Sitting on the Fence: Integrating the two worlds of scientific discovery and invention within the firm *

Lia Sheer †

April 28, 2022

Abstract

Applying a within-firm perspective to the topic of the division of innovative labor, I explore the organization of scientific discovery at the firm level - specialized or integrated with invention. Using data on inventors and authors related to U.S. publicly-traded science-performing firms for the period 1980-2015, the paper deepens our understanding of the determinants and the tradeoffs associated with the strategic choice of scientific discovery organization. I show that integration is related to a tradeoff between short-term applied R&D and long-term fundamental R&D; while integration is beneficial for invention, it has adverse effects on its scientific output, which decrease invention in the long run. The negative relationship between integration and publication reduces the direct increase in patents due to integration by approximately 90%. To better understand firms' R&D organizational choice, I present internal and external factors that have implications on the benefits and costs associated with integration: reliance on science, stage of technology, external market for technology, and R&D spillins. Finally, I present consistent implications in terms of market value and show that value creation is related to organizational structure.

Keywords: corporate research, division of labor, innovation, patents, R&D organizational structure

JEL Codes: D22,D23, O31, O32, O34

*Acknowledgments: I wish to thank my dissertation advisors Ashish Arora, Sharon Belenzon, and Wes Cohen, for insightful discussions, valuable comments, and continuous support and encouragement. I also thank Lee Fleming, and Bronwyn Hall for their support. I thank Michael Bikard, Dan Gross, and David Hsu for their valuable comments. I thank Larisa Cioaca, Bernardo Dionisi, Honggi Lee, Jungkyu Suh, and Dror Shvadron for their assistance with the data construction. I thank participants at the CCC Doctoral conference, and seminar participants at EPFL, Bocconi, ESMT, Harvard, UC Berkeley, UPF, Utah University, Purdue, Toronto, Tel-Aviv University, Hebrew University, and Duke for helpful comments and suggestions. All errors are my own.

†Tel Aviv University, liasheer@tauex.tau.ac.il

1 Introduction

Throughout the U.S. corporate history, there has been a constant debate on the "appropriate" organization of scientific discovery[1] within firms. Managers have mainly focused on whether integration between research and development practices is more effective than the specialization of research activity (Wise, 1985; Hounshell and Smith, 1988). This debate goes as far back as the early 20th century, when large firms established central corporate research labs (e.g., DuPont, GE, Xerox-PARC, and AT&T-Bell Laboratories). For example, Dupont had invested in both the "Eastern Lab", where researchers were working side by side with engineers, mainly on applied research related to the firm's immediate product and process improvements needs, as well as in the "Experimental Station" and its Central Research Department (CRD), where scientist initially focused on long-term-fundamental research, separate from the manufacturing and product lines (Hounshell and Smith, 1988).[2] The decline in central corporate research laboratories, starting in the 1980s (Mowery, 2009; Pisano, 2010; Arora et al., 2018, 2020) has only made the topic of the organization of scientific discovery within firms more relevant and important.

The literature presents two main views on the organization of scientific discovery.[3] On the one hand a specialized organization (Smith, 1776), supports a clear division of research and invention practices to increase productivity.[4] On the other hand, an integrated organization[5] supports the connectedness and the continuous interaction between research and invention

[1]Henceforth, I shall use scientific discovery interchangeably with scientific research or simply research.

[2]Dupont's CRD later changed its focus towards more applied research based on the needs of its business units, and accordingly, it was renamed the Central Research and Development Department. It continued to operate until 2016 when DuPont merged with Dow Chemical and decided to downsize and reorganize it as the Science & Innovation group, which was more closely aligned with its business units.

[3]While originally the concept of specialization and the division of innovative labor (Smith, 1776; Arora et al., 1994), was discussed in terms of different firms and organizations specializing in the stages of the innovation process where they have a comparative advantage, this paper takes a within-firm perspective of the topic.

[4]Smith (1776) suggests that productivity is enhanced by increasing dexterity and time saved by avoiding task changing: *"Subdivision of employment in philosophy, as well as in every other business, improves dexterity, and saves time. Each individual becomes more expert in his own peculiar branch, more work is done upon the whole, and the quantity of science is considerably increased by it."* (Smith (1776):12-13)

[5]Integration in this research should not be confused with *Vertical integration*, and with the concept of *Technology integration* (Iansiti, 1997), which relates to the capability of choosing among technological options (i.e., research outputs), and effectively integrating them into an application.

practices to increase productivity (Kline and Rosenberg, 1986; Rosenberg, 1990). As Rosenberg (1990) stated, *"When basic research in industry is isolated from the rest of the firm, whether organizationally or geographically, it is likely to become sterile and unproductive"*. These two views are not mutually exclusive; a firm can decide to implement a combination of both as well as to change its scientific discovery organization over time. For example, in 2014, after many years of clear separation between its research and product units, Microsoft became more integrated. It reassigned half of its research unit to a new group, MSR NEXT, where scientists work alongside engineers on applied projects with immediate impact on Microsoft's products (e.g., the Skype translator) rather than focusing on basic research initiatives (Bloomberg, 2016)[6]. Similarly, Alphabet has its hybrid AI teams (Spector et al., 2012), where researchers work closely with its product groups and focus on more short-term applied initiatives related to Google's current products, as well as DeepMind - a separate specialized AI research unit that focuses on basic research, such as the prediction of the protein folding structure that can assist drug discovery in the long run.[7]

Past research mainly highlighted the benefits of integration of research and development practices to the invention output (Henderson and Cockburn, 1994; Stern, 2004; Gittelman and Kogut, 2003; Bonaccorsi and Thoma, 2007; Breschi and Catalini, 2010; Sauermann and Roach, 2012). Building on Ronald Coase and Oliver Williamson's idea that *"All feasible forms of organization are flawed"*[8],the current paper examines both the benefits and the costs and the determinants associated with the organization of scientific discovery - either specialized or integrated - in a within-firm analysis.

I find that the choice of R&D organizational structure is related to the nature of technology and research in the field. Integration is higher for larger, science-based firms with early-stage technology. Furthermore, integration is positively related to internal use of science in invention

[6]https://www.bloomberg.com/features/2016-microsoft-research

[7]based on Google's research philosophy statement, in recent years, the company shifted from its hybrid research approach to an "open-ended, long-term research, driven more by scientific curiosity than current product needs." (https://research.google/philosophy - last accessed: September 2021)

[8]Interview with Oliver E. Williamson following the announcement of his 2009 Nobel Prize: https://www.nobelprize.org/prizes/economic-sciences/2009/williamson/interview/

- consistent with the idea that integration nurtures the development of scientists as boundary spanners. Finally, integration is positively associated with the availability of external technology but negatively related to external knowledge, which might indicate a weaker substitution between public science and private science.

I demonstrate the short-term benefits of integration in the form of higher invention productivity and the long-term costs of integration in terms of lower research productivity. I find that a one standard deviation increase in integration is related to an increase of 8.2 patents and a decrease of 1.5 scientific papers per firm-year. Building on the view that science is an input to invention (Bush, 1945; Rosenberg, 1990; Narin et al., 1997; Fleming and Sorenson, 2004; Arora et al., 2021a; Marx and Fuegi, 2020), the cumulative decrease in scientific discovery quality will, in turn, dilute the firm's invention quality and breakthroughs in the long-run - resulting in approximately a 90% reduction in the direct increase in patents due to integration. Finally, the organization of scientific discovery also has value implications for the firm. I find that the positive relationship between the firm's market value and its patent stock is stronger with integration, especially for large, science-based firms investing in early-stage technology. Conversely, the relationship between market value and scientific discovery stock decreases with integration, especially for firms with weak external technology sourcing opportunities.

It should be emphasized that the empirical results should not be interpreted as causal. The paper's main goal is to provide evidence about the determinants related to the organization of scientific discovery within firms and to document the empirical patterns related to firm-level innovation outcomes.

The paper contributes to the organizing for innovation literature that examines the relationship between internal organization and innovative output (Kay, 1988; Argyres and Silverman, 2004; Arora et al., 2014; Argyres et al., 2019; Aggarwal et al., 2020).[9] R&D organizational

[9]For example, Argyres and Silverman (2004) and Arora et al. (2014) suggest that R&D structure - centralized or decentralized (measured by resource allocation decision (Argyres and Silverman, 2004), and patents assignment between the headquarter and affiliates (Arora et al., 2014)) - is related to the nature of research the firm undertakes. Firms with decentralized R&D, managed at the business unit level with close interactions between research and production (similar to the concept of integration in the current paper) tend to produce more short-term applied research specific to their products and services. In contrast, firms with a centralized research lab are likely to conduct more general basic scientific research that benefits the firm as a whole. In more

structure, and more specifically, the organization of scientific discovery, is a strategic choice, and the current research documents and quantifies the determinants and tradeoffs associated with this choice. The results documented in this paper imply that managers must understand how to organize scientific discovery while balancing short-term and long-term R&D initiatives, as well as internal and external R&D sources. Furthermore, the paper contributes to the recent line of research on corporate science (Mowery, 2009; Pisano, 2010; Simeth and Cincera, 2016; Arora et al., 2018, 2020, 2021a), by offering an explanation of why conducting corporate basic research remains important and useful for corporate invention.

In terms of data, the paper offers a new and more concrete measure of the organization of scientific research at the firm level. I build on firm-level data, covering 35 years of publications and patents and their link to each other (Arora et al., 2021a,b) to measure the connectedness of research and invention practices within the firm. I compute a firm-level measure of integration: *the share of integrated-authors (of all authors)*. Integrated authors are scientists who perform both research and invention (i.e., both authors and inventors) and collaborate with specialized inventors (i.e., who only patent). To the best of my knowledge, this is the first paper to compile data on inventor-author organization at the firm level across 3.5 decades and for a wide variety of industries. The data enable me to perform cross-industry within-firm analyses and answer relevant questions on how integration relates to firm-level outcomes, which were limited in previous studies that mainly examined co-authorship data at the patent level (e.g., Bonaccorsi and Thoma (2007)), within networks of inventors (e.g., Breschi and Catalini (2010)), and for specific industries and limited years (e.g., Gittelman and Kogut (2003)).

The paper proceeds as follows. Section 2 presents the theory and literature review. Section 3 discusses the data and main measures, Section 4 presents the descriptive statistics and non-parametric evidence, and Section 5 summarizes the econometric results. Section 6 concludes.

recent work, Argyres et al. (2019) find that co-invention networks mediate the relationship between structure and innovative outcomes. That is, they show evidence that suggests, as I argue in the current paper, that co-invention practices within a firm are shaped by the choice of R&D structure.

2 Theory and literature review

2.1 The relationship between scientific discovery and invention within firms

Scientific discoveries and inventions are two distinct worlds (Dasgupta and David, 1994). The former focuses on the general principles and methods, and the latter on commercial application.[10]

The simplest view of the relationship between scientific discovery and invention was the so-called "linear model" associated with Bush (1945), who asserted that technical progress rests upon scientific advance. Over the years, this view was modified to a more complex relationship that includes an interactive connection between scientific discovery and invention - with both "demand-pull" and "discovery-push" as drivers of the technological innovation process (Marquis and Allen, 1966; Kline and Rosenberg, 1986; Rosenberg, 1994).

Within firms, the relationship between science and invention is shaped by the choice of R&D organizational structure. At one extreme, specialization, which imposes a clear division between scientific discovery and invention practices, results in a "discovery-push" research - high fundamental understanding of the underlying principals with no direct technological application (i.e., applying the terminology of Stokes (1997)'s quadrant model: Bohr's pure basic research quadrant). On the other hand, integration, which bridges between scientific discovery and invention practices within the firm, facilitates the transfer of science to invention as well as directs scientific research based on the firm's development needs. Integration thus results in "demand-pull" research with a direct technological application (i.e., either use-inspired basic research (Pasteur's Quadrant) or pure applied research (Edison's Quadrant)).

