The Rise of Absorptive Research in Corporate America: 1945-1980*

Ashish Arora† Sharon Belenzon‡ Jungkyu Suh§ Hansen Zhang¶

Abstract

We study the post–World War II "Golden Age" of American corporate research from 1945 to 1980, using multiple indicators of corporate research activity. We use an ensemble learning approach to classify firms as either Science Leaders, Absorbers or Followers. Our analysis reveals that only a small fraction of firms, whom we call Leaders, invest in internal research on the scientific frontier, with the objective to generate breakthrough inventions. Absorbers, more frequent than Leaders, invest in research principally to absorb external scientific discoveries to fuel their inventive activity. Followers mostly generate incremental innovations using older scientific knowledge. Consistent with this, we find Leaders were more likely to be at the technological frontier, enjoy greater market power, and benefit from government procurement contracts. As universities and startups began to commercialize academic discoveries, the need for "absorptive corporate labs" declined. The shift ultimately transformed the American innovation landscape, deepening the division of innovative labor between universities, startups, and incumbent corporations, with only a select group of Leader firms continuing to invest in basic science.

JEL Classification: O30, O31, O32, N82, L22, L25

Keywords: Corporate R&D; Corporate Science; Absorptive Capacity
Ensemble Learning

*We thank Dan Gross for sharing the DoD Contract, NRSTP and NER data. We also acknowledge support from the Fuqua School of Business, Duke University, the Stern School of Business, New York University, and the Alfred P. Sloan Foundation for financial support. All errors are our own.

†Duke University, Fuqua School of Business and NBER (ashish.arora@duke.edu)

‡Duke University, Fuqua School of Business and NBER (sharon.belenzon@duke.edu)

§New York University, Stern School of Business (js13220@stern.nyu.edu)

¶Duke University, Fuqua School of Business (hansen.zhang@duke.edu)

1 Introduction

The mid-twentieth century is often described as the "Golden Age" of American corporate R&D. While fewer than half of patenting firms in America operated an industrial lab in 1927, over 80% of firms would be operating one by 1956. Novel products in polymers, transistors and integrated circuits, software, and antibiotics originated in these labs and were brought to market. In the process, leading firms such as DuPont, AT&T, Xerox, GE, IBM, and Merck pushed the boundaries of science, on par with the top academic departments in their respective fields. What is often overlooked in recounting the history of the pioneering firms is that a majority of firms opening research labs did so not to push the frontiers of science but to absorb external knowledge that was being supplied by an expanding American university sector backed by federal funding (Auerswald & Branscomb, 2004; Cohen & Levinthal, 1990; Rosenberg, 1990). As startups began commercializing university science, and universities themselves sought industrial partners, thanks to policy changes such as the Bayh-Dole Act, (Arora et al., 2020; Arora et al., 2022; Arora et al., 2001, 2004) "absorptive corporate labs" lost their purpose, contributing to a broader shift of American firms away from research in favor of development activities after the 1980s.[1] Even as most firms came to rely on this division of innovative labor with academia, leading firms continued to invest in internal research, particularly where public science lagged the frontier.

In this paper, we draw on newly developed data to study the rise of mid-century corporate labs in America. We distinguish between firms that conducted research mainly to absorb academic knowledge and those that pursued new science beyond what universities or startups could supply. Our evidence shows that although corporate labs were increasingly common, only a minority engaged in "bona fide" science production. Most functioned as "absorptive" interfaces, bridging academic discoveries and industrial applications. By tracking how these roles evolved over time, we reveal that the decline of absorptive labs - rather than a collapse of all science-producing facilities - explains much of the contraction in corporate research activity since the 1970s. As universities and startups improved at commercializing scientific discoveries, large firms could acquire innovations directly rather than maintaining costly in-house research capabilities. This deepened a division of innovative labor: universities and startups now not only generate but also package new knowledge for industry, while established firms focus on development.

A key empirical challenge to conducting our analysis is that producing new scientific knowledge and absorbing external discoveries typically draw on the same inputs — scientists and engineers — and yield many of the same observable outputs such as scientific publications and patents. Indeed, we find in our own data that only about one-third of corporate scientists are actually engaged in scientific research, whereas the remaining two-thirds specialize in management or the downstream commercialization of R&D. This overlap makes it difficult to disentangle firms that create new science from those that merely adapt or translate outside discoveries. To address this challenge,

[1] Between 1990 and 2019, for instance, American businesses reduced the share of their R&D budgets devoted to research from about 30% to roughly 21% (Arora, Belenzon, Cioaca, et al., 2023; Arora et al., 2018; Boroush, Mark, 2021; National Science Foundation, 2019).

we assemble a comprehensive dataset that tracks both the inputs and outputs of corporate R&D and use an ensemble learning approach (Dong et al., 2020; Mienye & Sun, 2022) that combines multiple deep learning algorithms to classify firms into three distinct groups: Science Leaders, Science Absorbers, and Followers that exclusively commercialize technology.

Our panel consists of 2,904 patenting firms from the CRSP database of listed U.S. firms spanning the years 1945 to 1980. We develop new measures of corporate scientific activities: First, we match over 107,000 scientific publications to 779 firms (27% of the sample) based on author affiliations. Second, we identify 971 firms (33%) as operating an industrial lab at some point in our sample period. Third, we match scientists' and engineers' employment records from the National Register of Scientific and Technical Personnel (NRSTP) and the National Engineers Register (NER) surveys to 902 sample firms (31%).

We present four sets of results. First, we identify a small set of Science Leaders, comprising just 1.3 percent of firms in our sample (38 firms). Science Absorbers make up 6.1 percent (176 firms), and the remaining 92.6 percent (2,690 firms) are classified as Followers.[2] Scientific activity is highly concentrated even within this select group of Leaders: DuPont alone employed 8.7 percent of all U.S. corporate scientists, while the top five firms (DuPont, AT&T, GE, IBM, and Shell Oil) accounted for 20.7 percent. Concentration is particularly stark in electronics, where the top five (AT&T, IBM, GE, Westinghouse, and RCA) employed 65.4 percent of corporate scientists, and in chemicals, where the top five (DuPont, Shell, Union Carbide, Monsanto, and Dow) employed 41.0 percent of chemist corporate scientists.

Second, our classification reveals three distinct motivations behind corporate investments in research. Leaders treat internal research as direct input for novel innovations; Absorbers treat internal research as a funnel to source external knowledge from universities or rivals; Followers focus on incremental product improvement. Indeed, Leader firms are closer to the technological frontier compared to Absorbers and Followers. Leaders generate 2,200 more breakthrough patents than Followers, whereas Absorbers produce 22 more. Furthermore, both Leaders and Absorbers patents are 15-17 times more likely to rely on external scientific knowledge compared to Followers. Leaders are three times as likely to cite their own scientific output compared to Absorbers, while Followers rarely build on internal science. Leaders also show significantly closer links to internal science measured through citation distances (2.3 times closer than Absorbers and 12 times closer than Followers) and a greater resemblance to academic research (1.8 times more similar than Absorbers and 8 times more similar than Followers). Leaders also draw more directly on recent scientific advances.

Leaders were more likely to emerge when appropriability concerns were mitigated: our third result is that Science Leaders exhibit greater market power than both Absorbers and Followers (Aghion & Tirole, 1994). On average, they hold 27.2 times the assets of Followers, while Absorbers hold 9.3 times as many assets, implying that Leaders are 2.9 times larger than Absorbers. Market

[2]Examples of Science Leaders include AT&T, DuPont, IBM, Merck, General Electric, and Texas Instruments. These companies employed scientists who published groundbreaking work that garnered both Nobel Prizes and the President's National Medal of Science during the Golden Age of corporate research.

share follows a similar pattern: Leaders' shares exceed those of Followers by 3.6 percentage points, while Absorbers show only a 0.6 point advantage. Leaders are also more likely to be found in concentrated industries.

Our fourth set of results pertains to the contribution of the federal government. Reflecting on the lessons from wartime projects such as the Radar and the Manhattan Project and responding to the Sputnik shock, the federal government increased its subsidy for the production of human capital and knowledge in academia and became the buyer of the first resort for risky technologies with uncertain markets. While all firms gained from the expanded supplies of human capital, Leaders and Absorbers especially benefited from federal R&D and procurement policies. Leaders have 176% more government patents than Followers, while Absorbers have around 27% more.[3] We also find that Leaders exceed Followers by 11.1 percentage points in Department of Defense contract values normalized by sales (Absorbers: 8.0 points), adjusting for firm size and industry. Leaders are also 29.6 percentage points more likely to operate government-sponsored laboratories (Absorbers: 7.6 points).

We make several contributions to the literature. First, our findings provide a clue to why the Golden Age of corporate R&D gave way to a secular decline starting from the 1980s. Today's industrial innovation landscape presents a seemingly contradictory picture: on the one hand, many companies have shifted from research to development, with former Science Absorbers trimming their R&D to the point of resembling mere Followers; on the other hand, a few "superstar firms" have intensified their research efforts in domains such as large language models, robotics, autonomous vehicles, and quantum computing (Autor et al., 2020; De Loecker et al., 2020). Our results offer a potential reconciliation. Even in the mid-twentieth century, relatively few firms were true Science Leaders; the rest maintained labs largely to absorb external scientific knowledge (Absorbers) or attract skilled scientists and engineers (Followers). Over the past four decades, the rise of markets for technology (MFT) appears to have diminished the need for these "absorptive labs", prompting many former Science Absorbers to scale back research. Meanwhile, industrial research production has migrated from chemicals and electronics to newly emerging fields — AI, quantum computing, and robotics — where proprietary research can be of strategic and commercial value.

Second, by examining multiple R&D indicators simultaneously, our ensemble approach provides a more nuanced understanding of how firms engage with science. Consider two firms with ostensibly robust R&D metrics: one generates many patents but publishes little, whereas the other produces numerous scientific articles but relatively fewer patents. A single-indicator method might rate both firms similarly, missing the underlying differences in how they source and generate knowledge. In contrast, by blending various metrics — publications, patents, lab infrastructure, and scientist rosters — and allowing for non-linear interaction among the metrics, our ensemble models can detect distinct patterns. These have revealed, for example, that the first firm specialized in applying external knowledge and optimizing for marketable inventions, while the second

[3]Government patents refer to patents are either licensed to or directly held by government agencies (Gross & Sampat, 2024).

prioritized more open-ended, foundational research. This refined categorization is particularly valuable for understanding the historical evolution of corporate R&D and for informing contemporary strategies to foster innovation in different industrial settings.

Third, our finding that the largest firms tend to invest the most in advancing science underscores an important tension with current policy debates on monopoly power. In highly competitive markets, companies often lack the financial slack and long-term planning horizons necessary to support ambitious research. By contrast, firms with greater market power can afford significant R&D expenditures and are more likely to undertake science-intensive projects. This suggests that policymakers may need to tolerate some degree of concentration among "superstar" firms to maintain a robust ecosystem with industrial research. Striking the right balance between encouraging competition and preserving incentives for scientific investment is crucial: excessive market dominance can stifle innovation, but if regulations are too stringent, they might discourage the very research capabilities that drive breakthrough inventions.

The paper is organized as follows. Section 2 surveys the historical context; Section 3 describes the data; Section 4 presents the tripartite classification of corporate R&D strategies; Section 5 explains the causes and consequences of these divergent strategies. Section 6 concludes.

