The Effect of Public Science on Corporate R&D*

Ashish Arora† Sharon Belenzon‡ Larisa Cioaca§ Lia Sheer¶

Hansen Zhang‖

March 3, 2023

*Acknowledgments: We thank Wes Cohen, Annamaria Conti, Elia Ferracuti, Peter Hull, and Andrew Shipilov, as well as participants at the 2022 Open Innovation in Science (OIS) Research Conference held at CERN IdeaSquare, the 2022 REER Conference held at Georgia Tech, the 2022 Israel Strategy Conference held at Reichman University, and the 10th Annual Israeli IO Workshop held at Tel Aviv University for their helpful comments.

†Arora: Duke University, Fuqua School of Business and NBER

‡Belenzon: Duke University, Fuqua School of Business and NBER

§Cioaca: Duke University, Fuqua School of Business

¶Sheer: Tel-Aviv University, Coller School of Management (liasheer@tauex.tau.ac.il)

‖Zhang: Duke University, Fuqua School of Business

1 Introduction

Public investment in science in the United States has grown very significantly since World War II, as has scientific output. For the first four or five decades, private investment in scientific research also grew. However, over the last two or three decades, corporations have reversed course. Figure 1 shows that the recent decline in publications produced by firms (Arora, Belenzon, & Patacconi, 2018; Mowery, 2009) has coincided with a rise in publications produced by universities, other nonprofits, and federal government laboratories (henceforth, "public knowledge"), as well as a rise in trained PhD scientists (henceforth, "human capital"). Moreover, there is no decline in corporate patenting (Arora, Belenzon, & Sheer, 2021a), and corporate R&D intensity has remained stable. In other words, corporations are reducing the share of "R" in their R&D, even as public investment in scientific research is increasing.

However, there are notable exceptions to this broad trend. In several emerging technology fields, including artificial intelligence (AI) and quantum computing, leading corporations, such as IBM, Microsoft, and Google, continue to invest in upstream research. IBM and Microsoft together have produced more quantum computing publications than MIT during 2013-2020.[1] Some of the best-known AI researchers and quantum computing experts now work for corporations rather than universities.[2]

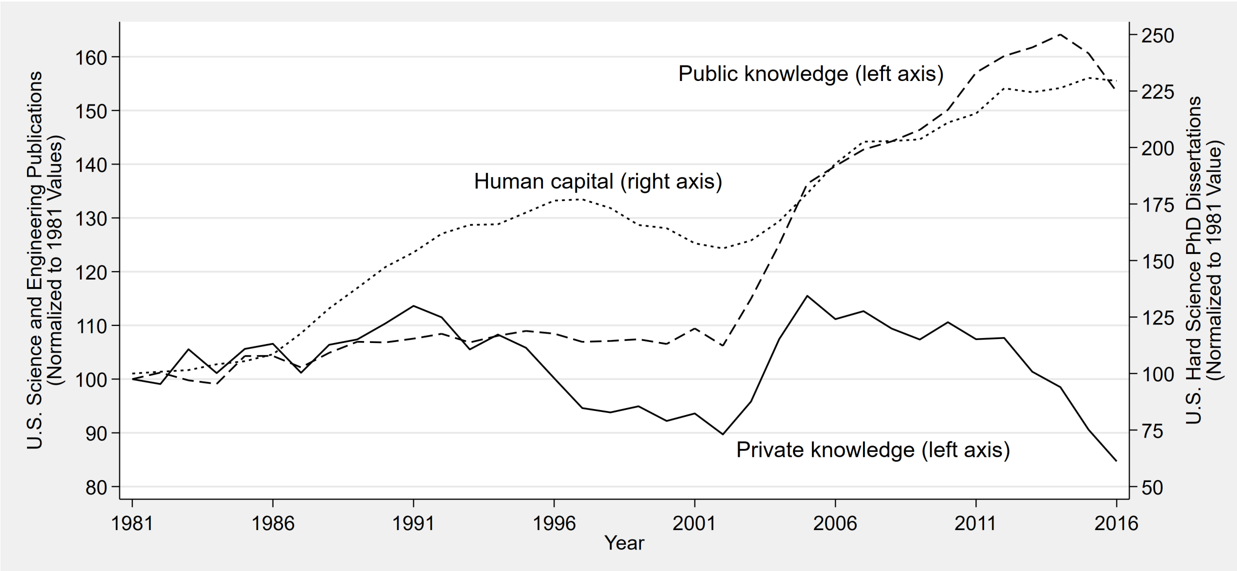
In this paper, we analyze how public science affects corporate innovation—both upstream research and downstream development. The relationships between public science and corporate scientific research and invention are complex. There are different dimensions to public science: human capital, in the form of trained PhD scientists, is produced along with disembodied knowledge, in the form of scientific publications. Universities also produce inventions that may be supplied to firms through different channels. We explore differences in firm responses to human capital (as measured by the flow of new PhDs) and disembodied knowledge (as measured by the stock of scientific publications). We also analyze differences between public knowledge covered by university patents and public knowledge that is in the public domain. And, we explore how the response of the leading firms differs from that of followers, and across different industries.

To capture the key economic forces behind the effect of public science, we develop a conceptual framework of private returns to investment in internal research, conditional on the state of public science. We distinguish between scientific knowledge and invention (Arora et al., 2021a). Innovation—the introduction of new products and processes—is the source of

[1]IBM, Microsoft, and MIT published 228, 165, and 283 publications, respectively, on "quantum computer/computing" during 2013-2020, based on data from Microsoft Academic Graph (Sinha et al., 2015; Wang et al., 2019).

[2]https://www.wsj.com/articles/universities-ai-talent-poached-by-tech-giants-1479999601

Figure 1: TRENDS IN PUBLIC AND PRIVATE SCIENCE, 1981-2016



Notes: This figure presents trends in the production of scientific knowledge (measured by U.S. science and engineering journal publications, left axis) and human capital (measured by U.S. hard science PhD dissertations, right axis) over time. Public knowledge includes publications from academia, other nonprofits, the federal government, and federally funded research and development centers (dashed line), while private knowledge includes publications from industry (solid line). Publication data for 1981-1995 are from Online Appendix Table 5-44 of Science and Engineering Indicators 1998 (National Science Board, 1998) and have been normalized by their 1981 values. Publication data for 1995-2003 are from Online Appendix Table 5-42 of Science and Engineering Indicators 2010 (National Science Board, 2010). Publication data for 2003-2016 are from Online Appendix Table 5-41 of the Science and Engineering Indicators 2018 (National Science Board, 2018). Dissertation data are from ProQuest Dissertations & Theses Global and have been normalized by their 1981 value.

profits. Innovations are based on inventions generated internally or on inventions acquired from outside. Scientific knowledge—from internal research and public science—reduces the cost of internal invention. Human capital lowers the cost of internal research.

How firms respond to an increase in public science depends on three key features. First, public knowledge could complement or substitute for internal research in firms, which would condition how internal research and invention respond to changes in public knowledge. Second, inventions from universities substitute for those generated inside firms, and thus lower the returns to internal research. Third, an increase in the supply of human capital reduces the cost of internal research. It may also directly reduce the cost of internal invention.

We consider two extensions to our conceptual framework. First, public science is available to all firms. Therefore, the nature of strategic interactions between competitors may affect the focal firm's innovation level and, hence, also its investment in internal research.[3]

[3]Strategic interactions describe situations where the outcome of one firm's choice depends on the actions of other firms.

Second, leading firms, operating closer to the technological frontier, derive greater returns from investing in internal research, and may respond differently than followers to increases in public science.

Understanding the mechanisms underlying the effect of public science is important, especially if public and private science are imperfect substitutes, as prior research suggests they are. Corporate research is more mission-oriented, more multi-disciplinary, and more closely tied to solving practical problems than university research (Rosenberg & Nelson, 1994). Corporations have access to specialized resources and expensive equipment, as well as links to manufacturing that can inform and subsequently scale useful research. Differences in incentives matter as well. While corporate researchers are rewarded for producing discoveries that are useful to their employers (Rosenberg, 1990), university researchers are rewarded for being first to make novel discoveries, even if those discoveries have little commercial application (Dasgupta & David, 1994). Potential heterogeneity in firm responses is important as well. If leaders respond to an increase in public science by increasing internal research and invention, whereas followers are more likely to pull back, then the increase in public science may widen the gap between leaders and followers, and possibly contribute to the rise of "superstar" firms.

Our empirical analysis includes all publicly traded firms headquartered in the United States that had at least one year of positive R&D expenditures, at least one granted patent, and at least three years of consecutive financial records from the first patent between 1980 and 2015. We measure corporate innovation using R&D intensity, patents, publications authored by corporate scientists, and the employment of scientists profiled in the American Men & Women of Science (hereafter, "AMWS scientists"). An important part of the analysis is measuring the public knowledge and human capital that are *potentially* relevant to the focal firm's innovation. We link publications to firms using non-patent literature (NPL) citations from patents granted in the same CPC subclasses as the patents of a firm. We link PhD dissertations to firms by using a state-of-the-art deep learning algorithm, SPECTER (Cohan, Feldman, Beltagy, Downey, & Weld, 2020), to measure the similarity between dissertations and the patents of a firm. We then use the publications of the thesis advisor(s) to link dissertations to funding from federal agencies.

Estimating the effect of public science on corporate innovation suffers from a classical endogeneity problem, whereby technological shocks that affect public science can also affect corporate innovation. This concern means that OLS estimates of the effect of public science on corporate innovation inputs (R&D intensity and employment of AMWS scientists) and outputs (patents and publications) would be upward biased. To deal with this, we use variation in funding for federal agencies to predict public science that is relevant to each firm.

Government funding affects the cost of public research and may, under some conditions, offer a source of variation that is orthogonal to technological opportunities. We measure each firm's potential exposure to federal funding using citations from patents granted in the subclasses in which the firm also has patents to publications from universities (and all other non-corporate sources). We link the sources of funding of these publications to specific federal agencies and use exogenous variation in agency funding to purge the endogeneity bias. We build our funding instruments for public knowledge and human capital by exploiting the windfall (or shortfall) in federal agency funding resulting from the congressional appropriations process (which is unlikely to be affected by technology shocks).