Therefore, the firm's choice of R&D organizational structure directly impacts the nature of research the firm undertakes and the firm's invention output. Specifically, while integration connects research to the immediate technical needs of the firm and facilitates the transfer of

[10]For the purpose of this paper, we can think of *Scientific Discovery* as research efforts that yield a scientific publication, and of *Invention* as downstream development of an artifact that results in a patent.

science to invention, it is unlikely to result in fundamental research, which is important for significant long-term breakthroughs (Nelson, 1959; Arrow, 1962). As Alphabet-Deepmind's founder, Demis Hassabis, puts it:

"A lot of research in industry is product led ... The problem with that is that you can only get incremental research. [That's] not conducive to doing ambitious, risky research, which, of course, is what you need if you want to make big breakthroughs." (Wired, 2019)[11]

2.2 The short-term benefits of integration

The concept of integration between research and invention practices relates to early work by Thomas Allen and Michael Tushman (Allen, 1966; Allen and Cohen, 1966; Allen, 1969; Tushman, 1977; Allen et al., 1979; Tushman and Katz, 1980). Integration nurtures the development of scientists as boundary spanners as well as gatekeepers - occupying a central position in diffusing knowledge between the firm's functions and between external scientific sources and the firm.[12]

Integration, therefore, gives the firm a competitive advantage to build on internal and external science in its inventions (Cockburn and Henderson, 1998; Singh, 2005). That is, for a given level of investment in internal research, a higher level of integration between science and invention practices increases the firm's absorptive capacity capabilities (Cohen and Levinthal, 1990; Rosenberg, 1990). Building on the view that science acts as a guide to invention, it follows that firms that are better able to use research should generate more or better-quality patents (Fleming and Sorenson, 2004; Arora et al., 2021a). In addition, integration connects research to the immediate technical needs of the firm, and thus, as asserted by Rosenberg (1990), it increases R&D productivity.

More recent work examines the innovative performance of corporate scientists, who both invent and do research (Stern, 2004; Gittelman and Kogut, 2003; Bonaccorsi and Thoma, 2007;

[11] https://www.wired.co.uk/article/deepmind-protein-folding

[12] In related work, Clark et al. (1987) and Holbrook et al. (2000) suggest that firms that apply cross-functional coordination mechanisms outperform those that do not. For example, Fairchild's early success and its significant breakthroughs in the planar process and integrated circuits are attributed mainly to the cross-functional integration between research and production (Holbrook et al., 2000).

Sauermann and Roach, 2012; Motohashi, 2020). Overall, these studies document that firms with integrated author-inventors have higher quality patents.[13] Gittelman and Kogut (2003), for instance, show for a sample of biotechnology firms that the availability of scientists who both publish and patent has more impact on the quality of a patent (examined at the individual patent level using forward patent citations to the focal patent) than the firm's stock of scientific publications. Focusing on the nanotechnology industry, Bonaccorsi and Thoma (2007) further show that corporate patents that include at least one inventor who is also an author are of higher quality than patents with inventors only - including receiving more forward citations and having a broader patent scope.[14] Consistent with these findings, I show in the current paper that higher integration between scientific discovery and invention practices increases downstream invention quantity and quality.

If integration between research and invention is, in fact, fruitful for invention – why don't all firms choose to organize their scientific discovery in such a way? What the aforementioned papers do not examine is the relationship between integration and scientific discovery. In this paper, I address this gap in the literature by examining both the benefits as well as the costs of integration.

2.3 The long-term costs of integration

Previous research sheds some light on the direct costs of integration in terms of reduction in scientific research output. From an opportunity cost perspective, integration draws scientists away from research, which results in lower scientific output. In terms of coordination costs, integration involves researchers teaming up with inventors, two very different groups in their

[13]past work on author-inventor has mainly focused on the individual scientist and was examined at the patent-level as opposed to the firm-level analysis of a firm's choice of integration and firm-level outcomes that I pursue in the current paper. The firm-level analysis provides a more holistic examination that advances our understanding of the link between the organization of scientific discovery and firm-level invention and scientific outcomes

[14]The concept of integration is also related to a line of research that examines the diversity of inventors' knowledge on invention teams and firms' recombinant capabilities (Henderson and Clark, 1990; Fleming, 2001; Singh and Fleming, 2010; Carnabuci and Operti, 2013; Aggarwal et al., 2020; Nagle and Teodoridis, 2020) and their relationship to invention output. For example, Singh and Fleming (2010) show that within-team diversity can trim poor outcomes, which is another mechanism through which hybrid teams of scientists and inventors can contribute to invention quality.

nature (Allen, 1984; Vincenti, 1990). Researchers and inventors use different technical language and coding schemes (Allen, 1984). Furthermore, they have different incentives and reward systems. Researchers are more concerned with non-pecuniary motives such as independence, autonomy in their research agenda, the publication of their work, and their reputation in the broader scientific community (Stern, 2004; Sauermann and Cohen, 2010). Inventors' goals, on the other hand, are more tied to the implementation of their work and their achievements within the firm (Ritti, 1971; Allen, 1984). Inventors' clearly defined research objectives based on development needs might contradict the researcher's basic motives. Integration can, therefore, also decrease the recruitment of researchers who prefer independence and have a taste for basic research (Stern, 2004), which will eventually lead to further deterioration in basic science.[15] Overall, these studies suggest that research output increases when there is a division of labor between research and invention activity within the firm.[16]

In addition to the direct cost of scientific discovery, integration also has an indirect cost in terms of long-term invention output. Internal research can enhance downstream invention both as a direct input to invention (Bush, 1945; Rosenberg, 1990; Narin et al., 1997; Arora et al., 2021a; Marx and Fuegi, 2020) as well as indirectly by guiding invention (Fleming and Sorenson, 2004). By reducing the productivity of research, integration, therefore, indirectly reduces invention in the longer run. Previous research shed some light on this indirect cost (Gambardella, 1992; Henderson and Cockburn, 1994; Li et al., 2017; Poege et al., 2019). For example, Gambardella (1992) shows a positive relationship between the quantity of basic research a pharmaceutical firm performs and the patents it produces. Similarly, examining research programs of major pharmaceutical firms, Henderson and Cockburn (1994) find that firms that encourage corporate scientists to publish scientific research show increased productivity as measured by

[15] While scientists' preferences are not directly considered in this paper, scientists select into organizations that fit their preference in terms of taste for either science or commercialization (Stern, 2004; Sauermann and Roach, 2012). Insofar as more productive scientists are more likely to work in firms with clearer separation between discovery and invention, this is another mechanism through which integration will reduce the scientific capability of the firm.

[16] Bikard et al. (2019) find that academic scientists achieve greater levels of specialization in their basic research when leaving the commercial aspects to their industry partner. This finding echoes the idea of specialization that I present in the current paper.

patent counts.

The Quantum Science Research Lab at Hewlett Packard (HP) demonstrates the importance of a specialized research group to the long-term breakthroughs of the firm. After many years that success at HP was considered as a 100% transfer rate from research to invention, in 1995, while their rivals were getting out of basic science, HP decided to adjust its research organization strategy and added a new specialized fundamental research group to the firm. HP recruited Professor Stanley Williams, at the time at UCLA, to establish the Quantum Science Research Lab. The efforts of this group resulted in several breakthroughs, including HP's 64-bit memory based on molecular switches that was announced in 2002. Furthermore, in 2006 the Small Times magazine evaluated patent portfolios of nanotechnology companies worldwide and awarded HP with the highest rank. In an interview following the award, Professor Williams emphasized the importance of a specialized research group (Cyrus and Williams, 2006):

"Quantum Science Research [lab] was created to get a bit more fundamental and long-term stuff into the HP culture. We were explicitly told that the time horizon was ten years: "You have the responsibility and the privilege of looking out at least ten years into the future and asking what's the fundamental science that we have to do now to establish a technology in that ten-year time frame that will be important at Hewlett-Packard Company"." (The Chemical Heritage Foundation, Interview with Prof R. Stanley Williams, 2006, p.49).

Consistent with these ideas, in the current paper, I find that integrating scientific discovery with invention practices lowers the quantity and quality of publications. I also show that scientific discovery complements the invention process. Taken together, I document the cost of integration in terms of long-term invention output.

2.4 The determinants of integration

The coupling between discovery and invention practices varies within industries and firms over time. To better understand firms' R&D organizational choice, I present internal and external factors that condition the benefits and costs associated with integration: reliance on science in

invention, stage of technology, and external R&D sources.

2.4.1 Internal Factors

Examining the nature of scientific research in the field, Cohen et al. (2020), find that in more applied-engineering fields (where research and invention practices are more connected), the opportunity cost for an academic scientist to be an author-inventor is quite low relative to more fundamental fields (e.g., physical sciences) that require the author-inventor to depart further away from the traditional research in the field. In the current paper, I extend their argument to the level of the firm. I argue that in fundamental science-based fields, where research results are further away from the commercial end, while the opportunity costs for research are high, integration is also more meaningful for invention (i.e., connecting two distinct practices).[17] Consistent with this reasoning, we should expect integration to be higher and increase invention value in more fundamental-based fields with high reliance on science.

Similarly, the benefits from integration are more significant in the early stages of a technology field (Lieberman, 1978; Faulkner, 1994), when technology is more knowledge-led, and there are more back and forth interactions between research and invention. Yet, it is less so in later stages, where inventions are incremental and more likely to result from trial and error than from research. Therefore, for firms with frequent introductions of new technologies, we should expect integration to be beneficial for both scientific discovery and invention value.

2.4.2 External Factors

External sourcing of R&D can substitute for private R&D. Firms can obtain external R&D through (i) markets for technology (MFT) (Arora et al., 2001; Serrano, 2010), (ii) markets for firms (MFF), (iii) mobility of people (Singh and Agrawal, 2011; Tzabbar et al., 2015), and (iv) spillins of science and technology (Arora et al., 2021a; Lucking et al., 2018; Bloom et al., 2013; Jaffe, 1986). The choice of scientific discovery organization is directly related to the availability

[17] In fact, in applied fields, where the practice of science is inseparable from the development (such as in the clinical trial phase in the pharmaceutical and biotech industry), we can expect corporate scientists to be both researchers and inventors in the same mind, such that there is no real need for integration.

of external sourcing opportunities. As external sources rise, the firm can increase integration, even at the cost of lowering internal scientific discovery and compromising on long-term internal breakthroughs. Furthermore, integration increases the absorptive capacity capabilities of the firm to discern external opportunities.[18]

In the empirical analysis, I further distinguish between external technology sources (proxied by markets for technology) and external knowledge sources (proxied by spillins from rivals' science and availability of external non-corporate science). While firms acquire external technology to substitute for internal development, the substitution between knowledge spillins and private science is imperfect. Specifically, public science can either substitute ("crowd-out") or complement ("crowd-in") private science (David et al., 2000; Dimos and Pugh, 2016; Arora et al., 2022a). When private science complements public science, integration would be more costly for firms.

Consistent with this reasoning, we should expect integration to be less costly for scientific discovery value with the availability of external sourcing (especially for external technology sources and less so for external knowledge sources).[19]

Table 1 presents the predicted relationship between integration and the main determinants. Table 2 summarizes how integration is related to main outcomes: scientific discovery, invention, and patent and publication value, and how those relationships vary by main determinants.

[18]In future work, I further explore this relationship between integration and external sourcing and show evidence that integration not only guides the firm's internal search process but also assists its external technology search, which further benefits the firm when external sources rise. This idea also fits with previous research findings that decentralized firms that have a stronger connection between science and invention at the unit level also tend to rely more on external acquisitions, while firms with centralized R&D draw more value from internal R&D (Arora et al., 2014).

[19]one can argue that as the market for technology increases, firms would increase their specialization in scientific discovery to become a seller of technology in the market. As my sample consists of large firms, it seems less likely that they are innovating to sell their inventions (Figueroa and Serrano, 2019). In fact, I find that only 13% of the firms in my sample sell more than 30% of their patents granted throughout the complete sample period.

Table 1: Determinants

VARIABLE	RELIANCE ON SCIENCE	EARLY STAGE TECHNOLOGY	EXTERNAL "D"	EXTERNAL "R"
Integration	Increase	Increase	Increase	TBD

Notes: TBD: to be determined in the empirical analysis.

Table 2: Outcomes

VARIABLE	INVENTION	SCIENTIFIC DISCOVERY	PATENT VALUE	PUBLICATION VALUE
Integration	Increase	Decrease	Increase	Decrease
Integration ×				
Reliance on science			Increase	Decrease
Early-stage tech			Increase	Increase
High External "D"			-	Increase
High External "R"			TBD	TBD

Notes: TBD: to be determined in the empirical analysis.