2 Historical Developments: 1945-1980

The rise of corporate participation in research after World War II was fueled by a confluence of supply- and demand-side factors. On the supply side, federal investments — most notably through the G.I. Bill, the National Science Foundation (established in 1950), and increased military and health-related grants — greatly expanded university research capacities, producing a larger pool of scientific talent and a surge in fundamental discoveries. Industrial research grew rapidly as a result. Between 1954 and 1970, the private sector employed 205% more science and engineering doctorates, with numbers rising from 11,181 to 22,915 (Figure 3). As illustrated in Panel (a) of Figure 1, the proportion of patenting firms operating dedicated research laboratories more than doubled — growing from 37% in 1927 to 83% by 1956.

[Figure 1 Here]

Despite this expansion on the extensive margin, firms producing scientific results pushing the boundaries of human knowledge were few and far between. Rather, the corporate research programs established during this period were largely tasked with absorbing and translating outside knowledge. An ascendant U.S. academic sector generated a steady flow of scientific discoveries, providing corporations with novel ideas to commercialize only if they possessed the ability to absorb and refine them. The "golden age of antibiotics" exemplifies this model, where pharmaceutical companies like Pfizer, Merck, and Eli Lilly systematically absorbed academic discoveries in microbiology and biochemistry, then developed and refined them into commercially viable medical treatments (Lee, 2003).

A small minority of leading firms such as AT&T and DuPont maintained research programs on par with the best research universities, partly because universities still often lagged the frontier in polymers, computer science, and solid state electronics. The "Leader" firms were also intimately involved in government procurement contracts for defense and space exploration, which created robust markets for innovation-intensive products such as IC chips and diode lasers. High-volume contracts and long-term funding commitments by the government allowed management at leading labs to tolerate fundamental science that promised blockbuster products.

This section therefore begins by characterizing the three approaches to corporate science - Leader, Absorber, and Follower - that had emerged by the 1960s. The second half then hypothesizes that a growing academic base, ample public funding, and demand for high-tech solutions contributed to this deepening trisection of R&D during the golden age of corporate science.

2.1 Expansion of Industrial Research

Even as firms began investing in science after 1945, the share of firms sustaining research at a scale sufficient to advance the scientific frontier stayed remarkably stable. For instance, the prevalence of very large corporate labs, indicated by labs employing over 100 scientists, reached 20% by 1946 but then plateaued — fluctuating between 18% and 29% through the next three decades. Panel (b) of Figure 1 confirms this pattern with corporate publications, showing that the share of firms publishing in scientific journals hovered around 25-30% throughout the period. Moreover, there is high concentration in the production of scientific knowledge: only 2-4% of firms publish 50 or more papers per year, while those producing high-impact publications never exceed 9%.

[Figure 2 Here]

The distribution of scientific activities among American corporations during 1945-1980 shows that levels of concentration grow as one moves closer to scientific knowledge production. Figure 2 illustrates this gradient through C4 concentration ratios: the top four firms during our sample period, for instance, own 18% of all corporate assets; research inputs such as industrial scientists demonstrate similar distributions (C4 = 19%). However, when examining research outputs, concentration increases substantially — The top four patenters account for 32% of corporate patents, and the top four in scientific publications account for 53% of all corporate publications. High-impact publications show even greater concentration at 65%. At the extreme, the top four firms account for all prestigious scientific awards given to corporate scientists (C4 = 1.00). Based on a historical reading of labs operated during this time, we conjecture three approaches to corporate science.

First, some leading laboratories positioned themselves as Leaders of fundamental scientific knowledge, with AT&T's Bell Laboratories, General Electric, and IBM as leading examples. Bell Laboratories became the premier industrial research institution through significant contributions to both basic science and technology applications. Their work in solid-state physics led to the transistor, transforming electronics. Shannon's mathematical theory of communication (Shannon, 1948) established foundations for digital communications. The lab later expanded into quantum

optics with laser spectroscopy research (Schawlow & Townes, 1958) and made contributions to cosmology by discovering cosmic microwave background radiation (Penzias & Wilson, 1965).

The significance of these scientific achievements during 1945-1980 is reflected in a concentration of Nobel Prizes in Physics. The laboratory's fundamental contributions were recognized through the 1956 Physics Prize awarded to Bardeen, Brattain, and Shockley for their 1947 invention of the transistor (Bardeen & Brattain, 1948). The 1978 Physics Prize shared by Penzias and Wilson for their 1965 discovery of cosmic microwave background radiation (Penzias & Wilson, 1965), and the 1981 Prize was awarded to Schawlow for his 1958 contributions to the development of laser spectroscopy (Schawlow & Townes, 1958). This remarkable clustering of Nobel-winning research, all conducted within a roughly twenty-year period at a single industrial laboratory, underscores Bell Labs' unique position in the history of industrial research and exemplifies the profound integration of fundamental scientific inquiry with technological innovation (Gertner, 2012; Nelson, 1962).

General Electric's research laboratory demonstrated exceptional strength across multiple scientific domains, particularly in physics and information systems. Their breakthrough research in superconductivity, notably Giaever's work on electron tunneling in superconductors (Giaever, 1960) advanced the fundamental understanding of quantum mechanics. GE's scientific excellence extended into computer science, recognized by Charles Bachman's 1973 Turing Award for pioneering work in database technology (Haigh, Thomas, 1973). In applied research, GE achieved a significant milestone with the development of the first civilian nuclear power plant (GE Vernova, 2024), exemplifying their capacity to translate scientific understanding into technological innovation.

IBM, meanwhile, emerged as the undisputed Leader in computer science and information technology, making seminal contributions to both theoretical foundations and practical implementations. Its development of the hard disk drive (Noyes & Dickinson, 1956) represented a fundamental advance in data storage technology, while their work in cryptography (Feistel, 1973) established essential principles for modern computer security. The laboratory's theoretical contributions to computational methods, particularly the Fast Fourier Transform (Cooley & Tukey, 1965), demonstrated their ability to advance both mathematical foundations and practical applications. IBM's scientific preeminence was recognized through prestigious awards, with John Backus receiving the Turing Award in 1977 for developing FORTRAN and programming language theory (Booch, Grady, 1977), and Kenneth E. Iverson being honored in 1979 for his work on mathematical notation and programming languages (Smillie, Keith, 1979).

Bell Labs, GE, and IBM, however, were the exceptions rather than the norm. Nelson and Langlois's historical analysis of industrial innovation (Nelson & Langlois, 1983) demonstrates that even during periods of widespread R&D expansion, meaningful scientific production remained concentrated among a select group of firms. Instead, firms hired scientists for various purposes, including better absorption of outside knowledge (Cohen & Levinthal, 1989). Scientific research in such cases focused primarily on applying existing scientific knowledge, serving as "listening posts" for research from leading firms and universities.

The case of Koppers Company (1945-1980) provides a revealing illustration of this application-

focused approach. Despite an impressive research infrastructure — 413 industrial laboratory scientists, 25 scientists listed in American Men & Women of Science, moderate annual publications (2.64), and substantial patents (46) — closer examination reveals that Koppers' resources were primarily directed toward applying established scientific principles rather than producing fundamental advances. In coal chemistry and materials science, their coal-gasification processes, documented in "Modern Gas Manufacturing/SNG Process Description/Koppers K-T Coal-Gasification Process" (), applied established thermodynamic principles and reaction kinetics to optimize coal-to-gas conversion. Similarly, their polymer science research exemplified this pattern of practical application over theoretical innovation, as evidenced by citation pattern analysis of Koppers' publications. Their 1966 paper "Axial Dispersion of Polymer Molecules in Gel Permeation Chromatography" by Hess and Kratz provides a clear example where the primary contribution is methodological — refining and improving gel permeation chromatography rather than developing new theoretical principles. The authors adapt existing dispersion models from chemical engineering to solve a practical problem: correcting chromatographic data for more accurate polymer characterization. Citation data confirms this applied orientation: the paper received 88 citations predominantly from applied technical journals — 23 in *Journal of Applied Polymer Science*, 8 in *Journal of Polymer Science Part C: Polymer Symposia*, and 7 in *Journal of Liquid Chromatography.* The paper itself relies heavily on practical information sources, with 5 of 15 references being private communications or technical data sheets. Rather than proposing fundamentally new theoretical frameworks, the authors emphasize practical necessity, stating their goal as addressing "an error for which correction is needed" in polymer analysis. Conducted within Koppers' industrial research setting, the work demonstrates a strong emphasis on data-driven correction rather than theoretical abstraction, testing the dispersion correction method on standard polystyrene samples to provide solutions directly usable by scientists and engineers.

This pattern of knowledge absorption is particularly evident in Koppers' patent filings, which frequently built upon external academic and industrial research. For example, Koppers' patent "Production of benzophenones" (US 3403183A), assigned to researchers Elmer Dobrotz and Alfred Kolka from Pittsburgh, directly leveraged scientific knowledge developed at nearby Pennsylvania State College. The patent explicitly cites work by Simons, Randall, and Archer published in the *Journal of the American Chemical Society* on "Hydrogen Fluoride as a Condensing Agent," (Simons et al., 1939) adapting their fundamental research on hydrogen fluoride chemistry to develop improved manufacturing processes for benzophenones.

In the field of polymer chemistry, Koppers' patent "Self-extinguishing expandable polymers containing volatile hydrocarbon, organic bromide, and n-nitroso secondary amine" (US 3274133A), inventor Paul Ingram explicitly cites research by J. Eichhorn of Dow Chemical Company published in *the Journal of Applied Polymer Science.* Eichhorn's work on "Synergism of free radical initiators with self-extinguishing additives in vinyl aromatic polymers" (Eichhorn, 1964) provided the scientific foundation that Koppers researchers then adapted and refined for commercial applications. The patent directly builds upon Eichhorn's discovery of synergistic effects between free radical

initiators and halogen compounds, applying this knowledge to develop improved flame-retardant polymer formulations.

Similarly, in Koppers' patent "Manufacture of Melamine" (US 2559617A), inventor Raymond Hartigan built upon foundational research conducted by P.P. McClellan of American Cyanamid Company. The patent incorporated knowledge from McClellan's published work on melamine preparation in the *Industrial and Engineering Chemistry* (Mccellan, 1940), specifically adapting processes for converting cyanuric halides to melamine through amination reactions. Rather than developing entirely new chemical pathways, Koppers refined existing processes by modifying reaction conditions and engineering more efficient production methods.

The company's role as an Absorber rather than Leader of science is further evidenced by its minimal presence in elite scientific achievements — only 2 members of the American Association for the Advancement of Science and no scientific awards. Moreover, while firms that produce science typically cite their own scientific work in patents, Koppers' patents make no mention of its science. These signals collectively suggest that while Koppers was highly skilled at applying established scientific principles to practical problems, it was not a significant Leader of scientific advances. This systematic approach to leveraging external scientific advances while focusing on process optimization and commercial applications exemplifies the absorptive model employed by many industrial research laboratories during this period.

The third approach to corporate science focused primarily on development and testing activities, employing scientists and technical staff for applying older knowledge rather than absorbing the latest scientific research. The National Biscuit Company (Nabisco) provides a telling example of this "Follower" approach. Despite maintaining a large research center in Fair Lawn, New Jersey with impressive facilities — including pilot plants and machine shops — and a substantial technical staff of 205 personnel (52 chemists, 30 mechanical engineers, and 94 technicians in 1960), the company's activities were firmly focused on routine development and testing rather than research.

While Nabisco reported to dedicate 75% of its activities to R&D and only 25% to testing and experimentation in the 1965 Industrial Research Laboratories directory, closer examination reveals these activities were primarily routine product development. The company's 1960 Annual Report shows its labs focused on practical manufacturing concerns: mixing and baking techniques, ingredient handling, and packaging methods ("National Biscuit Company (Nabisco) Annual Reports", 1960). A typical project during this period involved flour milling technology at the company's mills in Toledo, Ohio; Carthage, Missouri; and Cheney, Washington, which produced over 80% of the company's flour requirements. By 1965, Nabisco had established a Research Committee to set goals and priorities based solely on commercial potential, with little consideration for applying the latest scientific knowledge ("National Biscuit Company (Nabisco) Annual Reports", 1965).