We present three main findings. First, we document a crowding-out effect of public knowledge on corporate research and invention. Moving from the first to the third quartile of relevant public knowledge lowers firm publications by about 7% and firm patents by about 10%. This effect, however, varies substantially across firms. Firms operating on the technology frontier (henceforth, "frontier firms") continue to invest in upstream research and downstream development even when public knowledge relevant to their innovation is abundant. Second, we document a crowding-in effect of human capital on corporate research and invention, especially for frontier firms.

On average, growth in public knowledge leads firms to withdraw from internal research. However, growth in human capital leads firms to increase internal research. Moreover, the negative effect of relevant public knowledge on internal research is smaller for frontier firms. Because frontier firms continue to invest in internal research and to invent, the gap between leaders and followers can grow. These results mean that public science can have an uneven effect on internal research, crowding out research by followers, while increasing research by leaders.

Third, we document a negative effect of competition, as rival use of public knowledge reduces the focal firm's market value. However, we do not find evidence that strategic interactions in the product market moderate the effect of public science on corporate innovation (i.e., there is no effect of rival use of public science on the focal firm's patenting, publishing, or hiring behavior).

We make two main contributions to the innovation literature. First, we contribute to the literature that examines the effect of public science on corporate R&D. Over the past several decades researchers have investigated the effects of public R&D at various levels of aggregation, including the economy as a whole (Narin, Hamilton, & Olivastro, 1997), states (Acs, Audretsch, & Feldman, 1994; Jaffe, 1989), industries (Adams, 1990), firms (Lichtenberg, 1984; Myers & Lanahan, 2022; Wallsten, 2000), diseases (Azoulay, Graff Zivin, Li, & Sampat, 2019), and individuals (Goolsbee, 1998). Perhaps not surprisingly—given the diversity of approaches and levels of analysis—these studies have produced conflicting results (David, Hall, & Toole, 2000). At the firm level, Dimos and Pugh (2016) conducted a meta-regression analysis of 52 studies. They rejected crowding out of firm R&D investment yet found no evidence of substantial crowding in either. Beyond this average null effect, studies found substantial heterogeneity by industry (Mulligan, Lenihan, Doran, & Roper, 2022), channel of interaction between firms and government laboratories (Adams, Chiang, & Jensen, 2003), firm size (González, Jaumandreu, & Pazó, 2005), firm age (Cohen, Nelson, & Walsh, 2002), and firm R&D intensity (Szücs, 2020).

Our study differs from this vast literature in two important ways. First, we model *how* firms capture returns on their investment in upstream research and downstream development, and how public science—both disembodied and embodied in PhD scientists—conditions these investments. To our knowledge, ours is the first large-scale empirical analysis that considers both components of public science. We also explore how private returns to science depend on how close firms are to the technology frontier. And, we consider strategic interactions in the product market. Prior research has shown that spillovers to rivals affect the focal firm's R&D investments by changing the marginal returns to such investments (Arora et al., 2021a; Bloom, Schankerman, & Van Reenen, 2013; Lucking, Bloom, & Van Reenen, 2018). In our conceptual framework, the ability of rivals to draw on public science can affect focal firms' decisions to invest in internal research. Empirically, though, we find no effect of rival use of public knowledge on focal firm innovation.

Second, we make a data contribution by using funding acknowledgements and other bibliometric linkages to connect federal agency funding to publications and PhD dissertations. Other studies have leveraged linkages between federal funding for research, the resulting publications, and the patents that cite them (e.g., Azoulay et al., 2019). However, to our knowledge, we are the first to systematically link corporate innovation to public knowledge and human capital that is relevant to the firm. Moreover, we trace funding to science while accounting for the identity of firms and the nature of competition between them. This is important because public science is free for all, so its impact on corporate R&D may depend on the nature of strategic interactions between the firms that use it.

The paper proceeds as follows. Section 2 presents the conceptual framework that guides our empirical investigation. Section 3 discusses and summarizes the data, Section 4 outlines the econometric specifications, and Section 5 presents the results. Section 6 concludes and suggests directions for future work.

2 Conceptual Framework

We adapt the framework from Arora et al. (2021a) to focus on the effect of public science on internal research and invention. Public science has at least three dimensions that affect private innovation: (disembodied) knowledge captured in scientific publications, trained human capital, and university inventions. In general, all three encourage innovation, but there are some subtle differences in how they affect internal research (i.e., upstream R&D). For instance, public knowledge may either complement internal research or substitute for it. University inventions substitute for internal inventions. Human capital from universities, on the other hand, tends to increase internal research and internal inventions.

The focal firm's product market profits, $\Pi(d)$ depends on the number of inventions it introduces into the market, d, which is the sum of (i) internal inventions, d_1, and (ii) externally acquired inventions, d_2. Internal inventions are produced at a unit cost $\phi(r;u)$, where r is internal research and u is the stock of public knowledge that is relevant to the firm. The term ϕ, which is the inverse of invention productivity, decreases with r at a diminishing rate. We also assume that ϕ decreases with u.[4] The firm can acquire external inventions at a cost represented by $a_0 d_2 + \frac{1}{2}\alpha_1 {d_2}^2$.

The relationship between public knowledge and internal research in reducing the unit cost of internal invention is important for how investments in internal research relate to the stock of public knowledge. We assume that the cost of internal research depends on the supply of trained PhD scientists produced by universities.

2.1 Setup

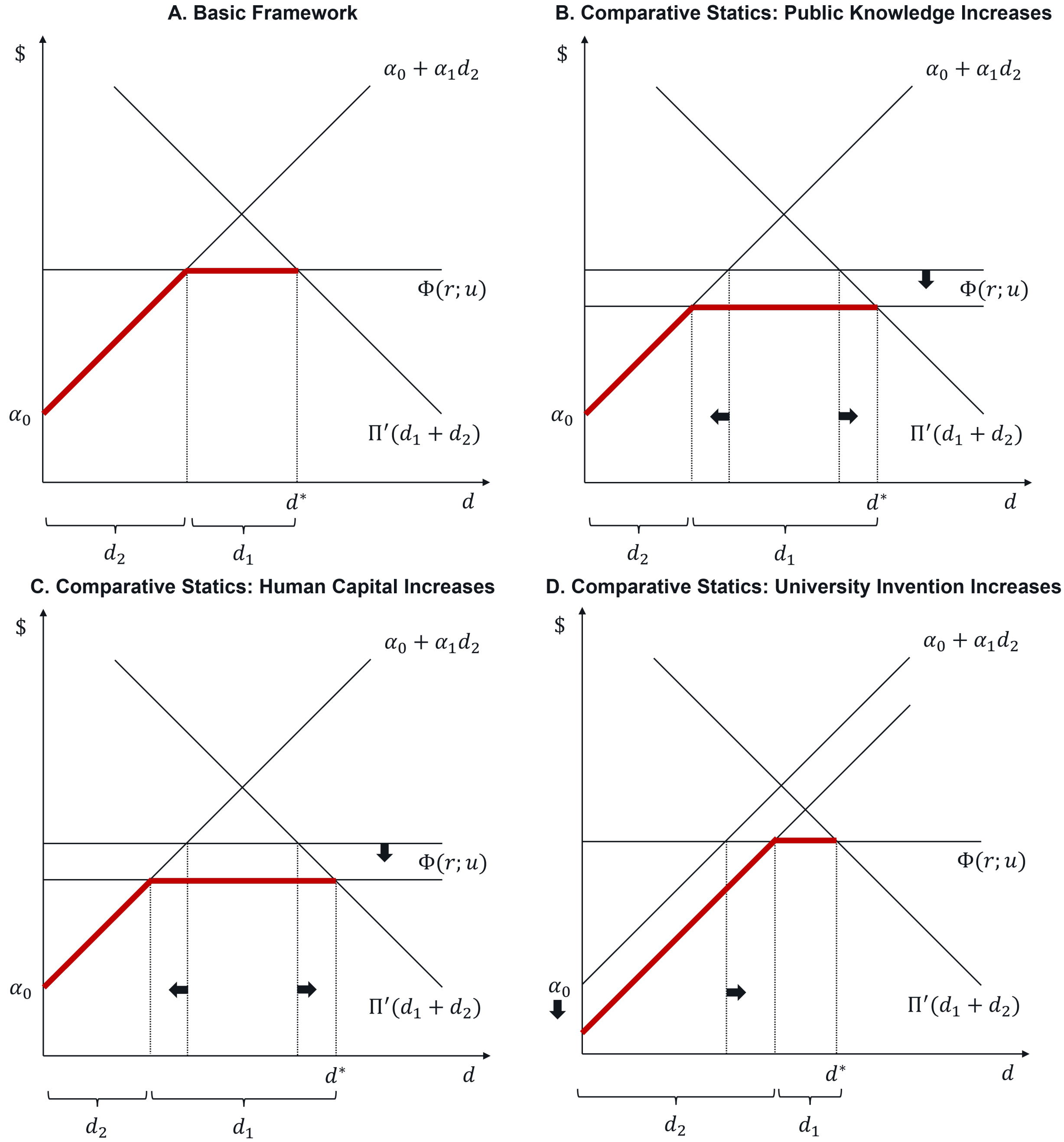
The value of the firm is $v = \max_{d_1,d_2,r}\{\Pi(d_1+d_2) - d_1\phi - \gamma(k)\frac{1}{2}r^2 - a_0 d_2 - \frac{1}{2}a_1 {d_2}^2\}$, where $\gamma(k)\frac{1}{2}r^2$ is the direct cost of internal research, and k is the supply of human capital (i.e., trained PhDs). Π represents the profits the firm gains in the product market due to its innovation output d. We assume that v is concave. Panel A in Figure 2 summarizes the elements of our basic conceptual framework.