3 Data and main measures

I combine data from 4 main sources: (i) firm-level data on publications and patents from DISCERN dataset[20] (Arora et al., 2021a,b), (ii) company and accounting information from S&P North American Compustat (Standard & Poor's, 2018), (iii) scientific publications and author information from Web of Science (WoS) (Clarivate Analytics, 2016), and (iv) patent and inventor information from PatStat (European Patent Office, 2016).

3.1 Accounting panel data

DISCERN dataset covers U.S.-headquartered publicly-listed firms and their subsidiaries over the period 1980-2015. For the purpose of this paper, which focuses on the organization of R&D, the sample is restricted to publishing manufacturing firms[21] with at least ten publications during

[20]The data can be downloaded from 10.5281/zenodo.3594642. The version used for the analysis in this paper is version 7.

[21]As in Bloom et al. (2013), manufacturing firms are based on SIC codes in Compustat segment file

the sample period.[22] The final sample includes an unbalanced panel of 1,530 ultimate owner parent companies and 24,878 firm-year observations over the sample period 1980-2015.

3.2 Main variables

Integration: Integration in this research is defined as the connectedness between scientific discovery and invention practices within the firm. While R&D organization structure is difficult to observe and measure at scale, I compile a measure that is correlated with it: *share integrated authors (of all authors)*. Integrated authors are corporate researchers (scientists and engineers) who perform both research and invention (i.e., both authors and inventors) and collaborate with specialized inventors (i.e., who only patent). That is, integration involves collaborations between individuals who work on research and those who specialize in invention.[23] These collaborations could be in the form of cross-functional interactions between separate research and development units[24], as well as within-unit interactions. Furthermore, these collaborations can be imposed by the firm's research organization strategy (e.g., Microsoft's MSR NEXT group), by the firm's organization and reporting structure, by R&D budget allocation decisions and workers' incentive plan (e.g., IBM's incentive plan change in 1989 that emphasizes patenting over scientific publication), as a result of physical co-location of authors and inventors, as well as a consequence of the technology focus of the firm on more applied initiatives. To construct the measure, I employ large-scale automatic matching techniques between the firm's inventors and authors. The Data Appendix B provides details on the data construction efforts.[25]

I pursue several robustness checks to understand better what my measure of integration captures. First, I find that my measure of integration is different from self-use of science in

[22]out of 2,653 firms in the sample with at least one granted publication throughout the sample period, 1,530 firms have at least 10 publications. The restriction on publications is to assure that there is a sufficient amount of publications, such that the measure of integration is meaningful, and the firm is, in fact, likely to face a choice regarding its scientific discovery organization. In a robustness check, I further show that results hold both for small and large firms in my sample.

[23]conditioning on collaborations with specialized inventors is essential for capturing real interactions between science and invention practices and not simply a lone-author or a group of authors who patent their discovery.

[24]Integration, in this sense, is closely related to Lawrence and Lorsch (1967)'s concept of "unity of effort among the various subsystems" and to Clark et al. (1987)'s cross-functional coordination.

[25]The measure is available upon request.

invention (Arora et al., 2021a). Specifically, I find that only 20% of the patents by scientists classified as "integrated" self-cite inventors' science, which supports the idea that integration is an outcome of direct interaction between science and invention practices rather than merely an outcome of researchers choosing to work on applied problems that are subsequently cited by their patents. In Appendix Table A.2 Columns 9-10, I further present robustness checks for integration using a measure of co-location of inventors and authors.[26] When examining the relationship between co-location and publication count at the firm level, I receive a positive and insignificant coefficient estimate for the co-location dummy. This finding reassures that my measure of integration does not simply capture interactions of research and inventions due to the co-location of research and invention activities.[27] Lastly, in section 4.3, I validate my measure of integration using the 1994 Carnegie Mellon Survey (CMS) data (Cohen et al., 2000). I find that my measure is correlated with cross-functional communication within firms.

Appendix Table A.1 illustrates the effectiveness of my measure in capturing changes in the organization of scientific discovery using case studies. Column 1 explores the major change in International Business Machines's (IBM) R&D organization strategy practices in the late 1980s (Gomory, 1989; Bhaskarabhatla and Hegde, 2014). Up to 1989, IBM's research under the academically oriented leadership of Ralph Gomory (1970-1986) and his successor John Armstrong (1986-1989) was mainly focused on basic research separated from development. In 1989, following a change in U.S. patent law in the early 1980s and with the appointment of James McGroddy as director of IBM Research, the company adopted pro-patent IP management practices. Simultaneously, it also shifted its focus towards more applied research initiatives, including joint research and development programs. Column 1 presents the trend in the integration measure for IBM. The sharp increase in integration from an average share of 0.2 pre-1990s to 0.5 post-

[26] to measure co-location, I look at the top city of authors and the top city of inventors for each 5-year cohort and compute a dummy variable that equals one if they are located in the same city.

[27] Appendix Table A.2 presents additional robustness checks for the integration measure. In Columns 1 and 2, integration is computed by excluding scientific publications from new journals post-1990. Results for both patent and publication equations continue to hold, which reassures that my results are not affected by new journals (and perhaps more applied journals) that were added in the latter part of the sample. Similarly, in Columns 3 and 4, integration is computed excluding applied scientific publications with below-median journal impact factor (JIF), and results continue to hold. Columns 5 and 6 include a dummy variable of integration, and Columns 7 and 8 lag the integration measure by five years (1 cohort).

1990s is consistent with the company's documented shift in scientific discovery organization. Furthermore, (Bhaskarabhatla and Hegde, 2014) find that in the decade following this shift, IBM increased its patent applications and decreased its publications, which is consistent with the results I document in this paper for integrated organization structure. Table A.1 Column 2 further explores the effect of Bell-Labs' separation from AT&T CORP in 1996 on integration. It shows that following Bell's acquisition by Lucent, AT&T became more integrated - reflecting the firm's loss of Bell's specialization in basic research.

Market for technology (MFT): I measure market for technology based on patent trading activity in invention classes relevant to the focal firm's patent portfolio (Serrano, 2010; Hochberg et al., 2018; Figueroa and Serrano, 2019; Arora et al., 2022b). Patent transactions are from the USPTO Patent Assignment dataset. [28] For each sample firm's patent portfolio, I compute the probability (averaged across related IPC classes) that a patent related to the focal firm's portfolio will be sold.[29]

Rival Spillins: incoming knowledge flows, SPILLSIC, is the sum of stocks of R&D, patents, and publications by other firms weighted according to the proximity of these firms to the focal firm in the product space.[30]

Relevant external science: Relevant external science stock is a cumulative measure of non-corporate scientific research used by patents in IPC classes relevant to the focal firm's patent portfolio in each cohort.[31] Patent citations are from PatStat and their matched publications

[28]The clean patent transaction data is based on Arora et al. (2022b). Following Figueroa and Serrano (2019), the transactions exclude patents that are reassigned due to pure M&As as well as deals with more than 25 patents transferred. That is, they may include acquisitions of small startups.

[29]Specifically, for each firm i, year t related IPC codes in its patents granted between [t,t-5], I compute the share of external patents sold up to year t out of all patents granted between [t,t-8]. The share is then averaged across IPC classes.

[30]Similar to Bloom et al. (2013) and Arora et al. (2021a), $SPILLSIC\_GRDit$ is the sum of weighted R&D stock by product market rivals and is computed as $\sum_j SEG_{ij} \times GRD_{jt}$. Where GRD_{jt} is the perpetual R&D stock of a potential rival firm j and SEG_{ij}, the Mahalanobis similarity of rival firm j from the focal firm i in the product market. For example, if firm i and firm j have similar sales shares across industry segments, the proximity score of the firms would be high. Industry segment data are from Compustat's operating segments database and are defined by 4-digit SIC codes. $SPILLSIC\_PUBit$ and $SPILLSIC\_PATit$ are computed similarly for publication stock and patent stock, respectively

[31]each paper is counted once per IPC at its journal year. I aggregate cited publications to the focal firm-year level based on the relative share of each IPC in the firm's patent portfolio in a 5-year cohort prior to the focal year.

from WoS.

Reliance on science in invention: I measure reliance on science by citations to external scientific publications located on the front page of a patent.[32] Patent citations are from PatStat.

Early-stage technologies: early-stage technologies are defined as patents granted no more than 15 years from the related IPC inception year. IPC's inception date is from PatStat.

4 Descriptive statistics and non-parametric evidence

The main sample and variables are at the parent company-year level. Table 3 presents descriptive statistics for the main variables over the sample period, 1980-2015. The sample includes a wide distribution of firms in terms of size and R&D investment: market value ranging from 24 million dollars (10th percentile) to 10 billion dollars (90th percentile) and R&D expenditures ranging from 3 million dollars (10th percentile) to 313 million dollars (90th percentile). The sample also varies in terms of R&D employees: authors range from 2 authors (10th percentile) to 240 authors (90th percentile) and inventors from 3 inventors (10th percentile) to 462 inventors (90th percentile). The firms produce, on average, 31 publications and 49 patents per year.

Integration varies across the sample ranging from zero (10th percentile) to 50% (90th percentile), with a mean of 19%, and it tends to be higher in science-based industries (Electronics and Semiconductors- 26%, Pharma 25%, whereas IT and Software 17%). Appendix Figures A1-A2 present trends in integration for main industry groups: (i) IT & Software, (ii) Electronics & Semiconductors, (iii) Telecommunication, (iv) Chemicals, (v) Energy, and (iv) Pharma & Biotech. There is substantial heterogeneity in integration over time by industry. Figure A1 shows an increase in integration in the first three groups (IT & Software, Electronics & Semiconductors, and Telecommunication). For life-science-related industries (Figure A2), integration trend is less clear, with a very similar rate at the end of the sample period as at the

[32] I rely only on citations to external science to make sure that the measure is not directly related to the dependent variable, annual publications.

beginning of the sample period.

4.1 Determinants of integration

Table 4 presents mean comparison tests for integration by above and below the median value of the main determinants of integration. It shows that firm-cohorts with above-median reliance on science in invention (measured by average citations to external science per patent) have a statistically significant higher share of integration (an average of 0.28 for low science-based versus an average of 0.16 for high science-based technologies). Similarly, integration is more prominent in firm-cohorts with early-stage technologies - measured by the share of patents granted no more than 15 years from the related IPC inception year out of all granted patents. When examining external sourcing, I find that one primary determinant of integration is external markets for technology. Integration is statistically significant higher in firms with above-median MFT (an average of 0.19 for low MFT versus an average of 0.25 for high MFT). Integration is also statistically significant higher in firm-cohorts with above-median relevant external public science and above-median spillins of rivals' R&D stock.

Section 5, explores these relationships empirically. Tables 7 and 8 explore the relationship between integration and main determinants. Tables 11 and 12 further show how the above determinants of integration have implications on the relationship between integration and patent and publication value.

4.2 Integration and the tradeoff between scientific discovery and invention

Table 5 presents mean comparison tests for differences in scientific discovery and invention between firms with high and low integration. It shows that firms with above-median integration share also have more and better-quality inventions (an average of 0.47 citation-weighted patents to R&D ($mm) per firm-year for low integration share versus an average of 0.77 citation-

weighted patents to R&D ($mm) for high integration share). Conversely, firms with above-median integration share have a significantly lower rate and quality of scientific discovery (an average of 0.66 citation-weighted publications to R&D ($mm) per firm-year for low integration share versus an average of 0.55 citation-weighted publications to R&D ($mm) for high integration share). These results are consistent with the idea that while integration is beneficial for internal invention, it has adverse effects on scientific discovery quantity and quality.