The company's technical output from the 1940s-1980s confirms this development-centric approach. Their publications in the *Journal of the Association of Official Agricultural Chemists International* were limited to routine quality control procedures like measuring sugar content and moisture levels in baked goods — standard industry practices requiring no advanced scientific

knowledge. Technical studies such as "Correlation of keeping properties of shortenings with keeping quality of biscuits" (Maveety, 1945) and "The Change in Thiamin Content of Breakfast Cereals During Cooking" (White et al., 1946) consisted mainly of routine testing procedures without engaging with university science. Similarly, publications in the *Journal of Food Protection* focused solely on implementing standard industrial sanitation practices and regulatory compliance (Barron, 1959), showing minimal engagement with advanced microbiological or chemical principles.

This pattern aligned with Nabisco's stated objectives in its 1960 annual report of focusing on "better products, better ways of making them, and better methods of getting them to consumers." Despite substantial laboratory facilities and technical staff, Nabisco's activities were confined to routine product development and quality control, demonstrating that even large labs with significant resources could operate primarily in the "D" of the R&D spectrum.

The stark contrasts among these three approaches to industrial research — from Bell Labs' Nobel Prize-winning discoveries to Koppers' sophisticated application of existing science to Nabisco's focus on routine development — raise fundamental questions about the organization of American industrial research during this period. What determined whether a firm would invest in producing fundamental science like Bell Labs, focus on applying existing scientific knowledge like Koppers, or concentrate on routine development like Nabisco? How prevalent was each type of research organization, and what factors influenced a firm's choice among these approaches?

2.2 Expansion of Federal Support for Corporate Research

The wartime development of radar, sonar, mass-produced penicillin, nylon parachutes and the atomic bomb convinced American policymakers of the importance of national scientific strength as a force multiplier in future conflicts. To their alarm, the onset of the Cold War seemed to show that the U.S. gap with the Soviet Union was closing rapidly, first with the detonation of the atomic bomb by the Soviet Union in 1949 and the hydrogen bomb in the 1950s. The launch of Sputnik in 1957 fuelled fears that the U.S. was scientifically surpassed by its rival. The National Defense Education Act (NDEA) of 1958 reinforced the development of scientific human capital that began with the postwar Servicemen's Readjustment Act of 1944 (otherwise known as the "G.I. Bill"). A dramatic expansion of human capital in American higher education supplied the scientists and engineers needed to staff all three types of industrial laboratories. Between 1949 and 1969, the number of degree-granting higher education institutions grew 36%, from 1,851 to 2,525, while total student enrollment grew more than threefold from 2.4 million to 8 million; the number of doctorates produced grew 4.6 times from 6,420 to 29,912.[4]

[Figures 3 and 4 Here]

Figure 4 illustrates how federal and state funding fueled this expansion. Adjusted to 2017 dollars, federal support for higher education jumped from $2.2 billion in 1945 to $15.6 billion in 1970, and further to $19.7 billion by 1980. Over the same period, state funding surged from $2.5

[4]Data for this analysis are sourced from the Biennial Surveys of Education in the United States.

billion to $28.7 billion in 1970, and to $46.7 billion by 1980. This substantial public investment created a robust foundation for training successive generations of scientists and engineers.

Firms absorbed these personnel to staff their research labs. As science and engineering doctorates expanded from 47,089 to 132,042 between 1954 and 1970, the share of those working at private corporations stayed steady between 19-28% (Figure 3).[5] The NRSTP data reveals a significant growth in industry-based research scientists over this period, with R&D scientists increasing from 12,239 in 1958 to 19,849 by 1970. Within this group, those specifically designated as research scientists grew from 9,293 to 15,036 over the same time. This represents a 62% increase in industry R&D scientists and a corresponding 62% growth in dedicated research scientists.

2.2.1 Expansion of Academic Knowledge

Whereas the growth of human capital lowered the costs of running labs for all firms, the expansion of academic knowledge had diverging effects: Follower firms that had little interest in leveraging the latest scientific discoveries were affected the least; Leader firms that were on par with or ahead of universities were also less sensitive; Absorber firms that built on outside knowledge, on the other hand, were most responsive to this rise.

The federal government redefined its role in funding research by founding the National Science Foundation (1950) and augmenting the National Institutes of Health (NIH) through the 1944 Public Health Service Act.[6] The government also reorganized the National Advisory Committee on Aeronautics and the Advanced Research Projects Agency, launching NASA and ARPA (later DARPA) in 1958, further reinforcing the federal government's central role in shaping scientific and technological progress. The federal laboratories established for World War II were redirected toward peacetime research and housed within universities as Federally Funded Research and Development Centers (FFRDCs).[7] These centers made significant contributions across a range of technologies. MIT's Radiation Laboratory, pivotal for wartime radar, evolved into Lincoln Laboratory, where early work on the Semi-Automatic Ground Environment (SAGE) helped to establish the foundations of modern computer networking (Fagen, 1978; MIT Lincoln Laboratory, 2025; Stambaugh, 2024). At Caltech's JPL, rocket research broadened into pioneering space exploration programs — including the Viking and Voyager missions (Jet Propulsion Laboratory, 2025; Redmond & Smith, 1983) — while Brookhaven National Laboratory, founded in 1947, introduced the Brookhaven Graphite Research Reactor to advance peaceful nuclear technology, making substantial strides in nuclear medicine and diagnostic imaging (Brookhaven National Laboratory, 2025a, 2025b, 2025c). Argonne National Laboratory, stemming from the University of Chicago's Metallurgical

[5] Workplace affiliations are taken from the National Register of Scientific and Technical Personnel (NRSTP) survey.

[6] A pivotal provision extended the National Cancer Institute's grants program to all of NIH by 1946. This extramural funding mechanism grew rapidly, from $4 million in 1947 to over $100 million by 1957, then surpassing $1 billion by 1974. The full NIH budget similarly expanded from $8 million in 1947 to over $1 billion by 1966. The period from 1955 to 1968, under the Leadership of Director James A. Shannon, came to be known as NIH's "golden years" (Office of NIH History and Stetten Museum, 2025).

[7] These included the Jet Propulsion Laboratory (JPL) at Caltech, Lincoln Laboratory at MIT, and the atomic research facilities at Brookhaven, Argonne, and Lawrence Livermore (Gross & Sampat, 2023).

Laboratory, led the development of nuclear reactors, most notably designing the propulsion system for the USS Nautilus (the first nuclear-powered submarine) (Argonne National Laboratory, 2025; Nuclear Engineering Division, Argonne National Laboratory, 2013). Lawrence Livermore National Laboratory, initially focused on nuclear weapons, later diversified into high-performance computing and fusion energy; its Shiva and Nova laser programs helped to pioneer inertial confinement fusion (Lawrence Livermore National Laboratory, 2025a, 2025b).

The immigration of top-notch scientists escaping Europe also helped (Moser et al., 2014). American universities expanded their scholarly output from approximately 2,000 natural science and engineering publications in 1925 to 44,000 by 1970, representing a 22-fold increase.[8] American science, once relegated to catching up to the European "frontier" up until the 1920s, established itself as a global Leader in scientific research.[9] Though abundant, the scientific knowledge produced by American universities would have been difficult to absorb for firms without prior investments. The incentives of Absorber firms to expand their research, therefore, likely grew together with the expansion of the American university sector.

2.2.2 Growing Market for Science-based Products

During World War II, American policymakers employed demand-side interventions to foster corporate innovation: the Sputnik shock reignited this mechanism. An expanded system of research contracts and defense procurement offered both direct support for foundational R&D and stable markets for advanced technologies. By reducing uncertainty about future demand, these programs incentivized most the Leader firms undertaking ambitious technological projects.

[Figure 5 Here]

The scale and scope of this federal commitment to technological advancement is illustrated in Figure 5, which tracks federal R&D obligations across different performing sectors from 1955 to 2022.[10] The data reveals a dramatic surge in total federal R&D funding during the post-Sputnik period, with industry emerging as the largest non-governmental performer. Federal funding to industry peaked in the mid-1960s at approximately $56.3 billion (in 2017 dollars), coinciding with the

[8] Publication data were collected from Web of Science and Dimensions databases using the OECD's Frascati Manual 2015 classification system. We included publications in the fields of Natural Sciences", Engineering and Technology", Medical and Health Sciences", and Agricultural Sciences", while excluding Social Sciences" and Humanities". The dataset was further refined to include only publications affiliated with U.S. universities. The Frascati Manual, available at https://www.oecd.org/en/publications/2015/10/frascati-manual-2015_g1g57dcb.html, provides standardized classifications for research activities and has been maintained by the OECD's National Experts on Science and Technology Indicators (NESTI) Group since 1963.

[9] U.S. institutions accounted for only 0-13% of scientific Nobel laureates in the early 1900s, their share rose dramatically to reach 52% in 1950-1954 and maintained a high level of 55-75% through the latter half of the century (diamond marked line in Figure 4.

[10] Obligations represent federal financial commitments through both grants and contracts, including amounts for orders placed, contracts awarded, services received, and similar transactions during a given period. These commitments are counted when the obligation is made, regardless of when funds were appropriated or when future payment occurs.

height of the Space Race and major defense initiatives such as the Minuteman Intercontinental Ballistic Missile program, the Polaris submarine-launched ballistic missile system, the development of advanced radar systems like SAGE (Semi-Automatic Ground Environment), and the early stages of the F-111 fighter-bomber project. While industrial funding experienced some fluctuations in subsequent decades, it consistently remained a substantial component of federal R&D investment, typically exceeding funding to universities and FFRDCs (Federally Funded Research and Development Centers). Simply put, sustained federal investment provided crucial support for corporate research capabilities during the postwar period.

Defense procurement was especially important, creating both immediate demand and long-term incentives (Mowery, 2010, 2012). As documented by Alic (2007), the United States spent approximately $1.3 trillion on military R&D following World War II (equivalent to $2 trillion in year 2000 dollars), with an additional $2.3 trillion on procurement outlays for equipment and systems based on that R&D. The most substantial expansion of defense research funding occurred in the late 1950s, when defense R&D expenditures more than doubled from $13.21 billion in 1953 to $40.22 billion in 1959 (American Association for the Advancement of Science, 2022). The Korean War served as the primary impetus behind the U.S. military's embrace of high technology, as the services sought new weapons systems including jet-propelled fighters and bombers, nuclear-powered submarines, and experimental aircraft like the Dyna-Soar. These programs not only pushed R&D spending upward but also demonstrated the national commitment to establishing technological advantages over the Soviet Union.

Throughout the 1960s, defense R&D maintained high funding levels between $48-55 billion annually, substantially exceeding non-defense R&D allocations. The Department of Defense strategically directed funds to multiple competing technologies supporting similar missions, thereby fostering innovation through competition both within military services and between defense contractors. Although the Vietnam War shifted some focus toward immediate operational needs, the "DOD's commitment to high technology almost regardless of cost remained on display" (Alic, 2007). By 1980, defense R&D stood at $43.32 billion, compared to non-defense R&D at $46.14 billion.