2.2 Public Knowledge

An increase in relevant public knowledge increases the value of the firm, v, by reducing the cost of internal invention. Formally, applying the envelope theorem, $\frac{\partial v}{\partial u} = -d_1\frac{\partial \phi}{\partial u} > 0$. If internal research complements public knowledge (i.e., $-\frac{\partial^2 \phi}{\partial r \partial u} > 0$), then an increase in public

[4]The cost function reflects a simple linear production function $d = \lambda(r,u)K$, where K is the number of inventors the firm employs. Thus, normalizing the wage rate of inventors to unity, $\phi = \frac{1}{\lambda(r,u)}$.

Figure 2: Conceptual Framework



Notes: This figure presents the elements of our basic conceptual framework (Panel A). The "demand" for firm innovation output, d, is represented by $\Pi'(d_1 + d_2)$. The "supply" of external inventions is represented by $\alpha_0 + \alpha_1 d$, while the "supply" of internal inventions is represented by $\Phi(r; u)$, where r is internal research, u is the stock of relevant public knowledge, $\gamma(k)\frac{1}{2}r^2$ is the cost of internal research, and k is the supply of human capital. Firm innovation output, d, is the sum of internal inventions, d_1, and external inventions, d_2. The figure also presents comparative statics for increases in relevant public knowledge (Panel B), human capital (Panel C), and university invention (Panel D).

knowledge will also increase internal research. If they are substitutes, then there are two opposing effects. Substitutability reduces the marginal return to internal research. However, a reduction in the cost of internal invention due to public knowledge increases the scale of internal invention, thereby increasing the marginal return on internal research.

The effect on internal invention follows a similar logic. The direct effect of an increase in public knowledge is to reduce the marginal cost of internal invention, ϕ, as shown in Panel B. As long as the marginal cost of internal invention decreases, overall innovation increases as the increase in internal invention is only partly at the expense of external invention.

2.3 Human Capital

In our framework, an increase in human capital supply reduces the direct cost of internal research, $\gamma(k)$. As with public knowledge, an increase in human capital supply increases firm value. Formally, $\frac{\partial v}{\partial k} = -\frac{r^2}{2}\frac{\partial \gamma}{\partial k} > 0$. An increase in k will increase internal research and, therefore, also increase internal invention because, with an increase in r, ϕ must decrease, as shown in Panel C. Since external invention substitutes for internal invention, the former will fall. These results hold *a fortiori* if an increase in human capital also reduces the cost of internal invention, ϕ, directly.

2.4 University Invention

An increase in the supply of university inventions can be modeled as a reduction in a_0, as shown in Panel D. This will increase firm value. Formally, $-\frac{\partial v}{\partial a_0} = d_2 > 0$. However, a reduction in the marginal cost of external invention will decrease internal invention, which will, in turn, decrease internal research. Intuitively, an increase in the supply of an input increases the firm's value. However, it will decrease the demand for substitute inputs. Table 1 summarizes the predictions of our basic conceptual framework.

Notice that university inventions may also result in startups that compete with the firm. If they are strategic substitutes for the firm's innovations (i.e., they reduce the marginal return to innovation, $\frac{\partial \Pi}{\partial d}$), the effect on firm research and invention will be similar: both will fall in response to an increase in university invention, as will firm value.

In the empirical analysis, we focus on public knowledge and human capital supply. We indirectly examine the role of university invention by distinguishing public knowledge that is covered by university patents from the knowledge that is fully in the public domain.

Table 1: Theoretical Predictions Regarding the Effect of Public Knowledge, Human Capital, and University Invention on Internal Research, Internal Invention, and Firm Value

(1)	(2)	(3)	(4)
		Relationship between public knowledge and internal research in the internal invention cost function	
Equation	Comparative statics	Complements or unrelated	Substitutes
A. Effect of higher public knowledge			
Publications	$\partial r/\partial u$	$\uparrow$	$\downarrow\uparrow$
Patents	$\partial d_1/\partial u$	$\uparrow$	$\downarrow\uparrow$
Firm value	$\partial v/\partial u$	$\uparrow$	$\uparrow$
B. Effect of higher human capital			
Publications	$\partial r/\partial k$	$\uparrow$	$\uparrow$
Patents	$\partial d_1/\partial k$	$\uparrow$	$\uparrow$
Firm value	$\partial v/\partial k$	$\uparrow$	$\uparrow$
C. Effect of higher university invention			
Publications	$-\partial r/\partial a_0$	$\downarrow$	$\downarrow$
Patents	$-\partial d_1/\partial a_0$	$\downarrow$	$\downarrow$
Firm value	$-\partial v/\partial a_0$	$\uparrow$	$\uparrow$

Notes: This table summarizes the theoretical predictions regarding the effect of higher public knowledge, human capital, and university invention on the publications, patents, and value of the focal firm.

2.5 Leaders and Followers

Even if the fruits of public science are available to all, they may not benefit all firms equally. It is plausible that for leading firms, which require "frontier" innovations, sourcing external inventions that match their needs is more difficult. By contrast, for follower firms, which are trying to "catch-up" to the technology frontier, external inventions may be more plentiful. If so, frontier firms would naturally rely to a greater extent on internal inventions and also invest more in internal research compared to follower firms. This suggests that frontier firms may also respond differently than followers to public science. Public knowledge may substitute for internal research for followers, but may complement internal research in frontier firms. Insofar as human capital reduces the cost of internal research, frontier firms would be more responsive to increases in human capital. On the other hand, followers may respond more to an expansion in the supply of university invention.

2.6 Competition

Introducing competition adds some additional considerations. Obviously, invention by competitors reduce the payoff of the focal firm. By extension, a reduction in the cost of invention for competitors, whether through an expansion in the supply of external inventions to competitors or through an increase in the knowledge relevant to them, will reduce the payoff of the focal firm. But whether the focal firm changes its behavior in response depends upon how competition affects its *marginal* return to innovation. In turn, this depends on whether invention by rivals is a strategic substitute for invention by the focal firm or a strategic complement. With strategic interactions, the effects of public science on private investment in internal research and invention are not straightforward to analyze. For instance, by expanding its internal invention, a firm may deter invention by its rivals when inventions are strategic substitutes. Thus, a firm may cut back on internal research if a rival benefits from an expansion in public science, even though its own research costs are unchanged. The patterns are similar to the single-firm case if inventions are strategic complements. An increase in public knowledge increases internal research and invention unless the marginal product of internal research falls with public knowledge. For completeness, we include our conceptual framework for competition in the AppendixA.

In our empirical analysis, we analyze separately how the focal firm's investments in internal research and invention respond to the use of public knowledge by its rivals, as well as the effect on the focal firm's value. Our results suggest that strategic interactions stemming from product-market rivalry are not significant empirically. A direct implication is that the single firm analysis of the effect of public science on internal research and invention, as summarized in Table 1, provides a reasonable conceptual framework.

3 Data

We combine data from five primary sources: (i) scientific publications (by research institutions and corporations), grants, patents, and citations from Digital Science's Dimensions project; (ii) scientists profiled in the American Men & Women of Science directory; (iii) PhD dissertations from ProQuest Dissertations & Theses Global; (iv) federal R&D budgets from the American Association for the Advancement of Science (AAAS); and (v) Budgets of the U.S. Government from FRASER, the digital library of the Federal Reserve Bank of St. Louis (U.S. Office of Management and Budget, 2022). Firm financial information is from S&P's Compustat North America. The construction of the main variables used in our econometric analyses is summarized below and detailed in Online AppendixB.

3.1 Corporate Science: Publications and AMWS Scientists

We measure corporate scientific research using (i) the number of publications authored by scientists affiliated with the firm (from Arora et al., 2021a) and (ii) the number of scientists employed by the firm that were listed in the American Men & Women of Science, a directory of accomplished North American scientists in the physical, biological, and related sciences (similar to Kim & Moser, 2021). Over the course of 39 editions published between 1906 and 2021, AMWS has profiled more than 300,000 people, including such information as field of specialty, education, professional experience, memberships, research information, mailing address, fax number, and email address. We identify 17,063 AMWS scientists who worked for 1,321 different firms (including their subsidiaries) in our panel during 1986-2015.

Taken individually, publications and AMWS scientists are noisy measures of corporate investment in research. The pairwise correlation between the number of corporate publications and the number of AMWS scientists employed in a firm year is 0.68, suggesting that there is a strong shared component. However, as Online Appendix Table C4 shows, 46% of the firms that publish do not employ AMWS scientists, while 10% of the firms that employ AMWS scientists do not publish. This suggests that using both measures in our analyses is warranted.

3.2 Public Knowledge: Relevant Publications

Public scientific investments yield new scientific knowledge as well as new scientists. We source scientific publications from Dimensions. Central to our analysis is determining which scientific publications are relevant to a firm's innovation. We identify relevant publications using non-patent literature (NPL) citations from patents granted by the U.S. Patent and Trademark Office (USPTO) in a subclass where the focal firm also has patents. Our firm-year measure of *Relevant publications* is the weighted sum of university (and other non-corporate) publications cited by one or more of the patents in each subclass. The weights are the focal firm's shares of patents in each subclass during the previous 5-year time cohort, as follows:

$$Relevant\ publications_{i,t} = \sum_{s \in S} Precohort\ share\ of\ patents_{i,s} \times Relevant\ publications_{s,t} \tag{1}$$

S is the full set of patent subclasses, identified using the first four digits of the current CPC classification of patents from the USPTO. $Precohort\ share\ of\ patents_{i,s}$ is firm i's share of patents in subclass s (relative to all firm patents) during the previous (i.e., lagged) time cohort, obtained by dividing (i) the number of firm patents granted in subclass s in the previous time cohort by (ii) the total number of firm patents granted in the previous time cohort. $Relevant\ publications_{s,t}$ is the number of university (and other non-corporate) publications published in year t and cited by at least one patent (whether a corporate patent or a non-corporate patent) granted in subclass s during 1980-2020.

We generate a stock measure of *Relevant publications* using a perpetual inventory method with a 15% depreciation rate (similar to Hall, Jaffe, & Trajtenberg, 2005).