Table 3: Summary Statistics for Main Variables

				Distribution		
	Obs.	Mean	Std. Dev.	10th	50th	90th
Integration	24,878	0.23	0.21	0	0.19	0.50
Authors	24,878	147	588	2	17	240
Inventors	24,878	229	832	3	35	462
Publications count	24,878	31	129	0	4	46
Publications stock	24,878	154	688	2	15	203
Patent count	24,878	49	200	0	6	94
Patent stock	24,878	225	931	2	25	439
R&D expenditures ($mm)	24,878	186	704	3	26	313
R&D stock ($mm)	24,878	803	3,390	6	91	1,252
Market Value ($mm)	24,878	6,437	30,010	24	465	10,214
Tobin's Q	24,878	5	6	0	2	20
Sales ($mm)	24,878	4,394	16,441	5	414	8,777
Assets ($mm)	24,878	3,299	14,419	4	230	5,904
MFT	24,878	0.06	0.02	0	0.06	0.08
Relevant public science stock	24,878	13,511	24,113	0	3,081	46,683
SPILLSIC, PUB	24,878	12,594	11,241	2,574	9,319	27,432
SPILLSIC, PAT	24,878	13,875	11,694	4,112	10,735	24,845
SPILLSIC, GRD	24,878	51,236	51,818	6,474	32,740	120,521
Citations to ext. science	24,878	158	772	0	7	277
Early stage technology	24,878	0.01	0.07	0.00	0.00	0.00

Notes: This table provides summary statistics for the main variables used in the econometric analysis. The sample is restricted to publishing firms with at least 10 publications during the sample period. The sample is at the firm-year level and includes an unbalanced panel of 1,530 U.S.-headquarter publicly-traded manufacturing ultimate owner parent companies over the sample period 1980-2015.

Table 4: **Mean Comparison: Determinants of Integration**

	(1)	(2)	(3)	(4)	(5)
	Diff. in means	High		Low	
	(3) minus (5)	No. Obs.	Mean	No. Obs.	Mean
Reliance on science in invention	0.12**	2,855	0.28	2,856	0.16
Early stage technology	0.07**	1,149	0.28	4,562	0.21
Market for technology	0.06**	3,408	0.25	2,303	0.19
Relevant external science	0.04**	1,390	0.25	4,321	0.21
SPILLSIC,GRD	0.03**	2,033	0.24	3,678	0.21

Notes: This table presents mean comparison tests for integration by high and low values of different determinants of integration. The unit of analysis is a firm-cohort, and yearly values are averaged over the cohort period. High and low are defined by above and below the median cohort value, respectively. See main text for variable definitions. ** p<0.05

Table 5: **Mean comparison - Integration, invention, and scientific discovery**

	(1)	(2)	(3)	(4)	(5)
	Diff. in means	High Integration		Low Integration	
	(3) minus (5)	No. Obs.	Mean	No. Obs.	Mean
Patent propensity	0.16**	764	0.59	755	0.43
Weighted-patent propensity	0.31**	764	0.77	755	0.47
Publication propensity	-0.19**	764	0.50	755	0.68
Weighted-publication propensity	-0.11*	764	0.55	755	0.66

Notes: This table presents mean comparison tests for firms with high integration vs. firms with low integration. The unit of analysis is a firm, and yearly values are averaged over the sample period. High and low share of integration are defined by above and below the median sample value of average per firm integration, respectively. Patents are weighted by IPC-year normalized forward patent citations. Publications are weighted by journal-year normalized forward publication citations. R&D expenditures are in $mm. * p<0.01 ** p<0.05

4.3 Validating share of integrated authors as a measure of organization of scientific discovery

To validate my measure of integration, I use the 1994 Carnegie Mellon Survey (CMS) data on industrial R&D (Cohen et al., 2000). As part of the survey, lab directors in R&D performing firms were asked about the relationship of their lab with other business functions in the 3 years period prior to the survey. I match the integration measure for the years 1996-2000 to the lagged CMS questions response related to the importance of inter-firm cross-function communication. Of the firms in my sample, 219 are also covered in the CMS (out of 1046 firms that are active in my sample in the relevant cohort).

Table 6 confirms that my measure of integration is related to cross-functional interactions within firms. Specifically, Column 1 shows that integration is related to project teams with cross-functional participation. There is a positive and statistically significant relationship between cross-functional teams and integration.[33]. Column 2 further suggests that integration is positively correlated with above-median communication between R&D units, though the coefficient estimate is significant at the 10 percent level.[34]

[33]based on CMS data Q.6b: "During the last three years, have project teams with cross-functional participation been used to facilitate interaction?". Dummy variable equals one if answered YES

[34]based on CMS data Q.5c: "How frequently do your R&D personnel talk face to face with personal from other R&D units?". Dummy variable equals one for weekly or daily communication

Table 6: Supporting Evidence from Carnegie Mellon Survey

	(1)	(2)
Dependent variable:	Integration	
	cross functional teams	communication between R&D units
Dummy for cross-functional teams	0.161**	
	(0.070)	
Dummy for frequent communication between R&D units		0.121*
		(0.061)
ln(Sales)	0.001	-0.001
	(0.013)	(0.013)
Industry dummies	Yes	Yes
Mean DV	.205	.21
Mean dummy	.947	.685
Number of firms	171	146
R-squared	.61	.61

Notes: This table presents OLS estimation results for the relationship between integration for cohort 4 (1996-2000) and the 1994 Carnegie Mellon survey (CMS) questions response (Cohen et al., 2000) related to the importance of inter-firm cross-function communication. Column 1 is based on CMS data Q.6b: "During the last three years, have project teams with cross-functional participation been used to facilitate interaction?". Column 2 is based on CMS data Q.5c: "How frequently do your R&D personnel talk face to face with personal from other R&D units?". The sample is restricted to survey firms that were matched to our sample. The number of observations varies based on the response rate to each question. Robust standard errors in parentheses. *** p<0.01 ** p<0.05 * p<0.1

5 Econometric analysis

5.1 Integration and determinants

This section tries to decompose the choice of integration by observables. Tables 7 and 8 provide econometric evidence supporting the relationship between integration and main internal and external determinants, respectively.

Table 7 Column 1 shows a positive relationship between the firm's size in terms of R&D investment and integration. Columns 2-3 support the idea that integration is more prominent in fundamental science-based firms. Column 2 shows a positive and statistically significant relationship between reliance on science in invention (measured by annual patent citations to

external scientific publications) and integration. Column 3 further indicates that integration is higher in fundamental fields (i.e., semiconductors, chemical and energy, and life sciences).

Column 4 shows a positive relationship between early-stage technology and integration. A two standard deviation increase in the share of early-stage technology (measured by patents granted no more than 15 years from the related IPC inception year out of all granted patents), will increase integration by 0.01 (a 5% increase relative to the mean value of integration).

Column 5 presents a positive relationship between integration and internal citations by the firm's patents to the firm's own science. Results are consistent with the theorized mechanism that integration nurtures the development of scientists as boundary spanners - diffusing knowledge between the firm's functions and giving the firm a competitive advantage to build on its internal science.

Column 6 indicates that novel patents (where originality is defined based on the uniqueness of technology classes combination reported in the patent) are positively correlated with integration.

Columns 7-8 support the idea that integration is related to a tradeoff between patents and publications. Column 7 shows that top cited patents (restricted to the top 2 percentile of corporate patents by IPC-year weighted forward citations) are positively related to integration, while Column 8 suggests that integration is negatively associated with top publications (restricted to the top 2 percentile of corporate publications by journal-year weighted forward citations).

Taken together, table 7 suggests that the optimal coupling between scientific discovery and invention is related to the nature of technology and research in the field.

Next, Table 8 examines the relationship between external determinants and integration. Column 1 examines the relationship between markets for technology (MFT) and integration. It shows that a one standard deviation increase in MFT will increase integration by 5% relative to the mean value of integration.

Columns 2 and 3 present specifications that control for potential incoming knowledge flows, *spill-ins*, including the stocks of R&D, patents, and publications by other firms weighted

according to the proximity of these firms to the focal firm in the product market. Column 2 regresses rivalry weighted total stock of R&D on integration and finds a statistically negative coefficient estimate. Column 3 further splits rivalry weighted R&D stock to publication stock and patent stock. Results show that science spillins are negatively related to integration while invention spillins are positively related.

Lastly, Column 4 examines the stock of relevant non-corporate external science (scientific research used by patents in IPC classes relevant to the focal firm's patent portfolio) and shows a negative relationship between integration and the availability of relevant public science.

Overall, Table 8 suggests that integration is positively related to the availability of external technology but negatively related to the availability of external knowledge, which might indicate a weaker substitution between public science and private science.

Table 7: Integration and Internal Determinants

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable:	Integration							
	Firm size	Reliance on science in invention	Industry	Age of technology	Citations to own Publications	Novel patents	Top patents	Top publications
$ln(R\&D\ stock)_{t-1}$	0.017**	0.005	0.014**	0.018**	0.014**	0.001	0.016**	0.018**
	(0.004)	(0.004)	(0.001)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
$ln(Citations\ to\ ext\ science)_{t-1}$		0.025**						
		(0.002)						
$IT\ \&\ Telecom$			-0.025**					
			(0.005)					
$Semiconductors$			0.052**					
			(0.004)					
$Chemicals\ \&\ Energy$			0.029**					
			(0.004)					
$Life\ sciences$			0.032**					
			(0.004)					
$Early\ stage\ technology_{t-1}$				0.078**				
				(0.021)				
$ln(Citations\ to\ int\ science)_{t-1}$					0.025**			
					(0.003)			
$ln(Novel)_{t-1}$						0.048**		
						(0.003)		
$ln(Top\ patents)_{t-1}$							0.036**	
							(0.005)	
$ln(Top\ publications)_{t-1}$								-0.010*
								(0.005)
Firm FE	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean DV	.227	.227	.227	.227	.227	.227	.227	.227
Obs.	23,344	23,344	23,348	23,344	23,344	23,344	23,344	23,344
Number of firms	1526	1526	1530	1526	1526	1526	1526	1526
R-squared	.57	.59	.054	.57	.58	.59	.58	.57

Notes: This table presents OLS estimation results examining the relationship between integration and internal determinants. Integration is defined as the share of a firm's authors who both published an article and were granted a collaborative patent with a specialized inventor during a 5-year-cohort period. Classification to main industries is based on SIC codes. See main text for the definition of control variables. One is added to logged control variables. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firms. * p<0.01 ** p<0.05

Table 8: Integration and External Determinants

	(1)	(2)	(3)	(4)
	Integration			
	MFT	Spillins GRD	Spillins PUB/PAT	Relevant external science
MFT_{t-1}	0.598**			
	(0.177)			
$ln(SPILLSIC, GRD)_{t-1}$		-0.062**		
		(0.021)		
$ln(SPILLSIC, PUB)_{t-1}$			-0.110**	
			(0.025)	
$ln(SPILLSIC, PAT)_{t-1}$			0.109**	
			(0.018)	
$ln(Relevant\ external\ sci.\ stock)_{t-1}$				-0.003**
				(0.001)
$ln(R\&D\ stock)_{t-1}$	0.017**	0.022**	0.017**	0.020**
	(0.004)	(0.004)	(0.004)	(0.004)
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Mean DV	.227	.227	.227	.227
Obs.	23,344	23,344	23,344	23,344
Number of firms	1526	1526	1526	1526
R-squared	.57	.58	.58	.57

Notes: This table presents OLS estimation results examining the relationship between integration and external determinants. Integration is defined as the share of a firm's authors who both published an article and were granted a collaborative patent with a specialized inventor during a 5-year-cohort period. Market for technology (MFT) is based on patent trading activity in invention classes relevant to the focal firm's patent portfolio. SPILLSIC is the product market distance weighted sum of all other firms' R&D/Publication/Patent stocks (as appropriate). Relevant external science stock is a cumulative measure of non-corporate scientific research used by patents in IPC classes relevant to the focal firm's patent portfolio. One is added to logged control variables. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firms. * p<0.01 ** p<0.05

5.2 Integration and the tradeoff between scientific discovery and invention

The relationship between scientific discovery and integration is specified as follows:

$$\begin{aligned} \ln(Publications)_{it} =& \beta_0 + \beta_1 Integration_{it-1} + \beta_2 \ln(R\&D\ stock)_{it-1} \\ &+ \boldsymbol{Z}'_{it-1}\boldsymbol{\gamma} + \boldsymbol{\eta_i} + \boldsymbol{\tau_t} + \epsilon_{it} \end{aligned} \tag{1}$$

In Equation 1, $Publications_{it}$ is the annual flow of publications by firm i in year t weighted by the number of forward citations each publication receives divided by the average number of citations received by all other publications published in the same journal-year. Integration is proxied by $Integration_{it-1}$, measured by the share of a firm's authors who both publish and patent out of all authors. $\boldsymbol{Z}_{it-1}$ is a vector of one-year lagged firm-year controls. The coefficient of interest is β_1. Following the prediction in Table 1, I expect $\hat{\beta}_1 < 0$.