This sustained commitment to defense procurement created a robust ecosystem for corporate innovation, providing both the resources and market stability necessary for long-term research investments. Major technological initiatives in stealth aircraft development, precision-guided weapons, and advanced computing systems demonstrated the Pentagon's willingness to fund multi-decade R&D programs. As Alic notes, weapons systems typically take "in excess of a dozen years and sometimes two decades or more" to develop, creating unique innovation cycles that differed profoundly from civilian technology development but ultimately contributed to it through "spin-off" and "spin-on" mechanisms (Alic, 2007).

The impact of defense procurement on corporate research was both widespread and transformative. Among the 877 firms in our sample operating an industrial laboratory in 1965, 356 (over 40 percent) conducted research under government contracts. A notable illustration is the American

Telephone and Telegraph Company (AT&T), home to one of the largest U.S. corporate labs.[11] Its Bellcomm division worked exclusively for the federal government, focusing on space exploration, missile systems, propulsion technologies, and communications for NASA missions. Even within Bell Labs itself, 44.7 percent of R&D activities were reported as government-funded in 1965 (IRL). IBM received approximately $500 million — their largest contract in the 1950s — to build 56 SAGE computers, advancing digital computing and networking technologies that would later influence modern computer architectures (Pugh, 2009).[12]

This "technological mobilization" extended beyond defense into civilian space exploration, as NASA's Apollo program directed $25.4 billion (approximately $257 billion in 2023 dollars) to corporate contractors between 1961 and 1972.(Dreier, 2022). North American Aviation, later part of Rockwell International, received a $36.9 billion (in 2016 dollars) contract to develop the Apollo Command and Service Module (CSM). This engineering challenge required breakthrough innovations in thermal protection systems capable of withstanding re-entry temperatures exceeding 5,000°F, development of regenerative life support systems for extended missions, and precision guidance computers with unprecedented reliability requirements (Brooks et al., 1980). The company also built the Saturn V's second stage, which demanded novel solutions in large-scale cryogenic propulsion technology and structural materials capable of containing super-cooled liquid hydrogen and oxygen (Bilstein, 1999).

The technological legacies of both ARPA and NASA highlight the critical relationship between federal funding and corporate innovation during this period. Through these programs, the federal government fulfilled a dual role — serving simultaneously as a research sponsor providing financial resources for ambitious R&D initiatives and as a lead customer creating guaranteed markets for resulting technologies. This dual function created a uniquely favorable environment that enabled Leader firms to sustain ambitious basic research programs throughout the Golden Age, assured that their technological developments would find both financial support and practical application.

In sum, we hypothesize the trichotomy of Leader, Absorber, and Follower firms that emerged in the 1950s was reinforced by a growing university sector and market for science-based innovations, both of which were fueled by the government's quest for scientific and technological dominance over its Cold War rival. To systematically test these hypotheses, we compile a comprehensive dataset capturing both the scientific inputs and outputs of firms. Rather than rely on a single indicator -such as laboratory size or publication counts, which can be misleading as shown by the cases of Koppers and Nabisco — we incorporate multiple dimensions of corporate research activities. We then apply a nonparametric clustering algorithm to classify firms along a latent "production vs. use of science" dimension. This novel approach enables, for what we believe to be the first time, a systematic analysis of industrial research strategies during this pivotal era in American

[11] According to the Industrial Research Labs of America Directory (IRL) of 1965, Bell Labs employed 720 physicists, 535 mathematicians, 296 chemists, 2,589 electrical engineers, 419 mechanical engineers, 282 other engineers, 29 metallurgists, 411 other professionals, 4,051 technicians, and 4,929 additional lab staff.

[12] The Semi-Automatic Ground Environment (SAGE) air defense system represented one of the era's most ambitious computing initiatives, with costs exceeding those of the Manhattan Project (Redmond & Smith, 2000).

innovation, highlighting the key determinants that shaped whether firms focused on creating or adopting scientific knowledge.

3 Data

Firms engage in science for two primary reasons: to generate breakthrough inventions by expanding the frontiers of scientific understanding, and to absorb external scientific discoveries (Cohen & Levinthal, 1990). But empirically distinguishing between the two is difficult. We consider multiple indicators of the inputs and outputs of firms' R&D efforts. On the input side, we use measures of human capital including the number of NRSTP and NER scientists employed, American Men and Women of Science listings, and industrial research laboratory size. For outputs, we analyze scientific publications authored by corporate researchers, publications citing Nobel Prize-winning research, patents, and corporate scientists' memberships in scientific societies, fellowships, and awards - indicators that reflect firms' tangible contributions to science and technology.

Our sample comprises 2,904 firms that appear in the Center for Research in Security Prices (CRSP) database and have registered at least one patent between 1945 and 1980 at the United States Patent and Trademark Office (USPTO). To construct this sample, we first identified all U.S.-headquartered firms listed in CRSP during this period and matched them with patent records from Google Patents. Our final dataset encompasses 644,877 patents granted between 1945 and 1980, matched to 2,868 unique publicly traded firms. For each firm in our sample, we collected relevant financial data from multiple sources. Market capitalization data was obtained from CRSP, while headquarters location, Standard Industrial Classification (SIC) codes, and additional financial metrics — including sales, market value, and total assets — were sourced from the CRSP-Compustat merged database.

3.1 Inputs of R&D Activities

3.1.1 Industrial Research Laboratories

The Industrial Research Laboratories (IRL) directory, published by the National Academy of Sciences/ National Research Council, provides information about corporate research laboratories in the United States, which are the physical locations and organizational units where firms conduct their R&D activities.[13] The data are collected through surveys sent to research directors, corporate communications officers, and additional administrative staff from the respective laboratories.

We expanded upon the work of Png (2019) by digitizing the IRL editions from 1920 to 1979, complementing their dataset covering 1975 to 1998. Our dataset includes 15 editions, with the number of laboratories ranging from 297 in 1920 to 6,293 in 1979, reflecting the substantial growth of industrial research facilities over six decades. For more details on the data collection and processing methodology, please refer to Appendix A.1.1.

[13]The publication was later taken over by R.R. Bowker in 1965, with volumes from 1970 to 1985 edited by Jaques Cattell Press.

3.1.2 American Men & Women of Science

The American Men and Women of Science (AMWS) is a comprehensive biographical directory profiling prominent living scientists in the United States. Established in 1906, AMWS has documented over 300,000 scientists across time. The directory has been used in prior research studying scientist labor market in the early and mid-twentieth century (Airoldi & Moser, 2024; Kim & Moser, 2021; Moser & San, 2020). We collect data on the 1949, 1955, 1961, and 1986 editions, which list 48,040, 47,875, 97,702, and 113,913 scientists, respectively. For each scientist, the directory provides detailed information including their educational background, professional affiliations, research interests, and notable achievements. We identified 46,527 unique scientists affiliated with 954 sample firms between 1945-1980. Of these, 49.8% (23,170 scientists) worked in chemical firms, 21.3% (9,910 scientists) in electronics firms, and 25.9% (12,050 scientists) in manufacturing firms. The remaining scientists were spread across other industries such as consumer goods, health, and miscellaneous sectors. Section A.1.2 provides details on how these scientists were matched to our sample firms.

3.1.3 National Register of Scientific and Technical Personnel

National Register of Scientific and Technical Personnel (NRSTP) is a comprehensive dataset compiled by the National Science Foundation (NSF) during the early Cold War period. The NRSTP data were collected in eight waves: 1954, 1958, 1960, 1962, 1964, 1966, 1968, and 1970. The register contains detailed information about scientists and engineers, including their skills and locations, and was designed to support government policy development and strategic planning in science and technology. To gather the information, the NSF surveyed a number of professional societies representing scientists and engineers. The surveys were then completed by the individual members of these professional organizations.

We matched employment data for each scientist from 1964 to 1970, a period for which NRSTP maintained detailed employment records. We focused on scientists who had earned Ph.D. degrees in natural sciences. We matched 26,245 PhD scientists to 855 firms, resulting in a total of 50,960 scientist-firm observations. In cases where a scientist worked for multiple firms, we created separate observations for each scientist-firm pair. The survey categorizes scientist employees into one of 14 job functions: (1) technical writing; (2) research; (3) management of R&D; (4) consulting; (5) management other than R&D; (6) sales, service, marketing; (7) development; (8) clinical, counseling, guidance; (9) operations; (10) test administration; (11) teaching; (12) weather; (13) urban/regional planning; (14) other. We analyze scientists engaged in research roles in corporations.

3.1.4 National Engineers Register

The National Engineers Register (NER) is a comprehensive database established by the Engineers Joint Council under contract with the National Science Foundation as part of the National Register of Scientific and Technical Personnel (NRSTP). The primary purpose of the NER was to ensure

that engineers could be quickly identified and mobilized during times of war or other national emergencies. The register contains detailed information about engineers, including their names, locations, birth dates and places, education (degrees earned, institutions attended, and years of graduation), work experience (current and previous employers, job titles, and years of employment), job functions (research, development, management, marketing, etc.), and areas of expertise (engineering disciplines, skills, and knowledge areas).

The NER data was collected in three waves: 1964, 1967, and 1969. The survey consisted of a random sample of names drawn from the membership rolls of various engineering societies. To be included in the NER survey, respondents had to meet specific criteria. They were required to hold a bachelor's degree or higher in engineering, be registered as a professional engineer or as an engineer-in-training, and maintain full professional membership in a recognized engineering society.

The NER categorizes engineers into one of 14 job functions: (1) development; (2) research; (3) management; (4) design; (5) consulting; (6) exploration; (7) quality control; (8) teaching, training; (9) data processing, technical writing; (10) product testing; (11) sales, marketing, service; (12) production and operations; (13) facilities; (14) other. We matched 3,603 engineers with PhD degrees with 475 firms, resulting in 5,996 person-firm observations. Consistent with our focus on research activities in the NRSTP data, we concentrate specifically on PhD engineers who occupied research positions within these corporations.

3.2 Outputs of R&D Activities

3.2.1 Corporate Scientific Publications

To identify scientific publications by publicly traded firms, we employ a novel matching approach using the Dimensions database, a comprehensive source of scholarly publications. Our matching procedure involves several steps. First, we standardize author affiliations using a Large Language Model (OpenAI's GPT-4o-mini) to address the heterogeneity in how organizational names are recorded. The model also classifies each affiliation into distinct organizational categories (e.g., universities, corporations, government agencies, hospitals, etc). This classification enables us to isolate publications with corporate authorship.

We then match these corporate publications to firms in the CRSP database. Our matching procedure identifies 1,434 publicly traded firms that published scientific research during the 1945-1980 period, producing 189,707 unique publications.

3.2.2 Nobel Prize Winning Publications and Corporate Publications Citing Them

First, we examine the affiliations of Nobel laureates at the time they published their prize-winning research to identify instances where the award-winning work was conducted within corporate laboratories. Our analysis covers 939 award-winning publications authored by 647 Nobel laureates from 1901 to 2023, spanning multiple scientific disciplines. Of these, 48 publications (5.1%) were affiliated with corporations at the time of publication, involving 38 unique Nobel laureates (5.9% of

all laureates in our dataset). These corporate-affiliated prize-winning publications are distributed across several prominent firms, with AT&T's Bell Laboratories leading with 11 Nobel laureates, followed by contributions from General Electric, Texas Instruments, DuPont, Dow Chemical, Monsanto, Varian Associates, Hercules, Exxon, Smith, Kline & French, and American Science and Engineering.[14]

Second, we trace how firms build upon fundamental scientific discoveries by identifying corporate publications that cite Nobel Prize-winning work. We construct citation links between corporate publications in our dataset and the 939 Nobel Prize-winning publications identified above. By tracking these citations, we can observe which firms are actively contributing to scientific dialogues surrounding these fundamental advances.

