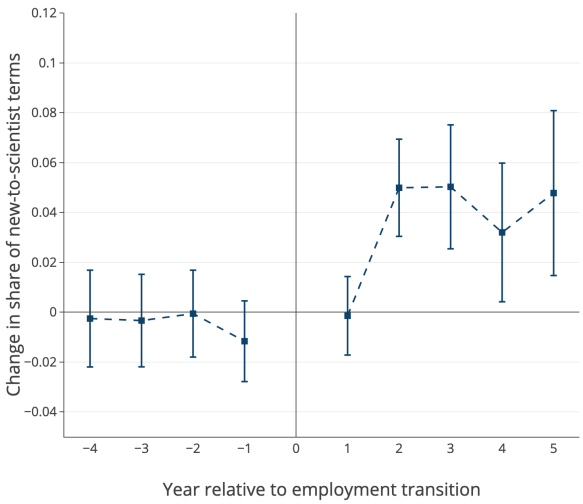
Notes: All estimates are based on panel OLS regression with year and scientist fixed effects. Models (1) and (2) replicate the results in the main paper. Control variables are the same as in Table 2 of the main paper. Sample mean of each dependent variable is showed in the table for comparison purposes. Standard errors (in parentheses) are clustered at the scientist level. ***p<0.001, **p<0.01, *p<0.5, $^{+}$p<0.1.

Figure A.1: Change in the shares of new-to-the-scientist MeSH terms in scientists' research around academia-to-industry employment transitions (alternative dependent variables)

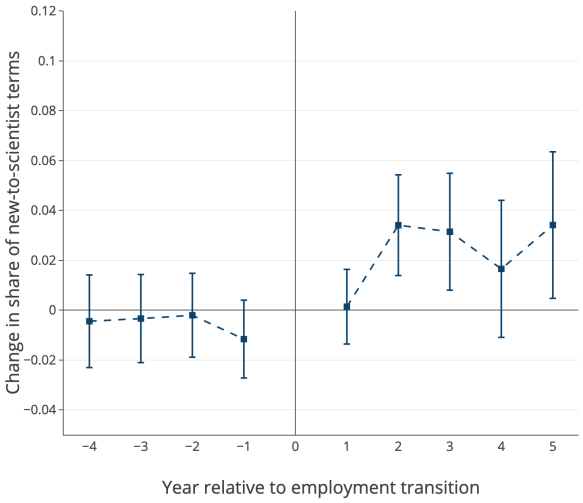
Panel A: % New Terms, including qualifiers



Panel B: % New Terms, 9-digit MeSH level



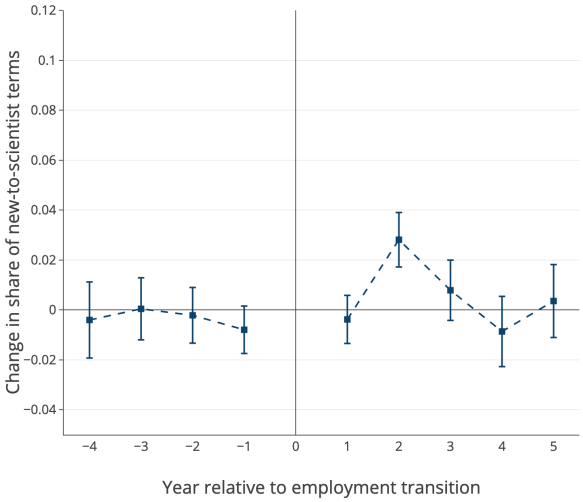
Panel C: % New Terms, 6-digit MeSH level



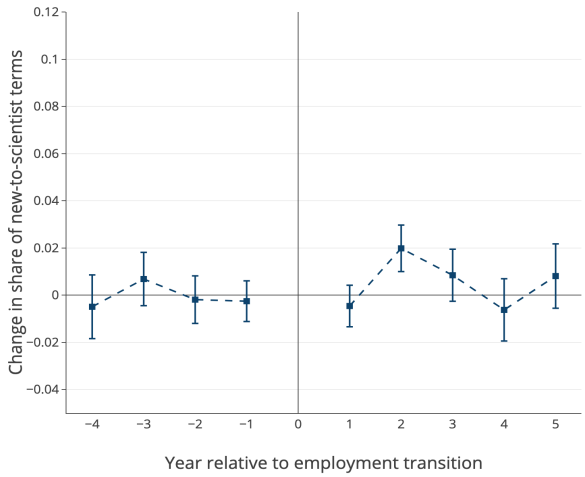
Notes: The figures replicate event studies from Figure 1 for transitions from academic to industry employment with alternative dependent variables (scientist, age, and year fixed effects are included in the models). Year 0 is the year prior to first acquiring an industry affiliation. Scientists transitioning 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 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.