We construct a stock measure of *Rival-relevant publications* for each focal firm as the weighed sum of the stocks of *Relevant publications* for firms operating in the same industry, where an industry is defined using the Text-based Network Industry Classifications (TNIC, calibrated to be as granular SIC3 codes) from Hoberg and Phillips (2010, 2016). The weights are the pairwise textual similarity scores of product descriptions included in firms' annual reports on Form 10K filings for 1988-2015. The scores range from 0 to 1, with higher scores indicating more similar product portfolios that year. We also construct an alternative stock measure of *Rival-relevant publications* as the sum of the stocks of *Relevant publications* for firms operating in the same SIC4 industry as the focal firm.

3.3 Human Capital: Relevant PhD Dissertations

We measure human capital using PhD dissertations sourced from ProQuest Dissertations & Theses Global (hereafter, PQDT), which includes more than 5 million dissertations and theses from thousands of universities around the world between 1900 and 2021. PQDT is the largest collection of multidisciplinary dissertations and is recognized by the U.S. Library of Congress as the official repository for dissertations.

We take several steps to remove "soft science" PhD dissertations and master's degree theses from our data. Doing so is challenging because the field that records dissertation subjects, "classterms," lists 308,862 different combinations of subjects. First, we manually create a list of 1,027 disambiguated subjects. Second, we drop dissertations with a "soft science" subject, such as "literature," "history," and "social sciences." Third, we discard PhD dissertations from non-U.S. universities and master's degree theses. Our final dissertation dataset includes 771,023 U.S. PhD dissertations defended between 1985 and 2016 in 394 "hard science" subjects.

The firm-relevant supply of human capital is based on the textual similarity between the abstracts of dissertations and the abstracts of firm patents, calculated using Google's Bidirectional Encoder Representations from Transformers (BERT) algorithm. We implement BERT as detailed in Online AppendixB. Our firm-time cohort measure of *Relevant PhD dissertations* is the weighted sum of unique PhD dissertations that are in the top 1,000 most similar dissertations for each of the firm's patents granted in the 5-year time cohort. The weights are the maximum similarity scores between dissertations and the focal firm's patents granted in the 5-year time cohort, as follows:

$$Relevant\ PhD\ dissertations_{i,t} = \sum_{d \in D} Maximum\ textual\ similarity_{d,i,t} \tag{2}$$

D is the set of PhD dissertations in the top 1,000 most similar dissertations for one or more of the patents granted to firm i during time cohort t. $Maximum\ textual\ similarity_{d,i,t}$ is the maximum textual similarity score between the abstract of dissertation d and the abstracts of all the patents granted to firm i during 5-year time cohort t. We build stock measures of *Relevant PhD dissertations* and *Rival-relevant PhD dissertations* using the same approaches as described for relevant publications.

3.4 External Inventions

External inventions are typically based on knowledge produced by universities, and may also be embodied in the scientists and engineers produced. We construct a measure of *Relevant university patents* and also divide *Relevant publications* into those that are related to university patents and those that are purely in the public domain. To do so, we first identify all the patents granted to universities between 1980 and 2015. For each university patent, we use BERT to calculate the textual similarity between its abstract and the abstracts of all publications published in years $[t-1, t+1]$ relative to the patent application year, t. We consider the top 3 most similar publications as "covered" by the focal university patent. Our firm-year measure of *Relevant publications covered by university patents* is the weighted sum of university (and other non-corporate) publications that (i) are cited by one or more of the patents in each CPC subclass and (ii) are "covered" by a university patent. Once again, the weights are the focal firm's shares of patents in each subclass during the previous 5-year time cohort. We assume that publications with knowledge covered by patents represent inventions that firms can acquire, either via technology licensing or by acquiring the relevant academic startup.

3.5 Federal Funding for Public Science

The U.S. government is a significant funder of public science. In constant 2012 dollars federal agencies' R&D budgets increased from \$1.046 trillion in the 1980s to \$1.561 trillion in the 2010s (American Association for the Advancement of Science, 2020). The Dimensions dataset includes more than 4.6 million publications that acknowledge federal funding. Relevant to our study, federal funding for public science varies substantially by agency and over time. As shown in Online Appendix Table B2, some agencies are significant funders of R&D (e.g., Defense, Health and Human Services), while others are not (e.g., Environmental Protection Agency, Department of Homeland Security). More importantly, the composition of federal R&D investments has changed over time. Defense-related R&D has dropped from 58% of all federal R&D budgets in the 1980s to just 49% in the 2010s. Conversely, human health-related R&D has increased from 11% of all federal R&D budgets in the 1980s to 23% in the 2010s.

We measure federal funding for publications and PhD dissertations using (i) funding linkages between publications, grants, and funding organizations from the Dimensions dataset, (ii) funding organization identifiers from the Global Research Identifier Database (GRID), and (iii) PhD student-advisor linkages from PQDT. We use federal funding data to create instrumental variables for *Relevant publications* and *Relevant PhD dissertations* (as described in Section 4), as well as to identify industries for our alternative identification strategy (the panel event study included in Online Appendix E).

3.6 Citations to Publications

We use NPL citations to measure use of public knowledge in innovation. *Citations to publications* is the sum of NPL citations from a firm's patents granted in the focal year to non-corporate publications published since 1980. *Rival citations to publications* is the sum of *Citations to publications* for firms operating in the same industry as the focal firm. Additional details are included in Online AppendixB.

3.7 Descriptive Statistics

Our sample consists of 3,371 publicly traded firms headquartered in the U.S. with at least one year of reported R&D expenditures and at least one granted patent during 1980-2015 (Arora, Belenzon, & Sheer, 2021b), resulting in 41,693 firm-year observations. Table 2 presents descriptive statistics for the main variables used in the econometric analyses. Our sample includes a wide distribution of firm sizes: sales range from \$3.96 million (10th percentile) to \$5,546 million (90th percentile), while R&D expenditures range from approximately \$0.56 million (10th percentile) to \$203 million (90th percentile). Approximately 65% of firms perform internal research (i.e., have at least one publication during 1986-2015). On average, firms produce 16 publications and 28 patents per year, and employ 5 AMWS scientists. Firms also vary in their potentially relevant public science. The annual stock of *Relevant publications* ranges from zero publications (10th percentile) to 21,908 publications (90th percentile), while the cohort-level count of textual similarity-weighted *Relevant PhD dissertations* ranges from zero dissertations (10th percentile) to 17,124 dissertations (90th percentile). There is a positive correlation between relevant publications and relevant dissertations. However, this correlation is not perfect and is subject to some degree of variability (as shown in Online Appendix Table C5).

Table 2: Summary Statistics for Main Variables

	(1)	(2)	(3)	(4)	(5)	(6)
				Distribution		
	Obs.	Mean	Std. dev.	10th	50th	90th
Relevant publications	41,693	6,954	14,032	0.00	1,436	21,908
Relevant PhD dissertations	41,693	6,413	9,709	0.00	2,762	17,124
Relevant publications, Corp Pubs	41,693	11,570	15,700	0.00	0	32,189
Relevant university patents	41,693	365	643	0.00	141	997
R&D expenditures ($ mm)	36,709	142	656	0.56	12	203
R&D stock ($ mm)	41,693	600	3,159	0.52	34	765
Patents	41,693	28	157	0.00	1	44
Publications	41,693	16	94	0.00	0	17
AMWS scientists	41,693	5	32	0.00	0	5
Award-winning AMWS scientists	41,693	1	5	0.00	0	1
Sales ($ mm)	41,458	3,174	14,514	3.96	202	5,546
Tobin's Q	36,800	34	688	0.39	2	16
Assets	41,032	2,388	12,497	2.62	100	3,705

Notes: This table provides summary statistics for the main variables used in the econometric analyses. The analysis sample is at the firm-year level and includes an unbalanced panel of 3,371 U.S.-headquartered publicly traded firms during 1986-2015.

4 Econometric Framework

We turn to the empirical investigation of the model predictions from Table 1.

4.1 R&D Intensity, Patents, Publications, and AMWS Scientists Equations

We estimate the following specification for the relationship between public science and corporate innovation (bold indicates vector representation):

$$\begin{aligned}\ln(Y)_{i,t} =& \alpha_0 + \alpha_1 \ln(\textit{Relevant publications})_{i,t-1} + \alpha_2 \ln(\textit{Relevant PhD dissertations})_{i,t-1} \\ &+ \boldsymbol{Z}'_{i,t-1}\boldsymbol{\omega} + \boldsymbol{\eta_i} + \boldsymbol{\tau_t} + \epsilon_{i,t}\end{aligned} \tag{3}$$

Y_{it} represents different measures of corporate innovation—*R&D intensity*, *Patents*, *Publications*, and *AMWS scientists*, respectively—for firm i in year t. *R&D intensity* is R&D expenditures divided by sales. *Relevant publications* is the stock of public knowledge that is relevant to firm i's innovation, while *Relevant PhD dissertations* is the 5-year time cohort count of human capital that is relevant to firm i's innovation (the construction of these variables is detailed in Online AppendixB). The vector $\boldsymbol{Z}$ includes time-varying controls, such as *ln(Sales)* for the R&D intensity equation and *ln(R&D stock)* for the patents, publications, and AMWS scientists equations (where we also add an unreported indicator variable equal to 1 for firms without R&D expenditures prior to the focal year). In all specifications, we account for a possible direct funding effect by adding an unreported control for lagged *ln(Awards to focal firm)*, a measure of federal procurement contract and grant dollars awarded to the focal firm and its subsidiaries. In the 2SLS specifications, we also add unreported indicator variables equal to 1 for firms without positive instruments prior to the focal year and an unreported control for lagged *Agency exposure*. The vectors $\boldsymbol{\eta}$ and $\boldsymbol{\tau}$ are firm and year fixed effects, respectively, and ϵ is an *iid* error term. When calculating natural logarithms, we add \$1 to variables measured in millions of dollars (e.g., *Sales, R&D stock*) and one unit to other variables (e.g., patents, publications, AMWS scientists). Standard errors are clustered at the firm level.

Our coefficients of interest are α_1 and α_2. If $\alpha_1 < 0$ in the patents, publications, and AMWS scientists equations, this would imply that internal research and public knowledge are substitutes in the internal invention cost function. We expect $\alpha_2 \geq 0$. In some specifications, we distinguish between (i) relevant publications that are covered by university patents and (ii) relevant publications that are not. The latter should reflect the pure effect of u, public knowledge, whereas the former likely also reflects the supply of external inventions. In other specifications, we distinguish between papers that are not cited by patents and those that are.