The organization of scientific discovery may vary across firms and industries. I thus include firm fixed effects as well as time-varying firm characteristics for scale, such as R&D stock. Furthermore, the temporal structure aims at mitigating concerns that the relationship between the number of yearly publications and integration is merely due to common shocks. Lastly, I report robust standard errors clustered at the firm level to account for possible within-firm error correlation across model years.

Table 9 Columns 1-3 present the estimation results. Column 1 shows results from a pooled specification with four-digit SIC dummies. There is a negative and statistically significant relationship between integration and the number of yearly publications. In Column 2, which adds firm fixed effects to the specification in Column 1, $\hat{\beta}_1$ slightly decreases, indicating that the relationship between integration and publications partly reflects a degree of heterogeneity across firms. Yet, $\hat{\beta}_1$ remains positive, both substantively and statistically: a one standard deviation increase in integration is associated with a 4.62% decrease in yearly publications - approximately 1.5 publications per year.

In Column 3, publications are weighted by the number of forward citations each publication

receives divided by the average number of citations received by all other publications published in the same journal-year. Results confirm that integration does not only decrease the quantity of publications but, more importantly, the quality of publications. Appendix Table A.3 presents additional robustness checks for the relationship between integration and publication.[35]

Whereas previous research has mainly focused on the positive relationship between integration and invention outcomes, Columns 1-3 suggest that integration significantly dilutes the firm's investment in scientific publications.

Next, I estimate a patent production function to assess the hypotheses that both research and integration increase downstream invention and R&D productivity:

$$\begin{aligned}\ln(Patents)_{it} =& \omega_0 + \omega_1 \ln(Publications\ stock)_{it-1} + \omega_2 Integration_{it-1} + \omega_3 \ln(R\&D\ stock)_{it-1} \\ &+ \boldsymbol{Z}'_{it-1}\boldsymbol{\gamma} + \boldsymbol{\eta_i} + \boldsymbol{\tau_t} + \epsilon_{it}\end{aligned} \tag{2}$$

In Equation 2, $Patents_{it}$ is the annual flow of patents weighted by the number of citations each patent receives divided by the average number of citations received by all other patents granted in the same IPC-year. The main variables of interest are $Publications\ stock_{it-1}$ and $Integration_{it-1}$. Other controls include the stock of R&D and author flow, both lagged by one year.

Internal research can enhance downstream invention both as a direct input to invention as well as indirectly by guiding invention. I thus expect firms with more scientific research stock to be more productive ($\hat{\omega_1} > 0$). Following the prediction in Table 1, if integration leads to more downstream invention, I expect $\hat{\omega}_2 > 0$. As shown in Table 9 Columns 4-6, both predictions

[35] Appendix Table A.3 Columns 1 and 2 distinguish between established publicly-traded scientific firms and more recent firms by splitting the sample into firms that entered before and after the year 1990, respectively. To address concerns that publication and patenting patterns might have changed throughout the sample period, in Columns 3 and 4, the panel is divided by firm-years prior and post the year 2000, respectively. To examine variation by firm size, in Columns 5 and 6 the sample is split by below and above-median sales, respectively. The level of integration the firm chooses varies by the nature of technology in the field. Columns 7 and 8 divide the firm sample based on below and above-median reliance on science in invention, respectively. Results indicate that the observed relationship is driven mostly by fundamental science-based technology. To address concerns that the results are driven by life science industry, Column 9 excludes life-science firms based on related SIC codes. Lastly, to confirm that results are robust to a change in specification, Column 10 presents results with Inverse hyperbolic sine transformation, and Column 11 presents a between-firm specification. Results are robust to all specifications.

are confirmed in the data.

Table 9 Column 4 presents results from a pooled specification with four-digit SIC dummies. There is a positive and statistically significant relationship between the count of yearly patents and both integration and publication stock. Column 5 presents the same pattern of results for a within-firm specification: a one standard deviation increase in integration is associated with a 16% increase in yearly patents - approximately 8.2 patents per year. Moreover, the marginal effect of an additional publication, evaluated at the sample mean, is equal to approximately 5 patents. The result supports the idea that scientific discovery complements the innovation process.

Integration increases not only the quantity of inventions but also the quality of inventions. Column 6 confirms that results hold when patents are weighted by IPC-year normalized forward patent citations and using a citation-weighted publication stock.

Taken together, the results in Columns 1-6 support the conjecture that integration is related to a tradeoff between short-term and long-term R&D initiatives - because integration is negatively related to publication (as illustrated in Columns 1-3), which in turn will have an adverse effect on inventions (as presented in Columns 4-6). The direct increase in patents from a one standard deviation increase in integration drops by approximately 90% (from 8.2 patents to 0.7 patents)[36] due to the indirect negative relationship between integration and publications.

5.3 Integration and market value

If integration is positively related to invention quality and negatively related to scientific discovery quality, it should be reflected not only in the level of patent and publication output but, more importantly, in the firm's value. Next, I examine the relationship between integration and firm stock market value and estimate the following Tobin's Q specification following Griliches (1986) and Hall et al. (2005) as well as more recent work by Simeth and Cincera (2016) and Arora et al. (2021a).[37]

[36] the net decrease equals to $8.2 - 1.5 \times 5 = 0.7$ patents

[37] Market value is the sum of common stock, preferred stock, and total debt net of current assets. Tobin's Q is market value over assets

$$
\begin{aligned}
\ln \frac{Value_{it}}{Assets_{it}} =& \alpha_0 \frac{G_{it-1}}{Assets_{it}} + \alpha_1 Integration_{it-1} * \frac{\ln(Publication\ stock)_{it-1}}{Assets_{it}} \\
&+ \alpha_2 Integration_{it-1} * \frac{\ln(Patent\ stock)_{it-1}}{Assets_{it}} \\
&+ \alpha_3 Integration_{it-1} + \boldsymbol{Z}'_{it-1}\boldsymbol{\gamma} + \boldsymbol{\eta_i} + \boldsymbol{\tau_t} + \epsilon_{it}
\end{aligned} \tag{3}
$$

In Equation 3, G is knowledge assets - measured as the perpetual stocks of publications and patents. The main interest is at coefficients α_1 and α_2, which estimate the interaction between integration and both publication stock and patent stock, respectively.

Consistent with the publication and patent equations results, I expect the positive relationship between the firm's market value and its stock of inventions to be stronger with integration. Conversely, the positive relationship between market value and scientific knowledge stock should be weaker with integration. Thus, $\hat{\alpha_1} < 0$ and $\hat{\alpha_2} > 0$.

Table 10 presents the estimation results. Building on Simeth and Cincera (2016) and Arora et al. (2021a), Column 1 shows the break up into publication and patent stocks, indicating a positive value for both patents and publication stocks. Column 2 adds the interaction between integration and citation-weighted patents and publication stocks. As expected, the coefficient estimate of the interaction with publication stock is negative ($\hat{\alpha_1} < 0$), and the estimate of interaction with patent stock is positive $\hat{\alpha_2} > 0$. Both estimates are statistically different from zero. In Column 3, the relationship endures even after controlling for firm fixed effects. Lastly, Column 4 presents a within-firm estimation result for a market value specification as a robustness check. Results hold, and both interaction estimates are statistically different from zero.

Overall, the results in Table 10 are consistent with the idea that value creation is conditioned by organizational structure (Arora et al., 2014). In particular, the private value of publications decreases, and the private value of patents increases when firms integrate scientific discovery with invention.

5.4 The determinants of integration and market value

Following the prediction in Table 2, the relationship between market value and scientific knowledge and invention is expected to vary based on factors that condition the benefits and costs associated with integration. Tables 11 and 12 examine how the firm's value varies based on main internal and external determinants of integration, respectively.

Table 11 presents the estimation results for internal determinants of integration and Tobin's Q. I start by examining the size of the firm. Columns 1 and 2 divide the firm sample based on below and above-median R&D stock. Results show that both the benefit of integration (in terms of patent value) and the cost of integration (in terms of publication value) are stronger for larger firms.

Columns 3-6 further explore determinants that are related to the nature of technology in the field. Columns 3 and 4 divide the firm sample based on below and above-median use of external science in invention, respectively. The results are consistent with the idea that the opportunity cost of integration for fundamental science-based firms is high (i.e., integration requires scientists to depart further away from the traditional research in their field). Yet, the benefits of integration, as reflected in the private value of patents, are also higher (i.e., by connecting two distinct practices) ($\hat{\alpha_2}$ is 0.09 for science-based firms, while the estimated coefficient for low reliance on science is statistically zero).

Next, I examine the stage of technology in the field. Columns 5 and 6 divide the sample based on the median share of investments in early-stage technology in each cohort. The results are consistent with the idea that the benefit of integration in terms of patent value is stronger in early-stage technology, as invention is more tightly connected to scientific research ($\hat{\alpha_2}$ is 0.194 for early-stage technology, while the estimated coefficient for mature technology is statistically zero). The results further suggest that the opportunity cost of integration in terms of publication value is lower in the early stages - supporting the importance of the back and forth relationship between research and invention in the early life cycle of the technology ($\hat{\alpha_1}$ is -0.098 for mature technology, while the estimated coefficient for early-stage technology is statistically zero).

Next, Table 12 examines how the external determinants of integration change the private value of integrated publication and patent stock. First, I explore how the market for technology(MFT) conditions the results. Columns 1 and 2 divide the sample into two groups based on the median value of MFT in each cohort. Following the prediction in Table 2, the results indicate that the cost of integration in terms of scientific discovery value is prominent for low MFT ($\hat{\alpha_1}$ is -0.155 for low MFT, while the coefficient estimate for high MFT is statistically zero). This result is consistent with the idea that when MFT is low, integration becomes more costly as firms rely on internal scientific discovery for long-term breakthroughs.

Similarly, Columns 3-6 examine how the value differs with the availability of knowledge spillins. Columns 3 and 4 are divided by below and above the median value of spillins from rivals' publications, respectively. Columns 5 and 6 are split by below and above the median value of relevant non-corporate external science stock, respectively. Results for rival spillins are similar to the MFT results in Columns 1-2. The results further suggest that for low rival spillins, the private value of patents increases with integration - that is, as external scientific opportunities decrease, integration is more valuable in guiding the firm's internal invention search process. Results for scientific spillins from relevant non-corporate science are less clear, indicating a weaker substitution between public science and private science.

Overall, Tables 11 and 12 suggest that firms relying on internal scientific discovery for value creation should find specialization more compatible with their objective. In contrast, larger firms that focus on more science-based, early-stage technology initiatives, possibly combined with external technology sourcing, are better served by a blend of integrated and specialized structure - correctly balancing applied and fundamental R&D initiatives.