Figure A.2. Change in the shares of new-to-the-scientist MeSH terms in scientists' research around academia-to-academia employment transitions

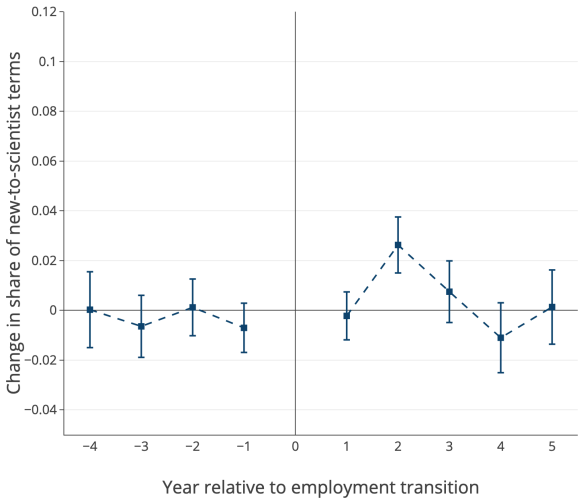
Panel A: % New Terms (core definition)



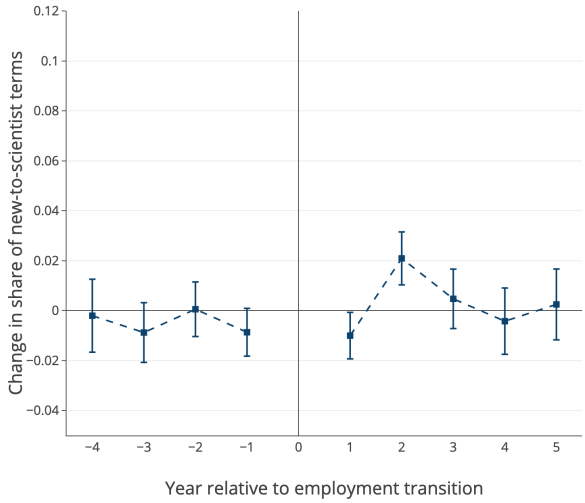
Panel B: % New Terms, including qualifiers



Panel C: % New Terms, 9-digit MeSH level



Panel D: % New Terms, 6-digit MeSH level



Notes: The figures replicate event studies from Figures 1 and A.1 for transitions between academic employers. Year 0 is the year prior to transition. Scientists changing employers are observed for up to five years before and five years after acquiring an industry affiliation. Scientists do not change employers are included as controls. Scientists making a second switch between employers are excluded from the sample from the year of second transition (inclusive). Scientists with industry affiliations during the sample period are excluded from the analysis. Error bars represent 90% confidence intervals.

Table A.3: Using total new terms as a dependent variable in Poisson models (IRR-reported)

	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*** (0.017)	1.068*** (0.016)	1.060*** (0.014)	1.070*** (0.019)	1.087*** (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 the same as in Table 2 of the main paper. 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: New-to-the-scientist terms 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*** (0.006)	0.024*** (0.005)	0.017*** (0.005)	0.025*** (0.005)	0.025*** (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 the same as in Table 2 of the main paper. 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: Changes in scientists' research on returning to academia (alternative definitions)

	% 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. Model (1) replicates the result in Model (3) of Table 2 in the main paper. Control variables and inclusion criteria are the same as in the main paper. Standard errors (in parentheses) are clustered at the scientist level. ***p<0.001, **p<0.01, *p<0.5, $^{+}$p<0.1.

Alternative Explanations for the Main Results

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: (1) The total number of scientist *i*'s publications in year *t*; (2) The mean number of citations scientist *i*'s publications in year *t*; and (3) The median CHI journal ranking of scientist *i*'s 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.

Table A.6: 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 (0.036)	-0.006 (0.035)	0.063 (0.072)	0.069 (0.073)	-0.026 (0.021)	-0.025 (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 A.7: 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*** (0.010)	0.034*** (0.009)	0.029*** (0.008)	0.023* (0.009)	0.018* (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 the same as in the main paper. There are fewer authors than in the full sample because some scientists who have a publication in Scopus between 1996 and 2000 only have publications not indexed in the PubMed database. This means that the dependent variables are undefined during this period. 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: 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 as in the main paper. Standard errors (in parentheses) are clustered at the scientist level. ***p<0.001, **p<0.01, *p<0.5, $^{+}$p<0.1.