We present results where *Relevant publications* and *Relevant PhD dissertations* are instrumented using different instrumental variables, as described in Section 4.2. We also examine heterogeneity in effects by firm proximity to the technology frontier and by main industry. To examine strategic interactions between firms, we also estimate specifications where rival use of public knowledge and rival-relevant dissertations are included in the patents, publications, and AMWS scientists equations.

4.2 Identification Strategy

A key econometric challenge is how to deal with the endogeneity of public science. Unobserved, time-varying technological shocks can influence both public science and corporate innovation. We deal with this problem by implementing an instrumental variable identification strategy. Specifically, we exploit changes in (i) federal agency R&D budgets, and (ii) federal agency windfall (or shortfall) in total funding resulting from the congressional appropriations process to predict firm-relevant publications and PhD dissertations. The construction of our instrumental variables is summarized below and detailed in Online AppendixB.

4.2.1 Instruments for Relevant Publications

First, we construct a measure of *Relevant R&D budget* by combining (i) the value of the federal R&D budget that is relevant to each patent subclass-agency-year with (ii) a focal firm's shares of patents granted in each patent subclass in the previous 5-year time cohort. This measure is used as an instrument for *Relevant Publications* for firm i in year t:

$$\begin{aligned} \textit{Relevant R\&D budget}_{i,t} = &\sum_{s\in S}\sum_{a\in A} \textit{Precohort share of patents}_{i,s} \\ &\times \textit{Reliance on public knowledge}_{s,a} \\ &\times \textit{R\&D budget}_{a,t} \end{aligned} \tag{4}$$

S is the full set of patent subclasses. A is the set of 12 main federal agencies, plus an "Other" category for smaller agencies. *Precohort share of patents*$_{i,s}$ is as previously defined. *Reliance on public knowledge*$_{p,a}$ is a share obtained by dividing (i) the number of NPL citations from patents in subclass s to non-corporate publications funded by agency a by (ii) the total number of NPL citations from patents in subclass s to all non-corporate publications. *R&D budget*$_{a,t}$ is the R&D budget appropriated by the U.S. Congress to agency a in year t.

Because agency R&D budgets may reflect a technological shock that is common to the R&D decisions of firms, our preferred instrument for *Relevant publications* uses data on the differences between the *proposed* budget and the *actual* budget enacted by Congress for each federal agency (similar to Dugoua, Gerarden, Myers, & Pless, 2022). We construct our measure of *Windfall-predicted R&D budget* by using the difference between (i) an agency's actual total budget enacted by Congress and (ii) the agency's proposed total budget included in the President's budget proposal to Congress to predict the agency's annual *R&D budget*, then using these predicted values in Equation 4.

The idea behind this approach is as follows. The Budget of the U.S. Government discloses, for each main federal agency, both the total amount proposed by the President and the total amount actually authorized by Congress. Demand for funding (the proposed amount) is a function of the common technological shock that can affect both public science and corporate innovation. However, the actual budget authorized by Congress includes a component that is independent of this shock. Hence, we can use an agency's windfall (or shortfall) in total funding as a source of exogenous variation in its R&D budget.

4.2.2 Instruments for Relevant PhD Dissertations

We construct analogous measures of federal R&D budgets to instrument each firm's *Relevant PhD dissertations*. We use information on each dissertation advisor's name, school, and field of study to match advisors to the Dimensions dataset and retrieve (i) the scientific publications authored by the advisors during the 5-year duration of the PhD program, (ii) the grant amounts and funding organizations for these publications. In our dissertation dataset, 1,310,774 dissertations have advisor information, producing 1,472,326 dissertation-advisor pairs (some dissertations have more than one advisor). We assume that grant funding received by the advisor(s) of a PhD student during the 5-year duration of the PhD program affects the direction and content of the dissertation. Our measure of *Advisors' R&D budget* combines (i) the value of federal R&D budget that is relevant to each dissertation with (ii) the maximum textual similarity between the dissertation and a focal firm's patents granted in a 5-year time cohort.

Our preferred instrument for each firm's *Relevant PhD dissertations* is *Advisors' windfall-predicted R&D budget*, which leverages the windfall (or shortfall) in total funding for federal agencies resulting from the political negotiation process in Congress. Additional details are available in Online AppendixB.

Our instrumental variable is constructed as the average of the windfall-predicted R&D budget for federal agencies, weighted by the relevance of the R&D budget of the agency for the focal firm. A key identifying assumption is that the share of a patent subclass in the firm's patenting is not related to variations in the R&D budget of the agencies. We empirically probe this issue by using an event study described in further detail in Online Appendix E.2, where we use the collapse of the Soviet Union as a source of exogenous change in the R&D budgets of various Federal agencies. We find that firms increased publication activity in response to a fall in the R&D budget of relevant federal agencies.

5 Estimation Results

5.1 R&D Intensity Equation

The general pattern of results in Columns 1-5 of Table 3 points to a negative effect of relevant publications and a positive effect of relevant dissertations on R&D intensity. However, the effects are not statistically significant in most specifications. Subject to the proviso that they reflect average effects, the 2SLS estimates in Columns 2 and 4 suggest that public science may only modestly affect R&D intensity overall.[5]

We next consider the effect of relevant publications covered by university patents and relevant university patents, respectively. The pattern of negative/positive results remains largely unchanged (Columns 6 and 7), implying that *Relevant publications* and *Relevant PhD dissertations* capture different effects of disembodied and embodied public science. Overall, these results also strongly suggest that relevant publications reflect both public knowledge as well as university inventions. The negative coefficient estimate on relevant publications may therefore either reflect substitution in invention between public knowledge and corporate research, or the substitution between internal and external inventions, or both.

5.2 Patents Equation

Table 4 presents the effect of public knowledge and human capital on firm patents (our measure of corporate invention). While we find a positive correlation between public knowledge and patents (Column 1), once we account for endogeneity, we find that an increase in public knowledge leads to fewer corporate patents (Column 2, p-value < 0.001). Moving from the first to the third quartile of relevant public knowledge decreases firm patents by 10%. In unreported specifications, we find a 15% decrease in patents for firms that publish at least one publication during 1986-2015 (p-value < 0.001). Conversely, an increase in the relevant human capital increases corporate patenting, as predicted by our conceptual framework (Columns 3 and 4). The same pattern of results holds when we include both relevant publications and relevant PhD dissertations (Column 5, p-values < 0.001).

We address the concern that publications and PhDs are jointly produced by including only relevant publications from national laboratories—which produce scientific knowledge for the public domain but do not train human capital—in Column 6.

Our patenting results may seem to contradict prior empirical findings. For example,

[5] In the 2SLS specifications, the first stage predicts *Relevant publications* against *Windfall-predicted R&D budget* (in Column 2) and *Relevant PhD dissertations* against *Advisors' windfall-predicted R&D budget* (in Column 4). The first stage results reported in Online Appendix Table D14 confirm that both public knowledge and human capital are positively related to federal agencies' R&D budgets (p-value < 0.001).

Table 3: Main Effect of Relevant Public Science on R&D Intensity

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	ln(R&D intensity)					
	OLS	2SLS	OLS	2SLS	2SLS	2SLS
ln(Relevant publications)$_{t-1}$	0.012	-0.100			-0.156	
	(0.010)	(0.053)			(0.063)	
ln(Relevant PhD dissertations)$_{t-1}$			0.015	0.079	0.158	0.179
			(0.010)	(0.042)	(0.057)	(0.062)
ln(Relevant publications, Corp Pubs)$_{t-1}$	0.006	0.001				
	(0.004)	(0.006)				
ln(Relevant university patents)$_{t-1}$						-0.198
						(0.077)
ln(Sales)$_{t-1}$	-0.104	-0.101	-0.104	-0.106	-0.107	-0.108
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Mean DV	4.63	4.63	4.63	4.63	4.63	4.63
Weak identif. (Kleibergen-Paap)		129.98		735.50	129.32	132.38
Firms	3,125	3,125	3,125	3,125	3,125	3,125
Observations	35,833	35,833	35,833	35,833	35,833	35,833
Adjusted R-squared	0.74		0.74			

Notes: This table presents estimation results for the relationship between relevant public science (including both public knowledge and human capital) and firm R&D intensity. In the 2SLS specifications, *Relevant publications*, *Relevant PhD dissertations*, *Relevant publications covered by university patents*, and *Relevant university patents* are instrumented using *Windfall-predicted R&D budget*, *Advisors' windfall-predicted R&D budget*, and a modified version of *Windfall-predicted R&D budget*, respectively. One is added to logged variables. Standard errors (in parentheses) are robust to arbitrary heteroskedasticity and allow for serial correlation through clustering by firms.

Jaffe (1989) uses state-level data to show that corporate patenting increases with university research expenditures. Azoulay et al. (2019) use disease-level data to show that total private-sector patenting increases with National Institutes of Health research funding. Conversely, we use firm-level data and find a negative effect in a sample of large, publicly traded firms. Both types of results could easily be reconciled if small, privately held firms (such as university spinoffs) were responsible for driving an increase in total patenting.

Indeed, our conceptual framework suggests that, insofar as relevant publications are also related to the supply of university inventions, there may be countervailing effects. Recall that an increase in university invention would directly reduce internal invention. When we include the stock of relevant university patents in Column 7, we find that university patents have a negative effect on patenting by public companies (p-value < 0.001).

These are average effects and may differ across firms. For instance, public knowledge may complement research conducted in firms operating on the technology frontier. Followers, by contrast, may be content to use established science, rather than invest in internal research. We examine how these effects vary across firms in section 5.5.