Patents Citing Firm's Internal Publications — We examine patents' connections to scientific knowledge through their citations to scientific publications following Marx and Fuegi (2020)'s approach. To measure how firms build upon their own scientific research in their technological development, we focus specifically on instances where a firm's patents cite scientific publications authored by the same firm's scientists, which we term "self-citations". Our matching identifies 2,943 patent-paper self-citations across 131 firms during 1945-1980. The distribution of these self-citations reveals significant variation across industries in how firms leverage their scientific research for technological development. The electronics sector shows the highest rate of self-citation with 2,102 instances, followed by the chemical industry with 392 citations and healthcare with 229 citations. Manufacturing firms show more moderate levels of self-citation (162 instances), while consumer goods firms have relatively few (53 citations). This industrial variation suggests different patterns in how sectors integrate scientific research into their technological development processes, with technology-intensive industries showing tighter coupling between their scientific research and patenting activities.

3.2.3 Patents

We construct our corporate patent dataset by matching patent assignee names from Google Patents to firm names in the Center for Research in Security Prices (CRSP) database. Our matching procedure extends the widely-used dataset developed by Kogan et al. (2017), identifying additional patents assigned to publicly traded firms during this period.

For each patent, we incorporate several complementary measures to capture different dimensions of technological innovation. The economic significance of patents is measured using Kogan et al. (2017)'s stock market-based value measure, which captures market reactions to patent grants. We assess technological impact through forward citation counts, measuring how frequently each patent is cited by subsequent innovations. To evaluate technological novelty, we employ Kelly et al. (2021)'s text-based innovation indicators, which assess each patent's similarity to prior and subsequent patents to identify truly novel and influential innovations.

Finally, we incorporate information about government-funded research from the Government

[14]See A.2.1 for details on our data collection process.

Patent Register developed by Gross and Sampat (2024). This database identifies patents arising from U.S. government-funded research from the early 20th century through 2020, which we match to our sample of CRSP firms. The register distinguishes between two crucial categories of government-involved patents: those where the U.S. government retained the patent title (government title patents), directly reflecting federal research initiatives, and those assigned to private entities but developed with government funding, where the government maintained license rights (government license patents).

This data is particularly valuable for our study period (1945-1980), as it encompasses the significant surge in government-funded corporate R&D during and after World War II. The inclusion of both title and license patents provides a more complete picture of federal involvement in corporate innovation than previously possible, as many license patents cannot be identified through traditional patent text analysis or other conventional sources. This granular classification of government involvement allows us to examine how federal funding shaped firms' technological trajectories and their engagement with scientific research during this transformative period in U.S. innovation history, when government funding played an increasingly crucial role in corporate R&D.

3.2.4 Department of Defense (DoD) Contracts

The Department of Defense (DoD) Defense Contract dataset contains comprehensive information on all contracts awarded by the DoD between 1966 and 1980. This rich dataset was initially collected by the DoD on microfilm and later converted into machine-readable CSV format by Daniel Gross, greatly enhancing its accessibility for research purposes.[15] The dataset offers granular details on each DoD contract during the specified period, including the contracting office identifier and name, contractor identifier, name, and location, contract number, date, and dollar amount, as well as product or service codes describing the nature of the purchase.

We matched the names of contractors in the defense data to the corresponding firm names in our sample, using fuzzy string matching techniques. Once the contractor names are successfully matched to the firm names, the defense contract variables can be aggregated for each firm-year observation in the panel. We then constructed the total dollar value of DoD contracts awarded to the firm annually, the proportion of the firm's total sales attributable to DoD contracts. To ensure the comparability of contract values across time, we deflated all amounts to 2017 dollars using the GDP Price Deflator from the Bureau of Economic Analysis.

3.2.5 Award, Fellowship, and Membership

We collected data on corporate recipients of prestigious science & engineering awards, and elected memberships. This includes the National Medal of Science, American Association for the Advancement of Science (AAAS) Fellowship, Membership of the National Academy of Sciences (NAS), the

[15]The DoD defense contract data used in this analysis was generously provided by Daniel Gross, who has undertaken the substantial task of converting the original DoD defense contract records, sourced from the National Archives and Records Administration and stored on microfilm, into a machine-readable CSV format.

National Academy of Engineering (NAE), the National Academy of Medicine (NAM), the American Academy of Arts and Sciences (AAA&S), and the American Philosophical Society (APS).

These awards and memberships honor individuals who have made outstanding contributions to science, engineering, and medicine. We collected data on National Medal of Science (NMS) recipients from the National Science Foundation website and gathered information on AAAS Fellows from the AAAS website. For NAS, NAE, and NAM members, we scraped their respective websites, and for AAA&S and APS members, we used each organization's online directory. This process enables us to identify the corporations and other institutions with which these accomplished scientists and engineers were affiliated at the time of their recognition. Further details on the data collection procedures, as well as the specific variables extracted for each award and membership, can be found in Appendix A.2.

3.3 Coveage and Summary Statistics of Variables

Table 1 displays the data-availability for key variables used in the classification between 1945 and 1980. It reveals complete coverage for patent counts, publication counts, AMWS, and awards across all years, while NRSTP & NER scientist counts are only available for specific periods (1964, 1966-1970), and IRL scientist data is intermittently available (1946, 1950, 1956, 1960, 1965, 1970, 1975, 1979).

[Tables 1 and 2 Here]

Table 2 presents summary statistics for our main variables based on an unbalanced panel of 2,899 U.S.-headquartered publicly traded firms from 1945 to 1980. The sample reveals substantial heterogeneity in firms' innovative and scientific activities.

On average, firms in our sample generate 12 patents per year (standard deviation = 51), though the distribution is highly skewed, with the median (50th percentile) firm producing no patents. Scientific publication activity shows similar concentration, with a mean of 4 publications per year (standard deviation = 31) and the 90th percentile at just 3 publications, indicating that scientific publishing was limited to a select group of firms.

Scientific human capital is captured through several measures. From our survey data, firms maintain an average of 20 scientists, with 11 specifically designated as research scientists and 2 as manager scientists. The American Men and Women of Science (AMWS) directory identifies an average of 49 scientists per firm (standard deviation = 283), with substantial variation across firms (10th-90th percentile range of 0-85 scientists).

Regarding research infrastructure, all firms that operate industrial laboratories in our sample maintain them throughout the sample period. A subset of 157 firms operate laboratories that conduct research for government contracts.

4 Classifying Corporate Leaders and Absorbers of Science

As discussed in Section 2 industrial labs serve one of three roles: Followers (focusing on product development and testing), Science Absorbers (primarily incorporating external scientific knowledge), or Science Leaders (generating new scientific insights). However, disentangling these three categories at scale is difficult using any single metric.

4.1 Limitations of Uni-dimensional Classification

Classifying firms with large laboratories as genuine research establishments can be misleading, as many such labs may devote most of their efforts to product development rather than fundamental research. For instance, Scott Paper Company housed extensive facilities, including pilot plants, and employed a sizable corps of technical specialists. Yet according to the 1960 Industrial Research Laboratories (IRL) directory, 80 percent of Scott's lab resources targeted production improvements. Nabisco (formerly National Biscuit Company) illustrates a similar pattern. In 1960, the firm operated a substantial research center in Fair Lawn, New Jersey, employing 205 technical staff — 52 chemists, 30 mechanical engineers, and 94 technicians — supported by pilot plants and machine shop equipment. Despite substantial laboratory facilities and a sizable technical staff, Nabisco's R&D efforts were largely confined to routine product development and quality control, showing minimal engagement with academic science or novel research findings.

An examination of Nabisco's technical output from the 1940s through the 1980s bears this out. Their publications in the Journal of the Association of Official Agricultural Chemists International focused on routine quality-control procedures — such as measuring sugar content and moisture levels in baked goods — that required little scientific knowledge beyond basic chemistry. Typical work, including "Correlation of Keeping Properties of Shortenings with Keeping Quality of Biscuits" (Maveety, 1945) and "The Change in Thiamine Content of Breakfast Cereals During Cooking" (White et al., 1946), involved little more than basic testing of manufacturing processes. Similarly, their contributions to the Journal of Food Protection addressed standard industrial sanitation practices and regulatory compliance (Barron, 1959), with scant attention to advanced microbiological or chemical principles. This pattern aligns with the company's 1960 annual report, which emphasized "better products, better ways of making them, and better methods of getting them to consumers."

In a similar vein, scale R&D operations alone cannot distinguish science-absorbing firms from science-advancing ones. The R&D scale at the Koppers Company was large yet its internal efforts were focused on absorbing scientific advances from elsewhere. It employed 413 laboratory scientists, of whom 25 were featured in the American Men & Women of Science, and produced 2.64 scientific publications and 46 patents per year. Koppers' patents frequently built upon external research, like their "Production of benzophenones" patent that directly leveraged academic work from Pennsylvania State College on hydrogen fluoride chemistry. Further indicators of its limited scientific leadership include minimal representation in elite scientific organizations, absence of ma-

jor scientific awards, and patents that rarely referenced in-house research — suggesting that despite impressive labs and technical expertise, Koppers functioned primarily as an Absorber rather than a Leader of scientific knowledge.

Beyond laboratory size, the sheer volume of scientific publications might seem a more direct measure of scientific production, yet it, too, presents challenges. Many corporate publications document routine testing ("American Cyanamid Company Technical Bulletins", 1988), quality control (Willis A. Jensen & Vining, 2018), or incremental technical improvements (Rinner, 2003) rather than advancing fundamental knowledge.[16] In this sense, a high volume of publications does not necessarily equate to scientific production in the same way as the research output of firms historically recognized as Science Leaders, such as AT&T's Bell Labs, IBM, or DuPont.

Such patterns appear in numerous industries. Pharmaceutical firms often publish studies on bioequivalence or manufacturing validation — intended mainly for regulatory compliance rather than scientific discovery. Similarly, engineering and manufacturing firms release case studies that highlight the application of existing methods to technical challenges rather than developing new theoretical insights.

A prominent example is Procter & Gamble (P&G). Most of P&G's work appears in applied research outlets, such as the Journal of the American Oil Chemists' Society (104 publications), Toxicology and Applied Pharmacology (44), and Journal of Dental Research (26). These publications largely focus on applying well-established scientific principles rather than pioneering new discoveries.

P&G's top-cited paper, "Percutaneous Absorption: On the Relevance of in Vitro Data" (Journal of Investigative Dermatology, 1975, 1,605 citations), refined experimental methods for skin absorption testing, leveraging established principles of membrane transport and diffusion rather than introducing new theoretical concepts. Similarly, "The Microscopy of the Liquid Crystalline Neat and Middle Phases of Soaps and Synthetic Detergents" (Journal of the American Oil Chemists' Society, 1954, 686 citations) employed standard optical microscopy techniques to characterize soap structures — sophisticated work, but grounded in existing liquid crystal theory. Likewise, "Quantitative Analysis of Fatty Acids by Gas-Liquid Chromatography" (Journal of Lipid Research, 1964, 221 citations) focused on refining chromatographic methods, illustrating P&G's strength in adapting analytical techniques rather than developing novel scientific frameworks. P&G's publications demonstrate its advanced capabilities in translating scientific knowledge, largely developed outside P&G, into commercial applications. This differs from the research trajectory of firms such as AT&T Bell Labs or IBM, where investigations frequently yielded foundational discoveries in physics, computer science, and information theory.