Table A.9: 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: Cohorts	*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 B: Career 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 scientist and year fixed effects. Other controls are included as per the core models in the paper. In Panel A, the sample is split according to whether a scientist has a first publication recorded in Scopus strictly before 1990. In Panel B, 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.10: 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)
Number of Co-authors	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
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)
Number of Publications	0.012***	0.013***	0.012***	0.012***	0.012***	0.013***	0.011***	0.012***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
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: Models follow the specifications in Table 2 of the main paper. Models in Panel A additionally control for the number of co-authors on a scientist's publications in a given year. Models in Panel B additionally control for a scientist's number of publications in a given year. Standard errors (in parentheses) are clustered at the scientist level. ***p<0.001, **p<0.01, *p<0.5, $^{+}$p<0.1.

Robustness Tests for the 2SLS Results

Table A.11: 2SLS results (alternative definitions)

	Corporate Affiliation	% New Descriptors	% New incl. Qualifiers	% New 9-Digit Level	% New 6-Digit Level
	(1)	(2)	(3)	(4)	(5)
Corporate Affiliation		0.141**	0.103*	0.120*	0.128*
		(0.048)	(0.047)	(0.052)	(0.052)
R&D Tax Credit Rate	0.210***				
	(0.056)				
Predicted Scientific	17.512***				
Labor Demand	(4.269)				
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
K.-P. F Statistic	29.8				
Hansen J Statistic		0.89	0.88	0.65	0.82
A.-R. Wald Test		0.003	0.060	0.004	0.003
Scientists	5,299	5,299	5,299	5,299	5,299
Observations	72,517	72,517	72,517	72,517	72,517

Notes: Models follow the specifications in Table 3 of the main paper. P-values are reported for Hansen J statistic and Anderson-Rubin Wald test for weak-instrument robust inference. Standard errors (in parentheses) are clustered at the state level. ***p<0.001, **p<0.01, *p<0.5, $^{+}$p<0.1.

Table A.12: 2SLS results (no controls and alternative standard errors)

	No Control Variables		Alternative S.E. Clustering	
	Corporate Affiliation	% New Terms	Corporate Affiliation	% New Terms
	(1)	(2)	(3)	(4)
Corporate Affiliation		0.172*		0.141*
		(0.076)		(0.064)
R&D Tax Credit Rate	0.220***		0.210***	
	(0.056)		(0.050)	
Predicted Scientific	17.512***		17.512***	
Labor Demand	(4.269)		(5.569)	
Controls	No	No	Yes	Yes
Age FE	No	No	Yes	Yes
Scientist FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
K.-P. F Statistic	30.9		17.5	
Hansen J Statistic		0.93		0.91
A.-R. Wald Test		0.052		0.063
Scientists	5,299	5,299	5,299	5,299
Observations	72,517	72,517	72,517	72,517

Notes: Models follow the specifications in Table 3 of the main paper. P-values are reported for Hansen J statistic and Anderson-Rubin Wald test for weak-instrument robust inference. Models (1) and (2) exclude all control variables (but retain the scientist and year fixed effects). In Models (3) and (4), standard errors are clustered at the scientist, rather than state, level. ***p<0.001, **p<0.01, *p<0.5, $^{+}$p<0.1.

Table A.13: 2SLS results excluding scientists transitioning both from industry to academic employment and the state of their affiliation

	Only Tax Credit Instrument		Only Employment Instrument		Both Instruments	
	Corp Aff	% New Terms	Corp Aff	% New Terms	Corp Aff	% New Terms
	(1)	(2)	(3)	(4)	(5)	(6)
Corporate Affiliation		0.262*		0.296*		0.274*
		(0.065)		(0.050)		(0.107)
R&D Tax Credit Rate	0.262**				0.143***	
	(0.124)				(0.040)	
Predicted Lab Demand			12.910***		8.212**	
			(3.497)		(2.831)	
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
K.-P. F Statistic	11.1		13.6		11.2	
Hansen J Statistic						0.77
A.-R. Wald Test		0.013		0.001		0.001
Scientists	5,299	5,299	5,299	5,299	5,299	5,299
Observations	70,085	70,085	70,085	70,085	70,085	70,085

Notes: Models follow the specifications in Table 3 of the main paper. Scientists are excluded from year t (inclusive) if in year t both: (1) their institutional affiliation changes from one state to another; and (2) they acquire a corporate affiliation in the same year. P-values are reported for Hansen J statistic and Anderson-Rubin Wald test for weak-instrument robust inference. Standard errors (in parentheses) are clustered at the state level. ***p<0.001, **p<0.01, *p<0.5, $^{+}$p<0.1.