Table 4: Main Effect of Relevant Public Science on Patents

Dependent variable:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	ln(Patents)							
	OLS	2SLS	OLS	2SLS	2SLS	2SLS	2SLS Only natl. lab pub.	2SLS
ln(Relevant publications)$_{t-1}$	0.012	-0.150			-0.217	-0.097	-0.363	
	(0.003)	(0.024)			(0.025)	(0.022)	(0.049)	
ln(Relevant PhD dissertations)$_{t-1}$			0.030	0.220	0.346	0.339	0.387	0.358
			(0.003)	(0.023)	(0.032)	(0.032)	(0.038)	(0.032)
ln(Relevant publications, Corp Pubs)$_{t-1}$	0.009	0.009				0.007		
	(0.002)	(0.004)				(0.004)		
ln(Relevant university patents)$_{t-1}$								-0.234
								(0.027)
ln(R&D stock)$_{t-1}$	0.244	0.271	0.238	0.199	0.191	0.020	0.189	0.195
	(0.017)	(0.020)	(0.017)	(0.015)	(0.016)	(0.005)	(0.017)	(0.015)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean DV	28.30	28.30	28.30	28.30	28.30	28.30	28.30	28.30
Weak identif. (Kleibergen-Paap)		147.90		746.86	182.18	109.26	92.79	194.93
Firms	3,371	3,371	3,371	3,371	3,371	3,371	3,371	3,371
Observations	41,693	41,693	41,693	41,693	41,693	41,693	41,693	41,693
Adjusted R-squared	0.86		0.86					

Notes: This table presents estimation results for the relationship between relevant public science (including both public knowledge and human capital) and firm patents. In the 2SLS specifications, *Relevant publications*, *Relevant PhD dissertations*, and *Relevant university patents* are instrumented using *Windfall-predicted R&D budget*, *Advisors' windfall-predicted R&D budget*, and a modified version of *Windfall-predicted R&D budget*, respectively. One is added to logged variables. Standard errors (in parentheses) are robust to arbitrary heteroskedasticity and allow for serial correlation through clustering by firms.

5.3 Publications Equation

Table 5 presents the effect of public knowledge and human capital on firm publications (a measure of corporate investment in internal research). Similar to the results for patents, we find a negative and statistically significant effect of public knowledge (Column 2). At the sample means, moving from the first to the third quartile of relevant public knowledge decreases firm publications by 7%. In unreported specifications, we find that the effect is a larger 11% decrease among firms that publish at least one publication during 1986-2015.

As predicted by our conceptual framework, relevant human capital has a positive and statistically significant effect on publications (Column 4, p-value < 0.001). The negative/positive effects disembodied/embodied public science remain when both dimensions are considered together (Column 5), and when we use only publications from national labs, which might provide a less contaminated estimate of the effect of public knowledge (Column 6).

The negative effect of public knowledge on corporate research might suggest that public knowledge impairs the effectiveness of internal research in reducing the cost of internal invention. Alternatively, our measure of public knowledge may also be related to the supply of external invention. Indeed, university patents have a negative effect on corporate research (Column 7), consistent with external invention substituting for internal invention. As was the case with patents, it is likely that these average effects disguise substantial heterogeneity across frontier firms and followers.

Table 5: Main Effect of Relevant Public Science on Publications

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable:	ln(Publications)							
	OLS	2SLS	OLS	2SLS	2SLS	2SLS	2SLS Only natl. lab pub.	2SLS
ln(Relevant publications)$_{t-1}$	0.005	-0.107			-0.134	-0.044	-0.266	
	(0.002)	(0.018)			(0.019)	(0.016)	(0.037)	
ln(Relevant PhD dissertations)$_{t-1}$			0.008	0.072	0.151	0.145	0.195	0.162
			(0.002)	(0.015)	(0.021)	(0.021)	(0.026)	(0.021)
ln(Relevant publications, Corp Pubs)$_{t-1}$	0.006	0.005				0.006		
	(0.002)	(0.003)				(0.003)		
ln(Relevant university patents)$_{t-1}$								-0.151
								(0.020)
ln(R&D stock)$_{t-1}$	0.148	0.166	0.149	0.136	0.130		0.128	0.132
	(0.013)	(0.015)	(0.013)	(0.013)	(0.013)		(0.014)	(0.013)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean DV	15.72	15.72	15.72	15.72	15.72	15.72	15.72	15.72
Weak identif. (Kleibergen-Paap)		147.90		746.86	182.18	108.35	92.79	194.93
Firms	3,371	3,371	3,371	3,371	3,371	3,372	3,371	3,371
Observations	41,693	41,693	41,693	41,693	41,693	41,698	41,693	41,693
Adjusted R-squared	0.88		0.88					

Notes: This table presents estimation results for the relationship between relevant public science (including both public knowledge and human capital) and firm publications. In the 2SLS specifications, *Relevant publications*, *Relevant PhD dissertations*, and *Relevant university patents* are instrumented using *Windfall-predicted R&D budget*, *Advisors' windfall-predicted R&D budget*, and a modified version of *Windfall-predicted R&D budget*, respectively. One is added to logged variables. Standard errors (in parentheses) are robust to arbitrary heteroskedasticity and allow for serial correlation through clustering by firms.

5.4 AMWS Scientists Equation

Table 6 presents the estimation results for the relationship between relevant publications and PhD dissertations and firm employment of AMWS scientists (our alternative measure of internal research). Evaluated at the sample means, the 2SLS estimates in Column 2 and 7 suggest that moving from the first to the third quartile of relevant public knowledge decreases employment of AMWS scientists by 2% for all firms, and by 4% for publishing firms.

Consistent with the publication results, the effect of relevant human capital is positive and significant (Columns 3-9), while the effect of public knowledge is negative and significant, even when using only publications from the national labs (Columns 6 and 8). Similarly, the supply of university patents has a negative effect on firm employment of AMWS scientists (Column 9).

Table 6: Main Effect of Relevant Public Science on the Employment of AMWS Scientists

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent variable:	ln(AMWS scientists)									
	All firms							Publishing firms		
	OLS	2SLS	OLS	2SLS	2SLS	2SLS	2SLS Only natl. lab pub.	2SLS	2SLS Only natl. lab pub.	2SLS
ln(Relevant publications$_{t-1}$)	0.004	-0.026			-0.039	-0.017	-0.070	-0.066	-0.110	
	(0.002)	(0.010)			(0.011)	(0.011)	(0.019)	(0.018)	(0.030)	
ln(Relevant PhD dissertations)$_{t-1}$			0.007	0.035	0.057	0.057	0.067	0.074	0.088	0.074
			(0.002)	(0.010)	(0.013)	(0.013)	(0.015)	(0.018)	(0.021)	(0.017)
ln(Relevant publications, Corp Pubs)$_{t-1}$	0.001	-0.002				-0.003				
	(0.001)	(0.002)				(0.002)				
ln(Relevant university patents)$_{t-1}$										-0.073
										(0.018)
ln(R&D stock)$_{t-1}$	0.036	0.042	0.034	0.028	0.027		0.027	0.029	0.027	0.032
	(0.007)	(0.008)	(0.007)	(0.008)	(0.008)		(0.008)	(0.011)	(0.011)	(0.010)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean DV	4.80	4.80	4.80	4.80	4.80	4.80	4.80	6.39	6.39	6.39
Weak identif. (Kleibergen-Paap)		147.90		746.86	182.18	108.35	92.79	106.21	61.87	152.60
Firms	3,371	3,371	3,371	3,371	3,371	3,372	3,371	2,255	2,255	2,255
Observations	41,693	41,693	41,693	41,693	41,693	41,698	41,693	31,154	31,154	31,154
Adjusted R-squared	0.93		0.93							

Notes: This table presents estimation results for the relationship between relevant public science (including both public knowledge and human capital) and firm employment of AMWS scientists. In the 2SLS specifications, *Relevant publications*, *Relevant PhD dissertations*, and *Relevant university patents* are instrumented using *Windfall-predicted R&D budget*, *Advisors' windfall-predicted R&D budget*, and a modified version of *Windfall-predicted R&D budget*, respectively. One is added to logged variables. Standard errors (in parentheses) are robust to arbitrary heteroskedasticity and allow for serial correlation through clustering by firms.

In summary, our key findings thus far are that (1) public knowledge, as measured by the stock of relevant publications, tends to crowd out corporate research and invention, whereas

(2) human capital, as measured by trained PhD scientists, tends to crowd in corporate research and invention. These results are robust across most industries, as reported in Online Appendix F. They are also robust to using modified versions of our measure of public knowledge (see Online Appendix Table G17), to constructing alternative measures of relevant public knowledge and human capital that rely on the textual similarity between publications and patents, as well as the grouping of publications in OECD subfields (see Online Appendix Table G19), and to using measures of high-quality corporate innovation as the dependent variables (see Online Appendix Table G20). In unreported specifications, we find similar results when we transform the dependent variable using an inverse hyperbolic sine transformation, calculated as $asinh(x) = ln(x + \sqrt{x^2 + 1})$.

However, it is likely that these effects differ in strength, and perhaps even in direction, in frontier firms versus followers. We investigate variation by firm proximity to the technology frontier next.

5.5 Frontier Firms Versus Followers

Frontier firms may differ from followers in the type of the inventions they produce or in terms of the value they derive from inventions, or both. We capture a firm's proximity to the technology frontier using its annual flow of novel patents, where patent novelty is based on unique IPC combinations, in Table 7 and its patent values from Kogan, Papanikolaou, Seru, and Stoffman (2017) in Table 8. Tables 7 and 8 present estimates from the second stage of 2SLS regressions using *Windfall-predicted R&D budget*, *Advisors' windfall-predicted R&D budget*, and their interactions with *Tech frontier* measures as instrumental variables for *Relevant publications*, *Relevant PhD dissertations*, and their interactions with *Tech frontier* measures, respectively. We also report results where patent novelty is based on being the first patent in a CPC subclass in Online Appendix Table G21. All *Tech frontier* measures yield similar results.

The coefficient estimates on the interaction terms suggest that there is substantial heterogeneity in the effect of public knowledge on internal research and innovation based on firm proximity to the technology frontier. While the estimates in Tables 4, 5, and 6 show that, on average, public knowledge crowds-out corporate patents, publications, and employment of AMWS scientists, firms operating on the technology frontier increase their patenting, publishing, and hiring when the stock of relevant non-corporate publications increases. Similarly, though both frontier firms and followers increase their patenting, publishing, and hiring in response to an increase in human capital supply, frontier firms do so to a greater extent.