Table 9: Integration and the Tradeoff between Science and Invention

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	ln(Publication count)			ln(Patent count)		
	Pooled	Firm FE	Citation weighted	Pooled	Firm FE	Citation weighted
$Integration_{t-1}$	-0.254**	-0.208**	-0.170**	1.165**	0.966**	0.952**
	(0.022)	(0.038)	(0.049)	(0.030)	(0.066)	(0.074)
$ln(Publication\ stock)_{t-1}$				0.049**	0.096**	
				(0.009)	(0.029)	
$ln(Weighted\ publication\ stock)_{t-1}$						0.078**
						(0.021)
$ln(R\&D\ stock)_{t-1}$	0.080**	0.084**	0.094**	0.401**	0.258**	0.244**
	(0.004)	(0.011)	(0.015)	(0.006)	(0.030)	(0.027)
$ln(Authors)_{t-1}$	0.730**	0.508**	0.468**	0.295**	0.198**	0.201**
	(0.004)	(0.012)	(0.016)	(0.008)	(0.015)	(0.017)
Firm FE	No	Yes	Yes	No	Yes	Yes
Industry dummies	Yes	-	-	Yes	-	-
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Mean DV	31.879	31.882	40.249	50.701	50.708	58.702
Obs.	23,348	23,344	23,344	23,348	23,344	23,344
Number of firms	1530	1526	1526	1530	1526	1526
R-squared	.84	.89	.82	.77	.87	.84

Notes: This table presents OLS estimation results for the relationship between integration and annual publications and patents. Integration is defined as the share of a firm's authors who both published an article and were granted a collaborative patent with a specialized inventor during a 5-year-cohort period. One is added to logged variables. All specifications include lagged dummies for zero publications &/or patents per year(as appropriate). Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firms. ** p<0.01 * p<0.05

Table 10: Integration and Market Value

	(1)	(2)	(3)	(4)
Dependent variable:	ln(Tobin's Q)			ln(Market Value)
	Base	Pooled	firm FE	firm FE
$Integration_{t-1}$ ×				
$Pub\ stock_{t-1}$ / $Assets$		-0.132**	-0.095**	
		(0.019)	(0.020)	
$Pat\ stock_{t-1}$ / $Assets$		0.102**	0.048*	
		(0.018)	(0.020)	
$ln(Pub\ stock)_{t-1}$				-0.119*
				(0.052)
$ln(Pat\ stock)_{t-1}$				0.133**
				(0.048)
$Integration_{t-1}$		0.176**	-0.123**	-0.268
		(0.040)	(0.043)	(0.190)
$Patent\ stock_{t-1}$ / $Assets$	0.044**	0.008	0.016	
	(0.004)	(0.007)	(0.008)	
$Publication\ stock_{t-1}$ / $Assets$	0.024**	0.041**	0.024**	
	(0.005)	(0.007)	(0.008)	
GRD_{t-1} / $Assets$	0.094**	0.077**	0.051**	
	(0.005)	(0.005)	(0.007)	
$Authors_{t-1}$ / $Assets$		0.054**	0.063**	
		(0.005)	(0.005)	
$ln(Weighted\ patent\ stock)_{t-1}$				-0.075**
				(0.027)
$ln(Weighted\ publication\ stock)_{t-1}$				-0.067*
				(0.028)
$ln(R\&D\ stock)_{t-1}$				0.025
				(0.032)
$ln(Authors)_{t-1}$				0.133**
				(0.019)
$ln(Assets)_{t-1}$				0.361**
				(0.022)
Industry dummies	Yes	Yes	-	-
Firm FE	No	No	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Mean DV	4.712	4.712	4.712	6775.582
Obs.	21,822	21,822	21,813	21,804
Number of firms	1518	1518	1509	1509
R-squared	.52	.52	.71	.87

Notes: This table presents OLS estimation results for the relationship between integration and value. Tobin's-Q is the ratio of market value to assets. Integration is defined as the share of a firm's authors who both published an article and were granted a collaborative patent with a specialized inventor during a 5-year-cohort period. Patents are weighted by IPC-year normalized forward patent citations. Publications are weighted by journal-year normalized forward publication citations. One is added to logged control variables. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firms. ** p<0.01 * p<0.05

Table 11: Internal determinants of Integration and Market Value

	(1)	(2)	(3)	(4)	(5)	(6)
	ln(Tobin's Q)					
Dependent variable:	Small firms	Large firms	Low Reliance on science	High Reliance on science	Mature technology	Early stage technology
$Integration_{t-1}$ ×						
$Pub\ stock_{t-1}$ / $Assets$	-0.073*	-0.146*	-0.087*	-0.099*	-0.098**	-0.078
	(0.033)	(0.059)	(0.039)	(0.042)	(0.033)	(0.091)
$Pat\ stock_{t-1}$ / $Assets$	0.025	0.126*	0.008	0.090*	0.039	0.194*
	(0.033)	(0.061)	(0.038)	(0.042)	(0.032)	(0.091)
$Integration_{t-1}$	-0.116	-0.126	-0.032	-0.263*	-0.113	-0.298
	(0.096)	(0.114)	(0.093)	(0.121)	(0.079)	(0.234)
$Patent\ stock_{t-1}$ / $Assets$	0.017	-0.000	0.017	0.013	0.013	-0.002
	(0.015)	(0.023)	(0.018)	(0.017)	(0.014)	(0.030)
$Publication\ stock_{t-1}$ / $Assets$	0.009	0.057*	0.038*	0.010	0.020	0.034
	(0.015)	(0.025)	(0.019)	(0.018)	(0.015)	(0.032)
GRD_{t-1} / $Assets$	0.053**	0.052**	0.051**	0.055**	0.060**	-0.001
	(0.012)	(0.018)	(0.016)	(0.013)	(0.011)	(0.025)
$Authors_{t-1}$ / $Assets$	0.070**	0.046**	0.047**	0.073**	0.064**	0.080**
	(0.010)	(0.013)	(0.012)	(0.011)	(0.009)	(0.024)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Mean DV	5.589	3.891	3.67	5.712	4.974	3.814
Obs.	10,548	11,265	10,677	11,136	16,829	4,952
Number of firms	857	652	743	766	1375	534
R-squared	.73	.69	.68	.72	.71	.78

Notes: This table presents OLS estimation results for the cross partial relationship between value and, integration and various internal determinants. Tobin's-Q is the ratio of market value to assets. Columns 3 and 4 are classified by below and above median use of external science in invention, respectively. Patents are weighted by IPC-year normalized forward patent citations. Publications are weighted by journal-year normalized forward publication citations. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firms. ** p<0.01 * p<0.05

Table 12: External determinants of Integration and Market Value

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	ln(Tobin's Q)					
	Low MFT	High MFT	Low Spillins	High Spillins	Low relevant ext. science	High relevant ext. science
$Integration_{t-1}$ ×						
$Pub\ stock_{t-1}$ / $Assets$	-0.155*	-0.060	-0.182**	-0.065	-0.147*	-0.072*
	(0.062)	(0.035)	(0.070)	(0.034)	(0.068)	(0.032)
$Pat\ stock_{t-1}$ / $Assets$	0.038	0.043	0.120*	-0.003	0.010	0.045
	(0.052)	(0.035)	(0.051)	(0.037)	(0.058)	(0.031)
$Integration_{t-1}$	-0.229	-0.089	-0.150	-0.130	-0.025	-0.157
	(0.130)	(0.099)	(0.095)	(0.120)	(0.104)	(0.118)
$Patent\ stock_{t-1}$ / $Assets$	0.021	0.013	0.001	0.036*	0.038	0.025
	(0.023)	(0.014)	(0.027)	(0.014)	(0.027)	(0.013)
$Publication\ stock_{t-1}$ / $Assets$	0.045	0.018	0.057*	0.015	0.047	0.020
	(0.024)	(0.015)	(0.028)	(0.015)	(0.028)	(0.014)
GRD_{t-1} / $Assets$	0.050**	0.051**	0.026	0.063**	0.014	0.075**
	(0.019)	(0.012)	(0.025)	(0.011)	(0.021)	(0.012)
$Authors_{t-1}$ / $Assets$	0.066**	0.059**	0.086**	0.049**	0.083**	0.042**
	(0.013)	(0.010)	(0.022)	(0.008)	(0.018)	(0.009)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Mean DV	3.644	5.732	2.777	6.611	3.042	6.107
Obs.	10,706	11,041	10,807	10,987	10,112	11,555
Number of firms	1153	1126	766	936	1138	1099
R-squared	.71	.75	.65	.72	.73	.71

Notes: This table presents OLS estimation results for the cross partial relationship between value and, integration for various external determinants. Tobin's-Q is the ratio of market value to assets. Market for technology (MFT) is based on patent trading activity in invention classes relevant to the focal firm's patent portfolio. Columns 1 and 2 are classified by below and above the median value of MFT, respectively. SPILLSIC is the product market distance weighted sum of all other firms' Publications. Columns 3 and 4 are classified by below and above the mean value of SPILLSIC-PUB, respectively. Relevant external science stock is a cumulative measure of non-corporate scientific research used by patents in IPC classes relevant to the focal firm's patent portfolio. Columns 5 and 6 are classified by below and above the median value of relevant external science stock, respectively. Patents are weighted by IPC-year normalized forward patent citations. Publications are weighted by journal-year normalized forward publication citations. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firms. ** p<0.01 * p<0.05

6 Conclusion and discussion

Firms differ in how they choose to organize scientific discovery. Using a novel firm-level measure of integration, I bridge two streams of literature - innovation and organization, to examine the organization of scientific discovery at the firm level. I find that the choice of R&D organization is related to the nature of technology and research in the field. While past research mainly highlighted the benefits of integration, I examine both its benefits and costs in the current paper. I show that while integration benefits the firm's short-term invention search process, specialization supports the long-term fundamental R&D initiatives. This forward-looking view is also captured in a stock-market valuation analysis of patent and publication stocks.

The results imply that firms might degrade their scientific capabilities if they tie their internal science too tightly to the firm's short-term invention needs, which eventually leads to the deterioration of long-run invention quality. Though specialization is important for long-term significant breakthroughs, firms cannot immediately appropriate the benefits of their investment in basic science (Nelson, 1959; Arrow, 1962). Thus, it might be the case that the tradeoff that I document in this paper is not immediately apparent to firms. If the feedback loop from basic scientific discovery to invention is only manifested in the long-term, firms might over-integrate or sequentially jump from one organization structure to another (Hounshell and Smith, 1988).

The results documented in this paper support a more balanced view between applied and fundamental R&D initiatives. In fact, it has been argued that the creation of Nylon at Dupont would not have been possible without the right blend of integrated research connected to the immediate needs of the product units, the view led by the research director at that time, Elmer K. Bolton, and that of the specialized basic research, led by Charles Stine (Davila et al., 2006). This balanced view is also embedded in the concept of ambidextrous organizations that pursue both exploration and exploitation (March, 1991; Tushman and O'Reilly III, 1996; Lavie et al., 2010; O'Reilly III and Tushman, 2013; Stettner and Lavie, 2014).[38] Future work can

[38]The research on ambidexterity proposed several approaches for balancing exploration and exploitation, including, simultaneously engaging in exploration and exploitation, temporal sequential engagement, and balance of exploration and exploitation across different organization modes.

design optimal structure for scientific discovery organization that would maximize the impact of corporate science on invention, and at the same time, protect its long-term properties.

I further present three main determinants that condition this tradeoff: reliance on science, stage of technology, and external R&D sources. Results suggest that in the presence of external technology sources, integration becomes less costly for firms. Over the years, there has been an increase in the availability of external R&D sources by small firms and universities (Pisano, 2010; Arora et al., 2020). There has also been a change in the research focus of universities towards more applied fields that might affect the preferences of scientists who join the industry and the availability of external technology sources. Furthermore, the nature of research in several fields has changed (e.g., the emergence of bio-science and biotechnology). These changes may partly explain the trends and relationships I document for integration. Future research should explore how these and other changes affect trends in the organization of scientific discovery.

I acknowledge the limitation of the paper in terms of relying on patent and scientific publications data. First, as acknowledged in the innovation literature, using patent data as a proxy of invention is not without problems. For example, some firms may choose to keep their inventions as trade secrets, and there is also variation in the use of patents across industries (Cohen et al., 2000). Furthermore, patents are also different from commercial success.

Second, these data are only sufficient for identifying instances of integration where an author-inventor successfully files patents and publishes scientific papers. As previously mentioned, these instances vary over time and across industries and technology fields. To mitigate this concern, my measure of integration does not rely on the quantity of publications but rather identifies an author as an individual who had at least one publication during a 5-year cohort period.

Furthermore, an obvious limitation of the paper is the lack of causality claim. Integration and investment research and invention can be affected by a common unobserved variable, which would bias the OLS estimation. For example, if a change in scientific opportunities affects both the rate of investment in scientific discovery and the collaboration opportunities between researchers and inventors, my OLS estimates of the relationship between integration

and publications would be upward biased. Another concern is that changes in firm strategy will affect both integration and investment in science. For example, a firm might be more inclined toward integration if it intends to focus on more applied invention in the future. This could coincide with the firm's decision to invest less in science, which would mean that the observed correlation between integration and research output is not causal. Nonetheless, the tradeoff I document may be informative for long-term strategies and orientation of firms. Future work should develop this idea further.

References

Aggarwal, V. A., Hsu, D. H., and Wu, A. (2020). Organizing knowledge production teams within firms for innovation. Strategy Science, 5(1):1–16.

Allen, T. J. (1966). Managing the flow of scientific and technological information. PhD thesis, Massachusetts Institute of Technology.

Allen, T. J. (1969). Roles in technical communication networks.

Allen, T. J. (1984). Managing the flow of technology: Technology transfer and the dissemination of technological information within the r&d organization. MIT Press Books.