Table A.14: 2SLS results excluding scientists transitioning both from industry to academic employment and the state of their affiliation (alternative definitions)

	Corporate Affiliation	% New Descriptors	% New incl. Qualifiers	% New 9-Digit Level	% New 6-Digit Level
	(1)	(2)	(3)	(4)	(5)
Corporate Affiliation		0.274*	0.209^{+}	0.210*	0.198^{+}
		(0.107)	(0.106)	(0.103)	(0.115)
R&D Tax Credit Rate	0.143***				
	(0.040)				
Predicted Scientific	8.212**				
Labor Demand	(2.831)				
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
K.-P. F Statistic	11.2				
Hansen J Statistic		0.77	0.92	0.64	0.82
A.-R. Wald Test		0.001	0.029	0.021	0.003
Scientists	5,299	5,299	5,299	5,299	5,299
Observations	70,085	70,085	70,085	70,085	70,085

Notes: Models follow the specifications in Table 3 of the main paper. P-values are reported for Hansen J statistic and Anderson-Rubin Wald test for weak-instrument robust inference. Standard errors (in parentheses) are clustered at the state level. ***p<0.001, **p<0.01, *p<0.5, $^{+}$p<0.1.

Alternative Definitions of Pre-Sample Impact

Table A.15: Heterogeneity in the changes in exploration of new-to-the-scientist concepts by scientists' prior impact (binary independent variable)

	% New Terms	N. Repeat Tech Terms	N. Repeat Non-Tech Terms
	(1)	(2)	(3)
Corporate Affiliation	0.022***	0.087**	-0.035**
	(0.005)	(0.029)	(0.012)
Corporate Affiliation*Pre-Sample Impact	-0.110***	-0.149*	0.114*
	(0.014)	(0.071)	(0.053)
Controls	Yes	Yes	Yes
Age FE	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Scientists	3,806	3,806	3,806
Observations	35,314	35,314	35,314

Notes: All estimates are based on Poisson regression with year and scientist fixed effects, except Model (1) which reports results from OLS regressions. The dependent variable in Model (1) is the same core dependent variable as Models (1) and (2) in Table 2. The dependent variablea in Models (2) and (3) are the same as in Models (4) and (5) of Table 2.. Pre-sample impact is measured as an indicator variable equal to one if the number of citations received by scientist i's papers that were published before 1996 relative to their cohort who are in the top percentile. Standard errors (in parentheses) are clustered at the scientist level. ***p<0.001, **p<0.01, *p<0.5, $^{+}$p<0.1.

Alternative Models of Transitions to Industry Employment

Table A.16: Linear probability models of transitions to industry employment

	Acquisition of First Industry Affiliation					
	Count-based Variables		Diversification Index Variables			
	(1)	(2)	(3)	(4)	(5)	(6)
Tech & Equipment	0.002*	0.005*	0.008*	0.001^{+}		
	(0.001)	(0.002)	(0.003)	(0.001)		
Chemicals & Drugs	0.003**	-0.002	0.003	-0.018***		
	(0.001)	(0.002)	(0.003)	(0.006)		
Phenom & Process	0.001	0.002	0.006*	0.001		
	(0.001)	(0.002)	(0.003)	(0.006)		
Diseases	-0.001	0.000	-0.003	0.001		
	(0.001)	(0.002)	(0.002)	(0.004)		
Anatomy	-0.001	-0.000	-0.003	-0.003		
	(0.001)	(0.002)	(0.003)	(0.006)		
Other Categories	0.001	-0.004	-0.006	-0.007		
	(0.001)	(0.002)	(0.004)	(0.007)		
All MeSH Terms					-0.004	-0.012
					(0.006)	(0.007)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	No	Yes	No	Yes	No
Scientist FE	No	Yes	No	Yes	No	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Scientists	4,830	4,830	4,830	4,830	4,830	4,830
Observations	62,049	62,049	62,049	62,049	62,049	62,049

Notes: All variables and inclusion criteria follow the definitions in Table 5 of the main paper. All results are from linear probability models with cohort fixed effects included in Models 1, 3 and r and scientist fixed effects included in Models 2, 4, and 6. 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. 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 Phenomena & Processes category comprises a range of physiological phenomena and 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.