Our results suggest that public knowledge may substitute for internal research, on aver-

Table 7: Variation by Firm Proximity to the Technology Frontier: Unique IPC Combinations

	(1) ln(Patents)	(2) ln(Pubs.)	(3) ln(AMWS Scientists)	(4) ln(Patents)	(5) ln(Pubs.)	(6) ln(AMWS Scientists)
ln(Relevant publications)$_{t-1}$ × $Tech\ frontier$	0.085	0.023	0.010			
	(0.002)	(0.002)	(0.001)			
ln(Relevant PhD dissertations)$_{t-1}$ × $Tech\ frontier$				0.059	0.020	0.012
				(0.002)	(0.003)	(0.002)
ln(Relevant publications)$_{t-1}$	-0.074	-0.089	-0.020	-0.041	-0.011	-0.007
	(0.012)	(0.016)	(0.010)	(0.003)	(0.003)	(0.002)
ln(Relevant PhD dissertations)$_{t-1}$	0.035	0.041	0.013	0.069	0.020	0.003
	(0.007)	(0.009)	(0.006)	(0.014)	(0.014)	(0.009)
ln(R&D stock)$_{t-1}$	0.074	0.103	0.016	0.117	0.108	0.011
	(0.008)	(0.011)	(0.007)	(0.010)	(0.013)	(0.009)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Mean DV	28.30	15.72	4.80	28.30	15.72	4.80
Weak identif. (Kleibergen-Paap)	178.61	178.61	178.61	450.01	450.01	450.01
Firms	3,371	3,371	3,371	3,371	3,371	3,371
Observations	41,693	41,693	41,693	41,693	41,693	41,693

Notes: This table presents the second stage of 2SLS estimation for the effect of public knowledge and human capital on firm patents, publications, and AMWS scientists when considering firm proximity to the technology frontier. We measure *Tech frontier* using a firm-cohort measure based on unique IPC combinations. We use *Windfall-predicted R&D budget* and its interactions with *Tech frontier* as instrumental variables for *Relevant publications* and its interactions with *Tech frontier*. We also use *Advisors' windfall-predicted R&D budget* and its interactions with *Tech frontier* as instrumental variables for *Relevant PhD dissertations* and its interactions with *Tech frontier*. One is added to logged variables. Standard errors (in parentheses) are robust to arbitrary heteroskedasticity and allow for serial correlation through clustering by firms.

age. However, public knowledge may well complement internal research for frontier firms. Alternatively, frontier firms may have a larger scale of operations. By reducing the cost of internal invention, public knowledge may increase the marginal return to internal research even if public knowledge substitutes for internal research in frontier firms. Furthermore, human capital has a larger positive effect on invention in frontier firms, as might be expected if frontier firms were more likely to invest in internal research in the first place. It may also be that frontier firms are less reliant on external inventions, leading to a larger scale of internal research and invention.

5.6 Rival Use of Public Science

Table 9 presents the estimation results for the relationship between rival-relevant public science (including actual use of public knowledge and rival-relevant PhD dissertations) and the focal firm's market value. Columns 1-2 present OLS estimation results. In Column 3, we use *Predicted rival-relevant PhD dissertations* as an instrument for *Rival-relevant PhD*

Table 8: Variation by Firm Proximity to the Technology Frontier: Patent Value

	(1) ln(Patents)	(2) ln(Pubs.)	(3) ln(AMWS Scientists)	(4) ln(Patents)	(5) ln(Pubs.)	(6) ln(AMWS Scientists)
ln(Relevant publications)$_{t-1}$ $\times$ *Tech frontier*	0.121	0.014	0.003			
	(0.003)	(0.002)	(0.001)			
ln(Relevant PhD dissertations)$_{t-1}$ $\times$ *Tech frontier*				0.094	0.020	0.008
				(0.005)	(0.004)	(0.003)
ln(Relevant publications)$_{t-1}$	-0.117	-0.109	-0.030	-0.029	-0.005	-0.002
	(0.020)	(0.018)	(0.011)	(0.004)	(0.003)	(0.002)
ln(Relevant PhD dissertations)$_{t-1}$	0.094	0.063	0.023	0.209	0.069	0.034
	(0.012)	(0.010)	(0.006)	(0.022)	(0.016)	(0.011)
ln(R&D stock)$_{t-1}$	0.193	0.142	0.033	0.171	0.130	0.026
	(0.014)	(0.013)	(0.007)	(0.014)	(0.013)	(0.008)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Mean DV	28.30	15.72	4.80	28.30	15.72	4.80
Weak identif. (Kleibergen-Paap)	178.25	178.25	178.25	546.33	546.33	546.33
Firms	3,371	3,371	3,371	3,371	3,371	3,371
Observations	41,693	41,693	41,693	41,693	41,693	41,693

Notes: This table presents the second stage of 2SLS estimation for the effect of public knowledge and human capital on firm patents, publications, and AMWS scientists when considering firm proximity to the technology frontier. We measure *Tech frontier* using a firm-cohort measure based on patent values from Kogan et al. (2017). We use *Windfall-predicted R&D budget* and its interactions with *Tech frontier* as instrumental variables for *Relevant publications* and its interactions with *Tech frontier*. We also use *Advisors' windfall-predicted R&D budget* and its interactions with *Tech frontier* as instrumental variables for *Relevant PhD dissertations* and its interactions with *Tech frontier*. One is added to logged variables. Standard errors (in parentheses) are robust to arbitrary heteroskedasticity and allow for serial correlation through clustering by firms.

dissertations, and *Predicted rival-relevant publications* as an instrument for *Rival citations to publications*. Consistent with the theoretical predictions from Online Appendix Table A1, firm value decreases with rival use of public knowledge for both OLS and 2SLS (Columns 2 and 3). However, the effect of *Rival-relevant PhD dissertations* is effectively zero.

So far, we have defined rivals as panel firms operating in the same Text-based Network Industry Classification (Hoberg & Phillips, 2010, 2016). An alternative set of rivals consists of small firms. In Column 4, we also consider the use of public knowledge by small rivals (i.e., standalone firms with fewer than 50 employees and within 10 years of incorporation that are included in the NETS High Tech 2020 database). We use *Predicted small rival-relevant publications* as an instrument for *Small rival citations to publications*. The 2SLS coefficient estimates suggest that the use of public knowledge by all rivals, whether large or small, decreases focal firm value.

In Column 5, we restrict our attention to a subsample of manufacturing firms, for which we can calculate product market distances between firms using pre-period sales in various business segments. Even after we control for spill-ins of knowledge from other firms to the focal firm (using the product market distance-weighted sum of all other firms' R&D stocks), we find that rival use of public knowledge has a negative effect on focal firm value. Results remain unchanged when we define rivals as firms operating in the same SIC4 industry as the focal firm (Column 6), suggesting a negative effect of competition.

Table 9: Firm Value Equation with Rival Use of Public Science

	(1)	(2)	(3)	(4)	(5)	(6)
	ln(Tobin's Q)					
	OLS (Baseline)	OLS (Add corp. rivals)	2SLS (Instrument for corp. rivals)	2SLS (Add instrument for small rivals)	2SLS (Add control for rival spill-ins)	2SLS (Alternative definition of rivalry based on SIC4 overlap)
Own-relevant publications$_{t-1}$ / $Assets_t$	0.018	0.016	0.017	0.016	0.014	0.014
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Own-relevant PhD dissertations$_{t-1}$ / $Assets_t$	-0.005	-0.002	-0.002	-0.002	-0.002	-0.004
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
ln(Rival citations to publications)$_{t-1}$		-0.107	-0.105	-0.101	-0.118	-0.051
		(0.013)	(0.017)	(0.019)	(0.019)	(0.019)
ln(Rival-relevant PhD dissertations)$_{t-1}$		0.001	0.002	0.002	0.006	0.009
		(0.004)	(0.005)	(0.005)	(0.005)	(0.008)
ln(Small rival citations to publications)$_{t-1}$				-0.008	0.010	-0.011
				(0.013)	(0.014)	(0.014)
R&D stock$_{t-1}$ / $Assets_t$	0.201	0.208	0.207	0.208	0.205	0.204
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)
ln(SPILLSIC R&D stock)$_{t-1}$					-0.162	-0.223
					(0.092)	(0.095)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Mean DV	33.80	33.80	33.80	33.80	28.51	28.51
Weak identif. (Kleibergen-Paap)			1,590.92	942.62	901.30	744.81
Firms	3,230	3,230	3,230	3,230	2,871	2,871
Observations	36,718	36,718	36,718	36,718	33,975	33,975

Notes: This table presents estimation results for the relationship between rival-relevant public science and Tobin's Q. In Columns 1-5, *ln(Rival citations to publications)* and *ln(Rival relevant PhD dissertations)* are constructed using the Text-based Network Industry Classifications and pairwise product portfolio similarity scores from Hoberg and Phillips (2010, 2016). In Columns 3-5 we use *Predicted rival-relevant PhD dissertations* as an instrument for *Rival-relevant PhD dissertations* and *Predicted rival-relevant publications* as an instrument for *Rival citations to publications*. In Columns 4 and 5 we also use *Predicted small rival-relevant publications* as an instrument for *Small rival citations to publications*. Column 5 is restricted to manufacturing firms. *SPILLSIC R&D stock* is the product market distance-weighted sum of all other firms' R&D stocks. In Column 6, we use alternative measures for *ln(Rival citations to publications)* and *ln(Rival relevant PhD dissertations)* based on SIC4 industry definitions. One is added to logged variables. Standard errors (in parentheses) are robust to arbitrary heteroskedasticity and allow for serial correlation through clustering by firms.

In Table 10, we turn our attention to the effect of rival-relevant public science on focal firm innovation. Unlike the effect on firm value, rival use of public knowledge does not affect corporate patenting, publishing, and hiring, on average. These results are consistent with changes in rival innovation not affecting the focal firm's marginal returns to innovation. In turn, this suggests that strategic interactions in innovation between rivals are modest.