Firms with moderate publication records demonstrate even less engagement with scientific principles. Chicago Bridge & Iron Co. exemplifies this pattern. Despite having 59 publications, the company's output shows little interaction with frontier scientific knowledge, focusing primarily on

[16] The Journal of Quality Technology (JQT) illustrates this distinction. Its inaugural editorial in 1969 explicitly stated a non-scientific mission, explaining that the journal would focus on techniques "already workable" rather than pioneering new statistical methods (Nelson, 1969).

practical engineering calculations and standard industry procedures. Its most frequent publication venues were the Journal of the Waterway Port Coastal and Ocean Division (12 publications) and the Journal of the Waterways Harbors and Coastal Engineering Division (8). These outlets generally document standard engineering calculations and practical design considerations.

The firm's most-cited paper, "Versatile Elastic-Plastic Stress-Strain Formula" (1975, 806 citations), published in the Journal of Engineering for Industry, provided a mathematical convenience for established engineering calculations. Likewise, "Wave Forces on Multiple Vertical Cylinders" (Journal of the Waterway, Port, Coastal and Ocean Division, 1978, 77 citations) and "Interaction of Waves with Large Vertical Cylinder" (Journal of Ship Research, 1975, 83 citations) validated existing theoretical predictions for practical design applications, rather than advancing fluid dynamics. Whereas P&G actively sought and applied scientific insights to create novel solutions, Chicago Bridge & Iron aimed primarily to streamline engineering calculations for standard design processes.

In sum, neither lab size nor publication volume — even when weighted by forward citations — provides a conclusive indicator of a firm's position on the R&D spectrum (Follower, Science Absorber, or Science Leader). To address these limitations, we propose a multi-dimensional classification strategy in the next section, mitigating the biases that arise from relying on single-indicator methods.

4.2 Multidimensional Classification: Ensemble Learning

We develop an ensemble learning approach that combines multiple deep learning algorithms. Ensemble learning leverages the combined strengths of multiple machine learning algorithms (Goodfellow et al., 2016) while mitigating their individual weaknesses (Dietterich, 2000; Dong et al., 2020), resulting in improved classification outcomes(Mienye & Sun, 2022).

The variables used in the ensemble model include publication metrics (count of all firm publications and those citing Nobel Prize-winning publications), patenting activity (count of all firm patents and those citing the same firm's publications), number of industrial laboratory scientists and NRSTP research scientists, inclusion of employees in the American Men & Women Science directory, awards received, membership in scientific and engineering societies.[17] The model analyzes these variables across three time periods: the full sample (1945-1980), early period (1945-1965), and later period (1965-1980).[18]

[17] Please see Appendix Table B.1 for further details on variable definitions

[18] Our implementation incorporates 16 distinct unsupervised deep learning methods, each potentially capturing distinct patterns in the relationships among our scientific metrics: Agglomerative Clustering; Attention Mechanism (Vaswani et al., 2017); Autoencoder (Pinaya et al., 2020); Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH) (Zhang et al., 1996); Convolutional Neural Network (CNN) (LeCun et al., 1998); Deep Autoencoder (Hinton & Salakhutdinov, 2006); Deep Belief Network (DBN) (Hinton et al., 2006); Deep Boltzmann Machine (DBM) (Salakhutdinov & Hinton, 2009); Gaussian Mixture Model (GMM) (Reynolds et al., 2009); Hierarchical Clustering (Murtagh, 1983); K-Means; Residual Network (ResNet) (He et al., 2016); Restricted Boltzmann Machine (RBM) (Smolensky et al., 1986); Self-Organizing Maps (SOMs) (Kohonen, 2013); Spectral Clustering (Ng et al., 2001); and Variational Autoencoder (VAE) (Kingma & Welling, 2013). While these methods operate in an unsupervised manner, we subsequently validate their performance against a "ground truth" set to select the most

By integrating three families of algorithms — classical clustering, representation learning, and specialized neural architectures — our ensemble approach offers a robust and flexible framework for analyzing firms' scientific capabilities.

Classical Clustering Algorithms. The first set of methods — encompassing Agglomerative Clustering, K-Means, Hierarchical Clustering, and BIRCH — is particularly adept at handling asymmetric capability patterns. Traditional classification approaches (e.g., Logistic Regression) rely on linear decision boundaries, making them prone to misclassifying firms with unusual profiles (such as very high publication counts but relatively few patents). While Gaussian Mixture Models (GMMs) relax some of these constraints, they still assume that multiple indicators follow particular distributions, limiting their usefulness for firms that deviate significantly from such assumptions. In contrast, clustering algorithms group observations based on overall similarity, free from strict distributional or structural constraints. This flexibility allows them to identify unconventional but important patterns of scientific engagement that might elude simpler models.

Representation Learning Models. A second set of techniques — Deep Belief Networks (DBN), Deep Boltzmann Machines (DBM), Restricted Boltzmann Machines (RBM), and Variational Autoencoders (VAE) (Bengio et al., 2013; Hinton et al., 2006; Salakhutdinov & Hinton, 2009) — are well suited for detecting hierarchical dependencies. Traditional dimensionality reduction methods, such as Principal Component Analysis (PCA) or Linear Discriminant Analysis (LDA), compress data into linear combinations of features and can therefore miss the subtle ways in which advanced scientific achievements build on more basic capabilities. Deep learning tools, by contrast, excel at discovering non-linear and layered relationships in the data. For example, they can detect that firms garnering elite awards or having Nobel laureates typically also maintain large research staffs, while noting that not all firms with substantial staffing reach those elite ranks. By preserving these multi-layered dependencies, representation learning methods reveal more nuanced patterns than linear techniques can provide.

Specialized Neural Architectures. Finally, models such as Convolutional Neural Networks (CNNs), Residual Networks (ResNet), Attention Mechanisms (Vaswani et al., 2017), and Deep Autoencoders excel at recognizing multiple pathways through which firms can achieve genuine science production. Unlike Decision Trees and Naive Bayes, which impose rigid splits or strong independence assumptions, these architectures capture overlapping, non-linear relationships among indicators such as publication counts, staff size, and elite scientific affiliations. For instance, some firms achieve high-level science production through exceptional publication activity and a large research workforce, while others may rely on a handful of Nobel Prize laureates or other advanced credentials. By learning these different routes to the same classification outcome, specialized neural models can accommodate diverse firm strategies and capabilities, going beyond what simpler methods can discern.

The methodological plurality of ensemble methods provides a safeguard against measurement error, a common issue in economic and business history research. Also, they help make the best

reliable classifiers for our specific context.

use of uneven data coverage. Patents, publications, and American Men of Science listings are available from 1945 to 1980, but other essential measures of industrial research appear only intermittently. For example, the Industrial Research Laboratories directory was published in 1946, 1950, 1956, 1960, 1965, 1970, 1975, and 1979, and the NRSTP survey's scientist counts exist solely for 1964–1970. Our ensemble approach helps mitigate these data constraints by merging models that rely on consistent annual indicators with those leveraging sporadic sources.

4.2.1 Estimation

We employ a two-stage selection procedure to identify which of the sixteen candidate models to include in our ensemble. In the first stage, we rank each model's classification outcomes against a ground truth set of 20 "Science Leaders" (e.g., AT&T, IBM, GE, RCA, DuPont) and 871 purely development-oriented firms. We define the Leader group by non-zero scores across all of our scientific indicators (such as notable award recipients and R&D staff with advanced qualifications), while the Follower group has zeroes on these metrics. Because the ground truth contains only these two categories — Leaders and Followers — we focus on how many Leaders each model correctly identifies (the "Leaders Correctly Identified" row in Table B4) and how consistently it keeps true Followers out of the Leader group (the "Followers Correctly Identified" row). Models that mislabel many Followers as Leaders or overlook too large a share of "ground-truth" Leaders are discarded.

After removing these under-performing approaches, we retain eleven methods: Agglomerative Clustering, Autoencoder, CNN, Deep Autoencoder, Deep Belief Network, Deep Boltzmann Machine, Gaussian Mixture Model, Hierarchical Clustering, K-Means, ResNet, and Variational Autoencoder. Table B4 and Appendix B.1 provide complete details on these methods' performance, including how many Leaders or Followers each one misclassifies. The retained algorithms collectively achieve strong "Leaders Correctly Identified" and "Followers Correctly Identified" rates and show an acceptable "Weighted Accuracy," indicating that they neither overlook the small set of science-oriented firms nor inflate the Leader category with development-only enterprises.

In the second stage, we combine these eleven models' three-class predictions — Science Leader, Science Absorber, and Follower — using a majority-vote scheme. Each firm is assigned to the category receiving the most votes. In cases of tie votes, we adopt a hierarchical rule favoring Science Leader status over Absorber, and Absorber status over Follower, so that contested firms are placed in the higher-intensity class. This voting procedure yields 38 Science Leaders (1.3%), 176 Science Absorbers (6.1%), and 2,690 Followers (92.6%). We validate these final labels by checking how the 20 "ground-truth" Leaders and 871 verified Followers fare under the ensemble. All of the latter remain correctly identified as non-Leaders, and 14 of the 20 "ground-truth" Leaders receive a Leader label. Although the "Absorber" category lacks an external ground truth, supplementary checks of patent-to-publication citations, PhD-staffed laboratories, and scientific affiliations confirm that Absorber firms occupy an intermediate zone of research engagement between Leaders and Followers. Full details on these additional checks and the final voting process appear in the Appendix B.1.

[Table 3 Here]

Table 3 reveals a shift in the composition of firm types over time. The share of Science Leaders increased from 0.69 percent in 1945–1965 to 1.87 percent in 1965–1980, suggesting a modest but meaningful rise in the number of firms significantly invested in generating new scientific knowledge. The proportion of Science Absorbers also grew, from 8.41 percent to 14.20 percent, indicating a broader uptake of research engagement across industries. Nevertheless, Follower — firms focused principally on technology development — remained the dominant group, comprising 83.93 percent to 92.63 percent of all firms during these intervals.[19]

[Figure 6 Here]

Figure 6 examines how firms cluster across multiple dimensions of scientific activity during 1945-1980. The scatter plots reveal distinct patterns among Science Leaders (red dots), Science Absorbers (green dots), and Followers (gray dots). Science Leaders consistently occupy the upper ranges across metrics, particularly evident in the publications versus industrial lab scientists plot, where they cluster in the upper-right quadrant. Science Absorbers form an intermediate group with moderate levels of scientific activity, while Followers predominantly cluster in the lower ranges.

The plots show substantial overlap between groups, especially in the lower ranges (0-2) of metrics like NRSTP research scientists and awards & membership. This overlap demonstrates why single metrics or simple thresholds are insufficient for classification. The multidimensional nature of scientific activity requires advanced classification methods like ensemble learning to effectively distinguish between these three groups.

4.2.2 Validation

We validate our classification results in three ways. First, we expect Absorber and Leaders firms to use external scientific knowledge in their inventions more than Followers. The difference in external science use between Absorbers and Leaders, on the other hand, should not differ significantly. Second, we expect Leaders to use internal science more than Absorbers. Third, we expect Leaders to use more recent science and science that is closer to academic science.

[Table 4 Here]

Table 4 presents OLS regression estimates examining the relationship between patent citations to scientific literature and firm classification types derived from the ensemble model. Columns 1 and 2 analyze firms' propensity to incorporate scientific research into their patenting activities. Specifically, Column 1 reports the probability of a firm's patents citing external scientific publications (those produced by other organizations), while Column 2 presents the probability of patents citing internal scientific publications (the firm's own scientific output).

[19] A feature-importance analysis using SHAP values identifies industrial laboratory scientists (0.047–0.067), patents (0.047–0.054), and listings in *American Men & Women of Science* (0.037–0.058) as the strongest predictors of a firm's classification, with publications (0.032–0.040) following closely and NRSTP research scientist counts (0.017–0.019) having moderate influence. Citations of Nobel Prize–winning research (0.003–0.004), patent self-citations (0.001–0.002), and scientist society memberships (0.002–0.004) are less impactful overall but still relevant to the ensemble classification.