Allen, T. J. and Cohen, S. I. (1966). Information flow in an r & d laboratory.

Allen, T. J., Tushman, M. L., and Lee, D. M. (1979). Technology transfer as a function of position in the spectrum from research through development to technical services. Academy of management journal, 22(4):694–708.

Argyres, N., Rios, L. A., and Silverman, B. S. (2019). Organizational change and the dynamics of innovation: Formal r&d structure and intrafirm inventor networks. Strategic Management Journal.

Argyres, N. S. and Silverman, B. S. (2004). R&d, organization structure, and the development of corporate technological knowledge. Strategic Management Journal, 25(8-9):929–958.

Arora, A., Belenzon, S., Cioaca, L., and Sheer, L. (2022a). The effect of public science on corporate r&d. Working Paper.

Arora, A., Belenzon, S., and Patacconi, A. (2018). The decline of science in corporate r&d. Strategic Management Journal, 39(1):3–32.

Arora, A., Belenzon, S., Patacconi, A., and Suh, J. (2020). The changing structure of american innovation: Some cautionary remarks for economic growth. Innovation Policy and the Economy, 20(1):39–93.

Arora, A., Belenzon, S., and Rios, L. A. (2014). Make, buy, organize: The interplay between research, external knowledge, and firm structure. Strategic Management Journal, 35(3):317–337.

Arora, A., Belenzon, S., and Sheer, L. (2021a). Knowledge spillovers and corporate investment in scientific research. American Economic Review, 111(3):1–28.

Arora, A., Belenzon, S., and Sheer, L. (2021b). Matching patents to compustat firms, 1980-2015: Dynamic reassignment, name changes, and ownership structures. Research Policy, 50(5):104217.

Arora, A., Belenzon, S., and Suh, J. (2022b). Science and the market for technology. Management Science.

Arora, A., Fosfuri, A., and Gambardella, A. (2001). Markets for technology and their implications for corporate strategy. Industrial and corporate change, 10(2):419–451.

Arora, A., Gambardella, A., et al. (1994). The changing technology of technological change: general and abstract knowledge and the division of innovative labour. Research policy, 23(5):523–532.

Arrow, K. J. (1962). Economic welfare and the allocation of resources for invention.

Bhaskarabhatla, A. and Hegde, D. (2014). An organizational perspective on patenting and open innovation. Organization Science, 25(6):1744–1763.

Bikard, M., Vakili, K., and Teodoridis, F. (2019). When collaboration bridges institutions: The impact of university–industry collaboration on academic productivity. Organization Science, 30(2):426–445.

Bloom, N., Schankerman, M., and Van Reenen, J. (2013). Identifying technology spillovers and product market rivalry. Econometrica, 81(4):1347–1393.

Bonaccorsi, A. and Thoma, G. (2007). Institutional complementarity and inventive performance in nano science and technology. Research policy, 36(6):813–831.

Breschi, S. and Catalini, C. (2010). Tracing the links between science and technology: An exploratory analysis of scientists' and inventors' networks. Research Policy, 39(1):14–26.

Bush, V. (1945). Science: The endless frontier. Transactions of the Kansas Academy of Science (1903-1945), 48(3):231–264.

Carnabuci, G. and Operti, E. (2013). Where do firms' recombinant capabilities come from? intraorganizational networks, knowledge, and firms' ability to innovate through technological recombination. Strategic management journal, 34(13):1591–1613.

Clarivate Analytics (2016). Web of science (wos) core collection xml, 1900-2016. Clarivate Analytics, Philadelphia, PA. Obtained from Clarivate Analytics by license in 2017.

Clark, K. B., Chew, W. B., Fujimoto, T., Meyer, J., and Scherer, F. (1987). Product development in the world auto industry. Brookings Papers on economic activity, 1987(3):729–781.

Cockburn, I. M. and Henderson, R. M. (1998). Absorptive capacity, coauthoring behavior, and the organization of research in drug discovery. The Journal of Industrial Economics, 46(2):157–182.

Cohen, W. M. and Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. Administrative science quarterly, pages 128–152.

Cohen, W. M., Nelson, R. R., and Walsh, J. P. (2000). Protecting their intellectual assets: Appropriability conditions and why us manufacturing firms patent (or not). Technical report, National Bureau of Economic Research.

Cohen, W. M., Sauermann, H., and Stephan, P. (2020). Not in the job description: The commercial activities of academic scientists and engineers. Management Science.

Cyrus, M. R. and Williams, S. R. (2006). R. stanley williams [interview]. Chemical Heritage Foundation, pages 1–64.

Dasgupta, P. and David, P. A. (1994). Toward a new economics of science. Research policy, 23(5):487–521.

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

Davila, T., Epstein, M. J., and Shelton, R. D. (2006). The creative enterprise: Managing innovative organizations and people. Greenwood Publishing Group.

Dimos, C. and Pugh, G. (2016). The effectiveness of R&D subsidies: A meta-regression analysis of the evaluation literature. Research Policy, 45(4):797–815.

European Patent Office (2016). The epo worldwide patent statistical database, 1980-2015. PATSTAT Global single edition 2016. Obtained from EPO by license in 2016.

Faulkner, W. (1994). Conceptualizing knowledge used in innovation: A second look at the science-technology distinction and industrial innovation. Science, technology, & human values, 19(4):425–458.

Figueroa, N. and Serrano, C. J. (2019). Patent trading flows of small and large firms. Research Policy, 48(7):1601–1616.

Fleming, L. (2001). Recombinant uncertainty in technological search. Management science, 47(1):117–132.

Fleming, L. and Sorenson, O. (2004). Science as a map in technological search. Strategic management journal, 25(8-9):909–928.

Gambardella, A. (1992). Competitive advantages from in-house scientific research: The us pharmaceutical industry in the 1980s. Research policy, 21(5):391–407.

Gittelman, M. and Kogut, B. (2003). Does good science lead to valuable knowledge? biotechnology firms and the evolutionary logic of citation patterns. Management Science, 49(4):366–382.

Gomory, R. E. (1989). Moving ibm's technology from research to development. Research-Technology Management, 32(6):27–32.

Griliches, Z. (1986). Productivity, R and D, and basic research at the firm level in the 1970's. American Economic Review, 76(1):141–154.

Hall, B. H., Jaffe, A., and Trajtenberg, M. (2005). Market value and patent citations. RAND Journal of Economics, pages 16–38.

Henderson, R. and Cockburn, I. (1994). Measuring competence? exploring firm effects in pharmaceutical research. Strategic management journal, 15(S1):63–84.

Henderson, R. M. and Clark, K. B. (1990). Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. Administrative science quarterly, pages 9–30.

Hochberg, Y. V., Serrano, C. J., and Ziedonis, R. H. (2018). Patent collateral, investor commitment, and the market for venture lending. Journal of Financial Economics, 130(1):74–94.

Holbrook, D., Cohen, W. M., Hounshell, D. A., and Klepper, S. (2000). The nature, sources, and consequences of firm differences in the early history of the semiconductor industry. Strategic Management Journal, 21(10-11):1017–1041.

Hounshell, D. A. and Smith, J. K. (1988). Science and Corporate Strategy: Du Pont R and D, 1902-1980. Cambridge University Press.

Iansiti, M. (1997). Technology integration: making critical choices in a dynamic world. Harvard Business school press.

Jaffe, A. B. (1986). Technological opportunity and spillovers of R&D: evidence from firms' patents, profits and market value. NBER Working Papers: No. 1815.

Kay, N. (1988). The r&d function: corporate strategy and structure. Technical change and economic theory, pages 283–294.

Kline, S. J. and Rosenberg, N. (1986). An overview of innovation. In The positive sum strategy: Harnessing technology for economic growth, page 640.

Lavie, D., Stettner, U., and Tushman, M. L. (2010). Exploration and exploitation within and across organizations. Academy of Management annals, 4(1):109–155.

Lawrence, P. R. and Lorsch, J. W. (1967). Differentiation and integration in complex organizations. Administrative science quarterly, pages 1–47.

Li, D., Azoulay, P., and Sampat, B. N. (2017). The applied value of public investments in biomedical research. Science, 356(6333):78–81.

Lieberman, M. B. (1978). A literature citation study of science-technology coupling in electronics. Proceedings of the IEEE, 66(1):5–13.

Lucking, B., Bloom, N., and Van Reenen, J. (2018). Have R&D spillovers changed? Technical report, National Bureau of Economic Research.

March, J. G. (1991). Exploration and exploitation in organizational learning. Organization science, 2(1):71–87.

Marquis, D. G. and Allen, T. J. (1966). Communication patterns in applied technology. American Psychologist, 21(11):1052.

Marx, M. and Fuegi, A. (2020). Reliance on science: Worldwide front-page patent citations to scientific articles. Strategic Management Journal, 41(9):1572–1594.

Motohashi, K. (2020). Science and technology co-evolution in ai: Empirical understanding through a linked dataset of scientific articles and patents. Science and Technology.

Mowery, D. C. (2009). Plus ca change: Industrial R&D in the "third industrial revolution". Industrial and Corporate Change, 18(1):1–50.

Nagle, F. and Teodoridis, F. (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.

Narin, F., Hamilton, K. S., and Olivastro, D. (1997). The increasing linkage between us technology and public science. Research policy, 26(3):317–330.

Nelson, R. R. (1959). The simple economics of basic scientific research. Journal of political economy, 67(3):297–306.

O'Reilly III, C. A. and Tushman, M. L. (2013). Organizational ambidexterity: Past, present, and future. Academy of management Perspectives, 27(4):324–338.

Pisano, G. P. (2010). The evolution of science-based business: innovating how we innovate. Industrial and corporate change, 19(2):465–482.

Poege, F., Harhoff, D., Gaessler, F., and Baruffaldi, S. (2019). Science quality and the value of inventions. Science advances, 5(12):eaay7323.

Ritti, R. R. (1971). The engineer in the industrial corporation. Columbia University Press New York.

Rosenberg, N. (1990). Why do firms do basic research (with their own money)? Research policy, 19(2):165–174.

Rosenberg, N. (1994). Exploring the black box: Technology, economics, and history. Cambridge University Press.

Sauermann, H. and Cohen, W. M. (2010). What makes them tick? employee motives and firm innovation. Management Science, 56(12):2134–2153.

Sauermann, H. and Roach, M. (2012). Taste for science, taste for commercialization, and hybrid scientists. Organisation, Economics and Policy of Scientific Research, Torino, Italy.

Serrano, C. J. (2010). The dynamics of the transfer and renewal of patents. The RAND Journal of Economics, 41(4):686–708.

Simeth, M. and Cincera, M. (2016). Corporate science, innovation, and firm value. Management Science, 62(7):1970–1981.

Singh, J. (2005). Collaborative networks as determinants of knowledge diffusion patterns. Management science, 51(5):756–770.

Singh, J. and Agrawal, A. (2011). Recruiting for ideas: How firms exploit the prior inventions of new hires. Management Science, 57(1):129–150.

Singh, J. and Fleming, L. (2010). Lone inventors as sources of breakthroughs: Myth or reality? Management science, 56(1):41–56.

Smith, A. (1776). The wealth of nations.

Spector, A., Norvig, P., and Petrov, S. (2012). Google's hybrid approach to research. Communications of the ACM, 55(7):34–37.

Standard & Poor's (2018). North america annual compustat, 1980-2015. Available to Duke University through Wharton Research Data Services (WRDS). https://wrds-web.wharton.upenn.edu/wrds/ds/comp/funda/index.cfm?navId=80 (Accessed August 2018).

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

Stettner, U. and Lavie, D. (2014). Ambidexterity under scrutiny: Exploration and exploitation via internal organization, alliances, and acquisitions. Strategic management journal, 35(13):1903–1929.

Stokes, D. (1997). Pasteur's quadrant: Basic science and technological innovation (the brookings institution, washington dc).

Tushman, M. L. (1977). Special boundary roles in the innovation process. Administrative science quarterly, pages 587–605.

Tushman, M. L. and Katz, R. (1980). External communication and project performance: An investigation into the role of gatekeepers. Management science, 26(11):1071–1085.

Tushman, M. L. and O'Reilly III, C. A. (1996). Ambidextrous organizations: Managing evolutionary and revolutionary change. California management review, 38(4):8–29.