Table 10: Patents, Publications, and AMWS Scientists Equations with Rival Use of Public Science

Dependent variable:	(1) ln(Patents)	(2) ln(Publications)	(3) ln(AMWS scientists)
ln(Own-relevant publications)$_{t-1}$	-0.201	-0.124	-0.035
	(0.024)	(0.019)	(0.011)
ln(Own-relevant PhD dissertations)$_{t-1}$	0.337	0.143	0.054
	(0.032)	(0.020)	(0.013)
ln(Rival citations to publications)$_{t-1}$	-0.017	-0.002	-0.002
	(0.013)	(0.011)	(0.007)
ln(Rival-relevant PhD dissertations)$_{t-1}$	0.023	0.017	0.006
	(0.003)	(0.002)	(0.002)
ln(R&D stock)$_{t-1}$	0.186	0.123	0.025
	(0.016)	(0.014)	(0.008)
Year FE	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
Mean DV	28.30	15.72	4.80
Weak identif. (Kleibergen-Paap)	92.21	92.21	92.21
Firms	3,371	3,371	3,371
Observations	41,693	41,693	41,693

Notes: This table presents the second stage of 2SLS estimation results for the relationship between rival-relevant public science and firm patents, publications, and AMWS scientists. *ln(Rival citations to publications)* and *ln(Rival relevant PhD dissertations)* are constructed using the Text-based Network Industry Classifications and pairwise product portfolio similarity scores from Hoberg and Phillips (2010, 2016). We use *Windfall-predicted R&D budget*, *Advisors' windfall-predicted R&D budget*, *Predicted rival citations to publications*, and *Predicted rival-relevant PhD dissertations* as instruments for *Own-relevant publications*, *Own-relevant PhD dissertations*, *Rival citations to publications*, and *Rival-relevant PhD dissertations*, respectively. One is added to logged variables. Standard errors (in parentheses) are robust to arbitrary heteroskedasticity and allow for serial correlation through clustering by firms.

6 Conclusion

This paper provides systematic evidence that most firms, across various industries, respond to an increase in relevant public knowledge by employing fewer AMWS scientists and producing fewer publications and patents. Conversely, firms respond to an increase in relevant human capital by employing more AMWS scientists and producing more publications and patents. We find that the disembodied and embodied components of public science appear to affect the

R&D investments of firms differently. Whereas public knowledge represented by publications appears to substitute for internal research by corporations, the trained scientists that are often produced alongside scientific knowledge enhance the payoffs to internal research by corporations. These potentially offsetting effects imply that the net effect of public science on corporate innovation may well be small. However, they also illuminate the different ways in which public and private R&D investments interact. Moreover, they suggest that firms may respond to rising public science differently depending on how close they are to the technology frontier.

Indeed, we find that firms respond differently to the rise in public knowledge based on their proximity to the technology frontier. Frontier firms continue to invest in internal research and invention even when public knowledge becomes abundant. The surge in corporate scientific research in several emerging technology fields (e.g., artificial intelligence and quantum computing) is consistent with these empirical findings. These differences may arise because the marginal returns from using internal research in innovation are greater for frontier firms than the marginal returns enjoyed by other firms. Put differently, public knowledge is less likely to substitute for internal research in frontier firms, whereas followers are more likely to use public knowledge to power their internal inventive activity. A complementary explanation is that external inventions generated from public science are less likely to substitute for internal inventions in frontier firms, whereas followers may be more reliant on such eternal inventions.

Understanding the mechanisms behind the effects of embodied and disembodied public science on corporate research and invention is important for understanding the evolution of the American innovation ecosystem, in which publicly-funded public knowledge and startups have steadily grown in importance as sources of innovation, at the expense of established corporate labs of yesteryears. To the extent that diversity in the sources and types of R&D activities is important for a vibrant innovation ecosystem, a reduction in corporate research could be problematic. Our results also point to the possibility that the expansion of public science may widen the gap between firms. Increases in the gap between frontier firms and followers have implications for product market competition, and for the rate and direction of technical advance, more broadly.

References

Acs, Z. J., Audretsch, D. B., & Feldman, M. P. (1994). R & D spillovers and recipient firm size. *The Review of Economics and Statistics*, *76*(2), 336–340.

Adams, J. D. (1990). Fundamental stocks of knowledge and productivity growth. *Journal of Political Economy*, *98*(4), 673–702.

Adams, J. D., Chiang, E. P., & Jensen, J. L. (2003). The influence of federal laboratory R&D on industrial research. *The Review of Economics and Statistics*, *85*(4), 1003–1020.

American Association for the Advancement of Science. (2020). *Historical Trends in Federal R&D.* (Available at https://www.aaas.org/programs/r-d-budget-and-policy/historical-trends-federal-rd. Accessed May 27, 2021.)

Arora, A., Belenzon, S., & Patacconi, A. (2018). The decline of science in corporate R&D. *Strategic Management Journal*, *39*(1), 3–32.

Arora, A., Belenzon, S., & Sheer, L. (2021a). Knowledge spillovers and corporate investment in scientific research. *American Economic Review*, *111*(3), 871–98.

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

Azoulay, P., Graff Zivin, J. S., Li, D., & Sampat, B. N. (2019). Public R&D investments and private-sector patenting: Evidence from NIH funding rules. *The Review of Economic Studies*, *86*(1), 117–152.

Belenzon, S., & Cioaca, L. C. (2021). Guaranteed markets and corporate scientific research. *National Bureau of Economic Research Working Paper*(w28644).

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

Cohan, A., Feldman, S., Beltagy, I., Downey, D., & Weld, D. S. (2020). SPECTER: Document-level representation learning using citation-informed transformers. In *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 2270–2282).

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

Dasgupta, P., & David, P. A. (1994). Toward a new economics of science. *Research Policy*, *23*(5), 487–521.

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

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

Dugoua, E., Gerarden, T., Myers, K., & Pless, J. (2022). Creating and steering innovators with supply versus demand policies.

González, X., Jaumandreu, J., & Pazó, C. (2005). Barriers to innovation and subsidy effectiveness. *RAND Journal of Economics*, 930–950.

Goolsbee, A. (1998). Does government R&D policy mainly benefit scientists and engineers? *American Economic Review*, *88*(2), 298–302.

Hall, B. H., Jaffe, A., & Trajtenberg, M. (2005). Market value and patent citations. *RAND*

Journal of Economics, 16–38.

Hoberg, G., & Phillips, G. (2010). Product market synergies and competition in mergers and acquisitions: A text-based analysis. *The Review of Financial Studies*, *23*(10), 3773–3811.

Hoberg, G., & Phillips, G. (2016). Text-based network industries and endogenous product differentiation. *Journal of Political Economy*, *124*(5), 1423–1465.

Jaffe, A. B. (1989). Real effects of academic research. *American Economic Review*, 957–970.

Kim, S. D., & Moser, P. (2021). *Women in science. lessons from the baby boom* (Tech. Rep.). National Bureau of Economic Research.

Kogan, L., Papanikolaou, D., Seru, A., & Stoffman, N. (2017). Technological innovation, resource allocation, and growth. *The Quarterly Journal of Economics*, *132*(2), 665–712.

Lichtenberg, F. R. (1984). The relationship between federal contract R&D and company R&D. *American Economic Review*, *74*(2), 73–78.

Lucking, B., Bloom, N., & Van Reenen, J. (2018). Have R&D spillovers changed? *National Bureau of Economic Research Working Paper*(w24622).

Milgrom, P., & Roberts, J. (1990). Rationalizability, learning, and equilibrium in games with strategic complementarities. *Econometrica: Journal of the Econometric Society*, 1255–1277.

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

Mulligan, K., Lenihan, H., Doran, J., & Roper, S. (2022). Harnessing the science base: Results from a national programme using publicly-funded research centres to reshape firms' R&D. *Research Policy*, *51*(4), 104468.

Myers, K. R., & Lanahan, L. (2022). Estimating spillovers from publicly funded R&D: Evidence from the US department of energy. *American Economic Review*(forthcoming).

Narin, F., Hamilton, K. S., & Olivastro, D. (1997). The increasing linkage between US technology and public science. *Research Policy*, *26*(3), 317–330.

National Science Board. (1998). *Science and engineering indicators 1998* (Tech. Rep. No. NSB-1998-1). National Science Foundation. Retrieved from https://wayback.archive-it.org/5902/20150627201913/http://www.nsf.gov/statistics/seind98/

National Science Board. (2010). *Science and engineering indicators 2010* (Tech. Rep. No. NSB-2010-1). National Science Foundation. Retrieved from https://wayback.archive-it.org/5902/20160210151754/http://www.nsf.gov/statistics/seind10/

National Science Board. (2018). *Science and engineering indicators 2018* (Tech. Rep. No. NSB-2018-1). National Science Foundation. Retrieved from https://www.nsf.gov/statistics/2018/nsb20181/

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

Rosenberg, N., & Nelson, R. R. (1994). American universities and technical advance in industry. *Research Policy*, *23*(3), 323–348.

Sinha, A., Shen, Z., Song, Y., Ma, H., Eide, D., Hsu, B.-J., & Wang, K. (2015). An overview of Microsoft Academic Service (MAS) and applications. In *Proceedings of the 24th* *international conference on world wide web* (pp. 243–246).

Szücs, F. (2020). Do research subsidies crowd out private R&D of large firms? Evidence from European framework programmes. *Research Policy*, *49*(3), 103923.

U.S. Office of Management and Budget. (2022). *Budget of the United States Government: 1921-2021.* (Retrieved from https://fraser.stlouisfed.org/title/54 on March 14, 2022.)

Wallsten, S. J. (2000). The effects of government-industry R&D programs on private R&D: The case of the Small Business Innovation Research program. *The RAND Journal of Economics*, 82–100.

Wang, K., Shen, Z., Huang, C., Wu, C.-H., Eide, D., Dong, Y., ... Rogahn, R. (2019). A review of Microsoft Academic Services for science of science studies. *Frontiers in Big Data*, *2*, 45.