The empirical results from Column 1 indicate that Science Leaders and Science Absorbers demonstrate significantly higher probabilities of citing external scientific publications (3.5% and 3.1% respectively) compared to Follower firms, which serve as the baseline category. These statistically significant positive coefficients suggest that both Science Leaders and Absorbers possess enhanced capabilities in identifying and integrating relevant external scientific advances into their innovation processes relative to Followers firms.

However, a notable difference appears in Column 2, where Science Leaders show a substantially higher probability of citing their own scientific work (0.6%) compared to Science Absorbers (0.2%), with both groups citing internal science more frequently than Followers firms (the omitted reference category). This threefold difference between Leaders and Absorbers suggests that Leader firms utilize science not only as a means to absorb external knowledge but also as direct inputs into their innovation processes.

Columns 3 and 4 examine the intellectual proximity between patents and scientific research using a more nuanced citation distance measure developed by Ahmadpoor and Jones (Ahmadpoor & Jones, 2017). This distance metric represents the shortest path between a patent and a scientific publication within the citation network. For instance, a distance of 0 indicates direct citation of a scientific paper by a patent, while a distance of 1 reflects a citation path requiring one intermediate step (e.g., patent A cites patent B, which cites scientific publication C).

Column 3 presents the average citation distance between a firm's patents and external scientific publications. The results show that both Science Leaders (-0.647) and Science Absorbers (-0.455) maintain significantly closer connections to external scientific literature than Follower firms. Similarly, Column 4 reports the average citation distance between a firm's patents and its own scientific publications. Here, the difference between firm types becomes more pronounced, with Science Leaders demonstrating substantially shorter citation distances (-1.072) compared to Science Absorbers (-0.460), and both groups showing closer intellectual proximity than Follower firms. This pattern provides additional evidence that Leader firms integrate their own scientific discoveries more directly into their technological development processes compared not only to Follower firms but also to Science Absorber firms.

Columns 5 and 6 analyze similarity between corporate and academic research using Field of Research (FOR) classifications (Australian Bureau of Statistics, 2020) at both main category and subcategory levels. To measure the similarity between corporate and university publications, we first compute pairwise similarity scores with all university publications in the same field published within the previous five years. Rather than using simple averages that might dilute strong connections or maximum values that could be skewed by extreme cases, we calculate the average similarity score of the top 10% most similar university publications.

The normalized similarity scores at both the main category (Leaders: 0.178, Absorbers: 0.095) and subcategory levels (Leaders: 0.255, Absorbers: 0.144) reveal that while both groups engage with scientific research, Leaders maintain closer alignment with academic scientific content. The higher similarity scores for Science Leaders, particularly at the more granular subcategory level,

suggest that these firms not only utilize scientific knowledge but also contribute to fundamental scientific discourse in ways that more closely resemble academic research.

Column 7 further reveals that Science Leaders engage more intensively with recent scientific knowledge. The significantly more negative coefficient for Leaders (-2.856 versus -0.751 for Absorbers) indicates they cite much more recent scientific research in their patents compared to Absorbers. This suggests that Science Leaders not only align more closely with academic research content, but also actively engage with and build upon the scientific frontier.

In summary, both Science Leaders and Science Absorbers incorporate scientific findings more frequently than Followers do. Although Leaders and Absorbers exhibit comparable tendencies to leverage external science, Leaders are notably more likely to draw on their own scientific output. This pattern suggests that Absorbers publish in peer-reviewed journals primarily to bolster their capacity to absorb outside knowledge, while Leaders publish with an additional goal in mind — namely, creating new knowledge that can be directly integrated into the firm's inventive activities.

5 What Accounts for the Three Strategies for Scientific Research?

We have so far shown the Golden Age of corporate science consisted of a trichotomy of Leader, Absorber and Follower firms. In this section, we empirically explore the factors that led to the emergence of these distinct research strategies: i) firms closer to the technological frontier (Leader & Absorbers) heavily relied on scientific inputs, ii) but Leaders were farther ahead of Absorbers such that they required scientific knowledge that universities may not provide; iii) firms with larger market power were better able to appropriate returns from such scientific inputs; iii) government demand for technologically sophisticated products may have subsidized the production of knowledge.

5.1 Technological Frontier and Need for Science

Table 5 examines the relationship between firms' scientific capabilities and their technological leadership, measured through two dimensions: the entry into new technological domains (first-in-CPC class patents) and the creation of "breakthrough" patents (patents in the top 20% of forward-to-backward similarity ratios (Kelly et al., 2021)).[20] We estimate OLS regressions comparing Science Leaders and Absorbers against Follower (baseline), controlling for firm size and patenting scale.

[Table 5 Here]

Science Leaders generate 289.75 more inaugural patents than Followers, while Science Absorbers generate only 3.54 more (Column 1). The disparity is even more pronounced for breakthrough inventions (Column 2). Science Leaders generate 2,200.88 more breakthrough patents than

[20] The forward-to-backward similarity ratio captures innovations that break from existing knowledge paradigms while establishing new directions that subsequent inventors follow.

Followers, while Science Absorbers generate only 22.00 more. These results suggest two complementary explanations for corporate scientific investments. First, proprietary scientific research may have allowed firms to gain competitive advantages even in areas where university research existed, through exclusivity and earlier access to scientific insights, especially in key areas such as computer science and semiconductors where the university ramp-up was slower, as also documented by (Arora, Belenzon, & Dionisi, 2023). Second, firms may have been conducting research in domains where university efforts were insufficient or absent altogether, in essence filling gaps in the public knowledge base. These findings highlight the strategic importance of in-house scientific capabilities for firms seeking to extend their technological leadership.

5.2 Appropriability of Science-based Products

Market Power — Columns 3 and 4 of Table 5 examines the relationship between firms' scientific capabilities and measures of market power. We find that both Leaders and Absorbers are significantly larger than Followers, with Leader assets approximately 27.2 times greater, while Absorbers are about 9.3 times larger (Column 3). Market shares reflect similar patterns: Leaders hold market shares 3.6 percentage points higher than Followers, while Science Absorbers show a 0.6 percentage point gain (Column 4). Furthermore, the gap between Leaders and Absorbers is statistically significant.

[Figure 7 Here]

Figure 7 tests whether Leader firms are more likely to operate in concentrated industries. We measure industry concentration using the share of industry assets held by the top 1% of firms (Panel (a)) and the share held by the top 10% of firms (Panel (b)).[21] Industries are categorized into low concentration (0-25%), medium concentration (25-75%), and high concentration (75-100%) based on 1% and 10% firm asset shares each year. The number Leader firms is then divided by total firms in a focal industry-year. The dots connected by dashed line in Figure 7 show the average Leader share in each concentration category (Absorber firm shares are connected by solid lines).

We find both Leaders and Absorbers are to be more prevalent in industries with higher asset concentration. Comparing the high vs low concentration industries in Panel (a), Leader firm shares increase by around 2 percentage points, while Absorber firm shares increase by around 6 percentage points. Notably, leaders firms are rarely found in industries with medium or low concentration. Using top 10% asset shares shows largely similar results in Panel (b).

With incomplete markets for technology, the ability to capture the value generated from scientific investments depends on control of markets for products. Our findings show that Leader and Absorber firms consistently exhibit greater size, market share, and industry concentration compared to Follower firms, and the clear gap between Leaders and Absorbers both confirm that a

[21] The variables are derived from IRS Statistics of Income data and obtained through Kwon et al. (2024). The asset shares are calculated for BEA subsectors, which are similar to 2-digit SICs.

robust control of product market could mitigate appropriability concerns for investing in scientific inputs.

Government Procurement — As recounted in Section 2, federal procurement policy also helped, especially when the downstream market was uncertain and amorphous. For products such as the laser, which was dubbed a "solution without a problem", firms could find a willing and able customer in the Department of Defense which first adapted lasers for artillery ranging and then for guided munitions. Table 6 analyzes how research strategies relate to government engagement across three dimensions, using Followers as the baseline category. In all three measures, Leaders demonstrate stronger government connections compared to both Absorbers and Followers.

[Table 6 Here]

The first measure examines the relationship between firms' scientific capabilities and their government patent counts, where government patents are defined as those licensed to or directly held by government entities (Gross & Sampat, 2024). Column 1 in Table 6 shows that relative to Followers, Absorbers have approximately 26.5% more government patents, while Leaders have roughly 175.7% more government patents.

Column 2 examines whether the firm operates a research lab wholly or substantially for government needs.[22] Column 2 shows that Science Absorbers are 7.6 percentage points more likely than Followers to operate government-sponsored laboratories, while Science Leaders have a 29.6 percentage point higher probability. Both effects are significant (standard errors of 0.066 and 0.117, respectively). These differences persist even after controlling for a firm's broader research capacity (measured by industrial-lab scientist counts), suggesting that Science Leaders hold a distinct edge in securing government research partnerships.

The third measure gauges defense contracting intensity, defined as the ratio of Department of Defense (DoD) contract values to firm sales.[23] Even after controlling for firm size and industry, Leaders demonstrate higher defense contracting intensity (11.1 percentage points above Followers) than Absorbers (8.0 percentage points above Followers), underscoring their stronger engagement with government-supported research.

In sum, dominant market positions for known downstream products and government sponsorship of the latest science-based innovations provided seems to have mitigated appropriability concerns for Leaders (and to a lesser extent, Absorbers) (Belenzon & Cioaca, 2021).

[22] In the 1965 edition of Industrial Research Laboratories of the United States, American Telephone and Telegraph Company (AT&T) appears as a prominent example of a "government-sponsored" research laboratory. Although AT&T is widely recognized as a private telecommunications giant, the IRL listing shows that nearly half (44.7%) of the company's R&D in 1965 was performed for the U.S. government. In addition, Bellcomm Inc. — noted in the same entry as a joint venture with Western Electric and AT&T — devoted its entire research effort to U.S. government contracts (in this case, largely for NASA).

[23] Data on DoD defense contracts come from Daniel Gross, who converted the original contract records — sourced from the National Archives and Records Administration and stored on microfilm — into a machine-readable CSV format.

5.3 University Entry into Invention

Around the 1980s, science programs at corporate laboratories began to unravel, culminating in the closure and spinoff of basic research programs even at leading firms such as Xerox PARC and DuPont's Purity Hall. However, what is less known is that this withdrawal from science was even more pronounced for Absorber firms that trailed the celebrated Leaders. As Figure 8 indicates, while Leader and Follower firms generally maintain steady (and in some instances slightly increasing) publication-to-sales ratios over the observed period, Absorbers exhibit a clear rise in their research output during the late 1970s or early 1980s, followed by a marked downturn thereafter.[24]

[Figure 8 Here]

This is consistent with our characterization of this research strategy as primarily concerned with acquiring outside knowledge: as reforms such as the Bayh-Dole Act and the steady entry of universities into inventions, absorbing academic science became easier, reducing the need for research at Absorber firms. Followers, who did not rely on the latest scientific developments, or Leaders, who relied more heavily on internal science, were less likely to have responded to the "forward integration" of universities into the business of science-based invention.