Tzabbar, D., Silverman, B. S., and Aharonson, B. S. (2015). Learning by hiring or hiring to avoid learning? Journal of Managerial Psychology, 30(5):550.

Vincenti, W. G. (1990). What Engineers Know and How They Know It. The Johns Hopkins University Press, Baltimore, MD.

Wise, G. (1985). Willis R. Whitney, General Electric and the Origins of US Industrial Research. Columbia University Press.

Appendix A Supplementary Results

Table A.1: Integration Case Studies

COHORT	(1) IBM	(2) AT&T
1980-1985	0.23	0.20
1986-1990	0.21	0.20
1991-1995	0.30	0.27
1996-2000	0.42	0.38
2001-2005	0.50	0.43
2006-2010	0.53	N/A
2011-2015	0.60	N/A

Notes: The table presents the measure of integration for each firm-cohort.

Figure A.1: **Trend in Integration - Technology**

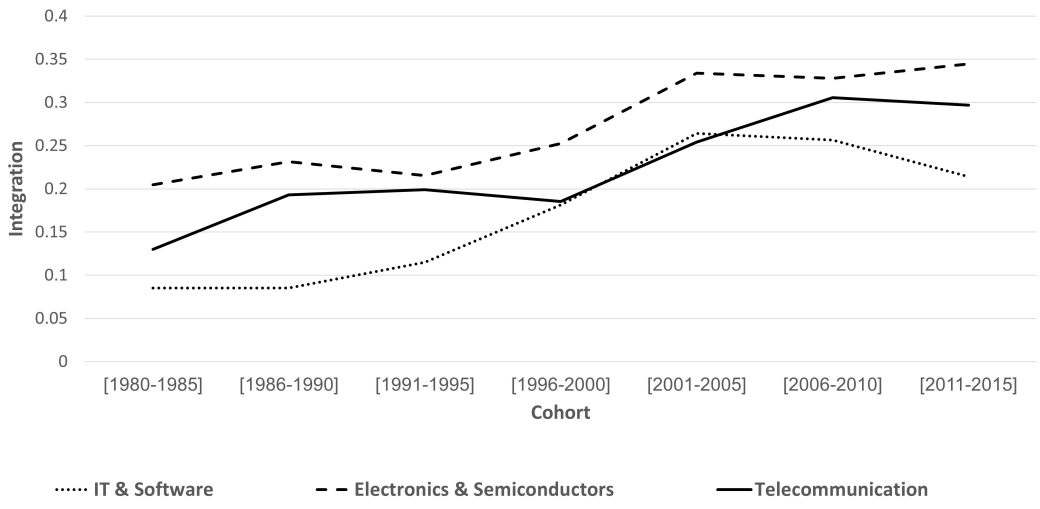

Figure A.2: **Trend in Integration - Life Science**

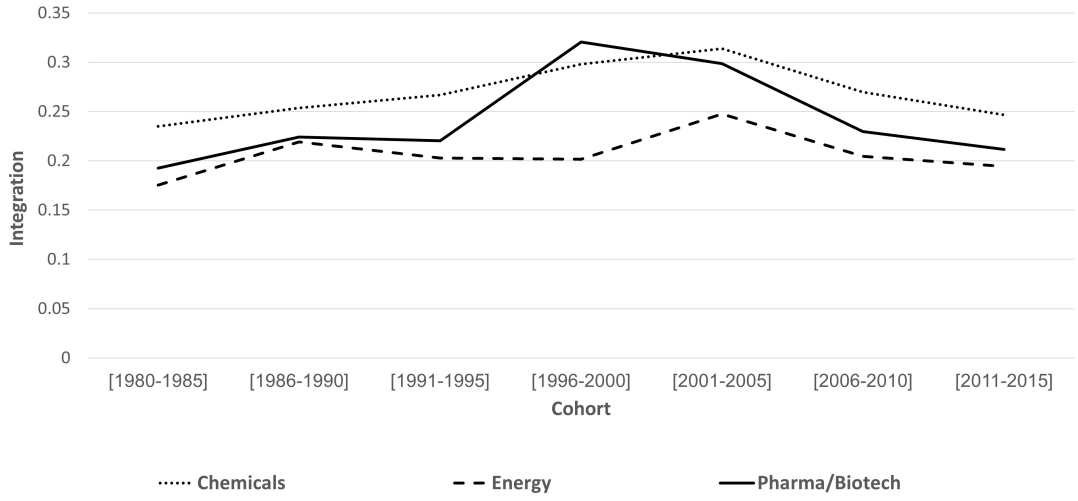

Table A.2: Robustness Checks for Integration Measure

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent variable:	Excluding new journals		Excluding low JIF journals		Dummy variable		Lag integration		Co-location	
	ln(Pub count)	ln(Patent count)	ln(Pub count)	ln(Patent count)	ln(Pub count)	ln(Patent count)	ln(Pub count)	ln(Patent count)	ln(Pub count)	ln(Patent count)
$Integration_{t-1}$	-0.230**	0.694**	-0.217**	0.505**			-0.122**	0.114*		
	(0.031)	(0.056)	(0.033)	(0.054)			(0.041)	(0.049)		
$Integration\ dummy_{t-1}$					-0.140**	0.244**				
					(0.020)	(0.033)				
$Co-location_{t-1}$									0.008	0.093**
									(0.018)	(0.031)
$ln(Publication\ stock)_{t-1}$		0.167**		0.122**		0.113**		0.252**		0.101**
		(0.037)		(0.026)		(0.031)		(0.035)		(0.031)
$ln(R\&D\ stock)_{t-1}$	0.072**	0.288**	0.078**	0.307**	0.082**	0.262**	0.260**	0.287**	0.081**	0.262**
	(0.011)	(0.030)	(0.012)	(0.034)	(0.011)	(0.030)	(0.023)	(0.032)	(0.011)	(0.031)
$ln(Authors)_{t-1}$	0.493**	0.145**	0.500**	0.171**	0.530**	0.161**	0.047**	-0.014	0.505**	0.206**
	(0.011)	(0.016)	(0.013)	(0.017)	(0.013)	(0.017)	(0.010)	(0.011)	(0.012)	(0.016)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean DV	25.852	55.673	24.364	65.319	31.882	50.708	31.882	50.708	32.002	50.927
Obs.	20,909	20,909	17,349	17,349	23,344	23,344	23,344	23,344	23,242	23,242
Number of firms	1333	1333	1105	1105	1526	1526	1526	1526	1512	1512
R-squared	.88	.87	.87	.88	.89	.87	.85	.86	.89	.87

Notes: This table presents robustness checks for the integration measure. In Columns 1 and 2, integration is computed by excluding scientific publications from new journals post 1990. In Columns 3 and 4, integration is computed excluding scientific publications with below-median JIF. Columns 5 and 6 include a dummy variable of integration. Columns 7 and 8 lag the integration measure by 5 years (1 cohort). In Columns 9 and 10, integration is defined based on the co-location of inventors and authors. All specifications include lagged dummies for zero publications per year and/or zero patents per year. One is added to logged control variables. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firms. ** p<0.01 * p<0.05

Table A.3: Robustness Checks for Integration and Scientific Discovery

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Dependent variable:	ln(Publication count)										
	Established scientific firms (pre-1990)	Young scientific firms (post-1990)	Sample pre-2000	Sample post-2000	Small firms	Large firms	Applied firms	Fundamental firms	Excluding Life-sciences	Inverse hyperbolic sine	Between firms
$Integration_{t-1}$	-0.263**	-0.176**	-0.200**	-0.250**	-0.255**	-0.207**	-0.107**	-0.345**	-0.175**	-0.243**	-0.183*
	(0.050)	(0.056)	(0.057)	(0.051)	(0.045)	(0.063)	(0.041)	(0.067)	(0.043)	(0.045)	(0.075)
$ln(R\&D\ stock)_{t-1}$	0.097**	0.091**	0.085**	0.106**	0.072**	0.082**	0.052**	0.087**	0.073**	0.088**	0.012
	(0.020)	(0.016)	(0.015)	(0.019)	(0.015)	(0.015)	(0.015)	(0.015)	(0.013)	(0.013)	(0.009)
$ln(Authors)_{t-1}$	0.426**	0.571**	0.516**	0.360**	0.402**	0.593**	0.422**	0.594**	0.520**	0.606**	0.788**
	(0.015)	(0.016)	(0.015)	(0.013)	(0.013)	(0.016)	(0.013)	(0.018)	(0.013)	(0.013)	(0.012)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Industry dummies	-	-	-	-	-	-	-	-	-	-	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Mean DV	12.365	47.918	33.146	30.799	6.381	56.69	3.603	60.205	26.229	2.174	22.85
Obs.	10,529	12,815	11,361	11,917	11,511	11,833	11,681	11,663	17,836	23,344	1,468
Number of firms	929	597	1074	1155	933	593	756	770	1032	1526	1468
R-squared	.82	.92	.91	.89	.76	.93	.67	.91	.88	.87	.92

Notes: This table presents robustness checks for the relationship between publication and integration. Columns 1 and 2 distinguishes between firms that enter the sample before and after 1990, respectively. In Columns 3 and 4 the panel is split by firm-years prior and post the year 2000, respectively. In Columns 5 and 6 the sample is split by below and above-median sales, respectively. Columns 7 and 8 divide the firm sample based on below and above-median reliance on science in invention, respectively. Column 9 excludes life science related firms based on SIC codes. All specifications include lagged dummies for zero publications per year. One is added to logged control variables. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firms. ** p<0.01 * p<0.05

Appendix B Data Appendix

To compute the integration measure, I start by identifying all authors listed in the scientific publications and inventors listed on the patent documents related to the corporate firm sample. My goal is to match authors and inventors related to the same ultimate owner (UO) firm during a 5-year cohort period in order to identify individuals who both patent and publish.[39] Any individual who had at least one publication during a 5-year cohort period is considered an author; any individual who had at least one patent during the same period is considered an inventor.

One challenge I face using WoS data is that not all authors are linked to an institutional affiliation address.[40] This can cause a problem for collaborative publications that consist of authors from several institutions. To overcome this challenge, I compiled three lists of author names- (i) the complete list of authors listed on each corporate publication, (ii) a list of all authors that were linked in WoS to their UO firm, and (iii) a list of all authors related to non-collaborate scientific publications (i.e., where all institutions listed on the publication are related to a unique UO firm). In addition, for the purpose of compiling the inventor name list, I exclude from the sample patents with multi-assignees, which are far less prominent than collaborative scientific publications (less than 2% of patents in the sample). I then use this inventor list to match the above author list (i). In other words, limiting the sample of inventors provides me the certainty that all matched author-inventor individuals are related to the focal firm. Finally, after matching, I use author lists (ii) and (iii) combined with the matched author-inventor results to compute the total number of unique authors for each firm-cohort.

One other challenge is that patent data report inventors' first names and last names, whereas WoS data list the last names and initials of authors. To resolve the differences, I first standardized inventor and author names in a similar way – last names and initials. Since the match is done within firm-cohort, I am less exposed to mismatches that identify different individuals as the same person, which could lead to an overestimation of integrated scientists. Nonetheless, I conduct extensive manual checks to verify the matches, especially of common and short names.[41]

I divide the sample period into seven cohorts of 5 consecutive calendar years. I fuzzy match the standardized list of inventor names with the list of standardized author names to identify for each UO firm-cohort individuals responsible for both a patented invention and a scientific publication. Next, I merge the matched results back to the patent level data to identify integrated scientists who perform both research and invention (i.e., both authors and inventors) and are part of an inventor team, which includes at least one specialized inventor (i.e., only inventor). By doing so, I essentially exclude from my measure of integration cases of paper-patent pairs as well as cases of lone inventors, where there is no real interaction between author and inventor teams.

Lastly, I aggregate the data to the firm-cohort level by counting the number of unique integrated scientists in each UO firm-cohort and dividing by the total number of unique authors

[39] The measure is per 5-year period as the main assumption is that integration is a long-term organizational feature.

[40] for example, in some cases, WoS only documents the link between the reprint author and her institution address, while in the original publication, all authors are linked to an address

[41] I also run the match with different levels of restrictions for false-positive matches. I confirm that results hold even when I use a very conservative definition for matches

in each UO firm-cohort, to compute my main measure for the analyses- *share integrated authors of all authors*. Then, in the regression analysis, I use a running average of the measure.