[Table 7 Here]

Table **??** tests the hypothesis that the greater ease of acquiring inputs to innovation diminished Absorber firms' incentives to invest in science. Column 1 regresses corporate publications against ensemble category dummies that are interacted with exposure to markets for technology (MFT). MFT exposure is defined at the firm-year level by weighting the share of traded patents in CPC patent class-years by how heavily the firm patented in the focal CPC.[25] Firms operating in technology classes with frequent market transactions — such as those in electronics or chemicals — are assigned higher MFT Intensity scores, reflecting their reliance on external knowledge acquisition.[26]

Column 1 results shows that Absorber firms operating in fields with 1% higher MFT exposure reduced publications by 6.5%. Leaders, by contrast, exhibited a statistically insignificant 4.3% increase in publications. The decline of Absorber firm science in Figure 8 therefore seems correlated with the growing ease of acquiring external innovation inputs.

Column 2 tests the obverse of this prediction - whether the abstention of universities from invention prior to institutional changes of the 1980s in key fields can explain the rise of Absorber

[24] Firms founded after 1980 do not appear in Figure 8. Hence, the estimates shown refer solely to those firms we classify as Leaders, Followers, or Absorbers.

[25] Traded patents are identified through reassignment records from the USPTO Patent Assignment Dataset. Full detail on the identification process is available in (Arora et al., 2022).

[26] For instance, in 2005, Sun Oil Co. operated across five technology classes with varying degrees of participation in the market for technology (MFT). Its most prolific patent class was C01F (compounds of beryllium), where 28.6% of its patents were granted. 83.3% of all patents in C01F that year were traded, contributing to an MFT Exposure of 29.9%, over five standard deviations above the sample mean that year.

science. "University Patent" is defined at the scientific field-year level and measures the share of knowledge that is translated by academics into inventions. For each scientific field, we weight the annual counts of patents assigned to universities in each CPC by the share of NPL citations originating from that patent class (to the scientific field) from 1945-1980. We then sum the weighted university patents across all patent classes for each field-year, which yields "University Patent". Intuitively, university patents serve as substitutes for Absorptive research, as scientific knowledge is more readily packaged into patents that can be acquired through MFT.

The interaction coefficient estimate in Column 2 show that a unit increase in university patenting is correlated with a 9.4% drop in Absorber publications relative to the sample average. Leaders on the other hand show no statistically meaningful change with respect to university patents. This suggests that Absorbers increased internal research during the Golden Age especially in areas where university participation in inventions were weak.

The results suggest that reforms such as the Bayh-Dole Act that lowered transaction costs for accessing academic patents reduced the marginal value of in-house scientific production. Leaders, however, whose innovation strategies depended on proprietary science, were insulated from — and occasionally leveraged — these shifts. The Column 1 result clarifies the mechanism: Absorbers retreated precisely in subfields where universities or markets supplied relevant, externalizable knowledge, whereas Leaders maintained their scientific foothold. This pattern mirrors the dynamics in Figure 8, where Absorbers' publication-to-sales ratio rises briefly in the late 1970s — a last gasp of internal research — before collapsing as institutional changes reshape the knowledge economy. The divergent trajectories of Absorbers and Leaders underscore the strategic reallocation of innovative effort in response to evolving institutional frontiers.

6 Discussion and Conclusion

Our study integrates a comprehensive historical dataset with ensemble learning method to illuminate the changing roles of American corporations in producing and using scientific knowledge. By classifying firms according to the nature of their scientific engagement — Science Leaders, Absorbers and Followers — we shed new light on why, during the mid-twentieth-century "Golden Age," a small but highly influential cadre of large firms devoted substantial resources to basic research, and a broader population of companies maintained R&D laboratories primarily to absorb and adapt external discoveries for commercial applications.

We show that Leaders – firms using cutting-edge research as a source of internal inventions – were (and remain) relatively rare. They operate on the technological frontier in technologically advanced industries, and have the scale, market power and government contracts to sustain long-horizon R&D investments. Despite their small numbers, these firms have had an outsized impact on the American innovation ecosystem, developing transformative technologies, and in doing so, making significant scientific advances. Meanwhile, other firms — Absorbers — have shifted emphasis over time from in-house science to the acquisition of external knowledge, reflecting the growing

ability of universities and startups to translate new discoveries into near-market prototypes.

Critically, our analysis reveals that the overall decline in corporate science is far from uniform. Rather, it is marked by the disappearance of many once-active science-absorbing labs, coupled with continued (and even intensified) investments by a handful of frontier Leaders. This nuanced perspective clarifies why some industries, such as chemicals and electronics, show significant retrenchment in basic research capacity, whereas others, including information and communications technologies (ICT), are witnessing a new wave of science-led corporate R&D. It also highlights that size and market concentration consistently matter across different eras and sectors: firms that continue to engage in science production often occupy dominant positions, suggesting that a degree of market power may be necessary to sustain costly, long-term research efforts.

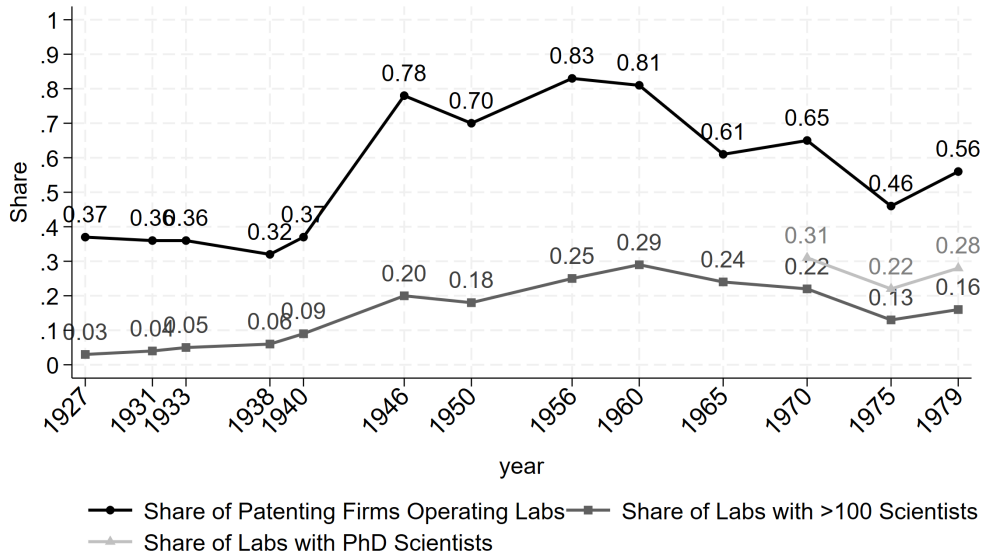
Our second contribution is methodological. By employing ensemble learning techniques and a broad portfolio of indicators — spanning scientific publications, patents, employee rosters from the National Register of Scientific and Technical Personnel (NRSTP), and directories like Industrial Research Laboratories (IRL) and American Men & Women of Science — we can distinguish firms that truly push the scientific frontier from those that rely on assimilating existing research. This novel classification moves beyond binary measures of "research versus development," providing a clearer view of how firms differ in the depth and purpose of their scientific engagement.

Third, our work also contributes valuable new data on corporate science activity during the Golden Age. By linking multiple archival and administrative sources, we document the substantial heterogeneity across sectors and firms in how science was (and still is) produced, absorbed, or bypassed. This dataset opens important avenues for further research: scholars can explore the mobility of scientists between corporations and academia, investigate the structure of collaboration networks among science-producing firms, and assess the impact of corporate research on the rate and direction of technological change.

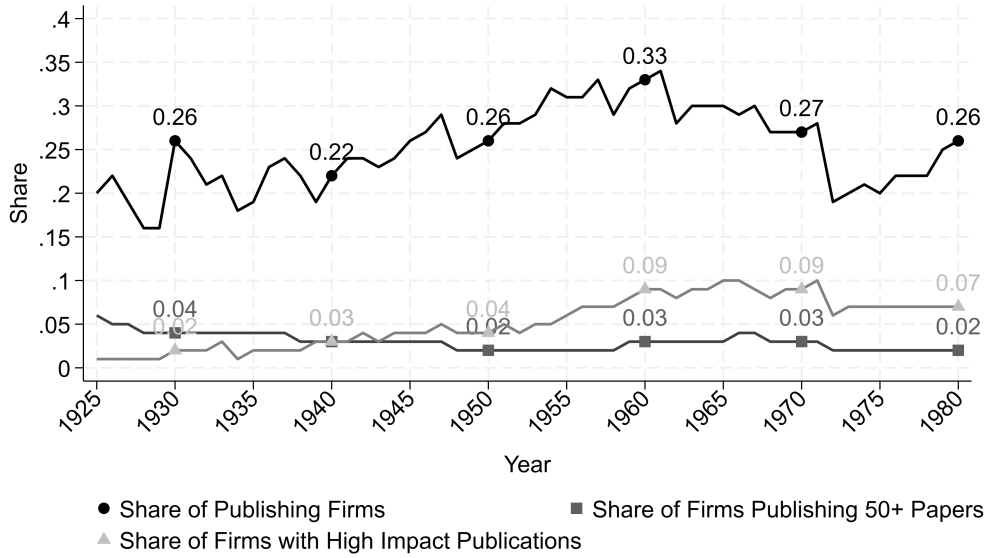
Looking ahead, our findings have significant implications for both policy and theory. Policies that aim to foster corporate scientific activity must weigh the trade-offs between competitive market structures and the realities of funding risky, long-term science. Furthermore, as universities and startups become more sophisticated in commercializing their discoveries, incumbent firms will likely continue to focus on downstream development — except in those sectors operating at or near the scientific frontier, where proprietary, specialized knowledge is still crucial. Ultimately, our results help explain both the historical arc of corporate science withdrawal and the conditions under which large firms still find it profitable — and essential — to invest in pioneering scientific research.

Figure 1: Growth of Industrial Labs and Corporate Publications, 1925-1980

(a) Share of Patenting Firms Operating Labs, by Type of Lab

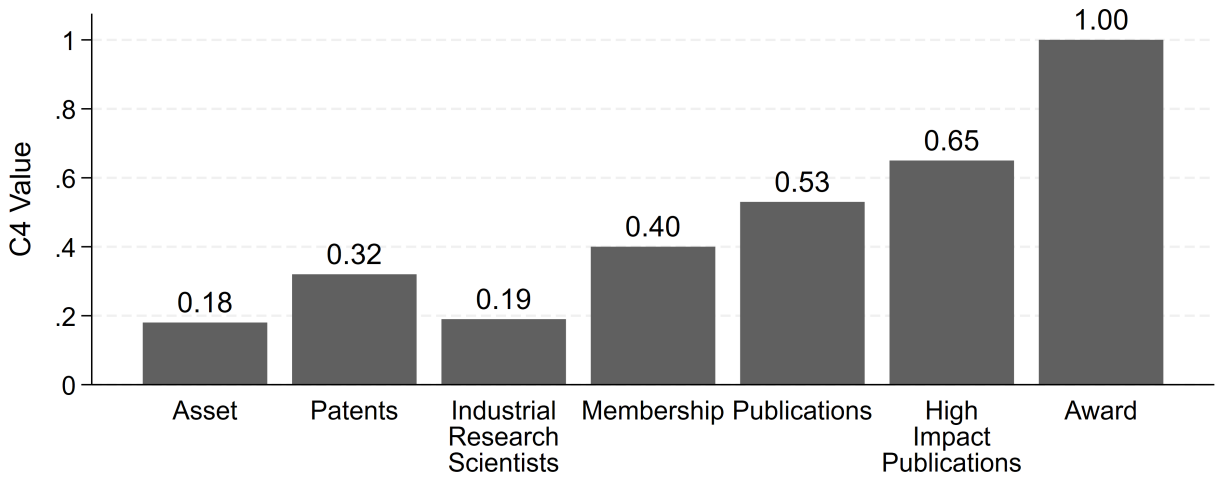


(b) Share of Publishing Firms, by Type of Publication



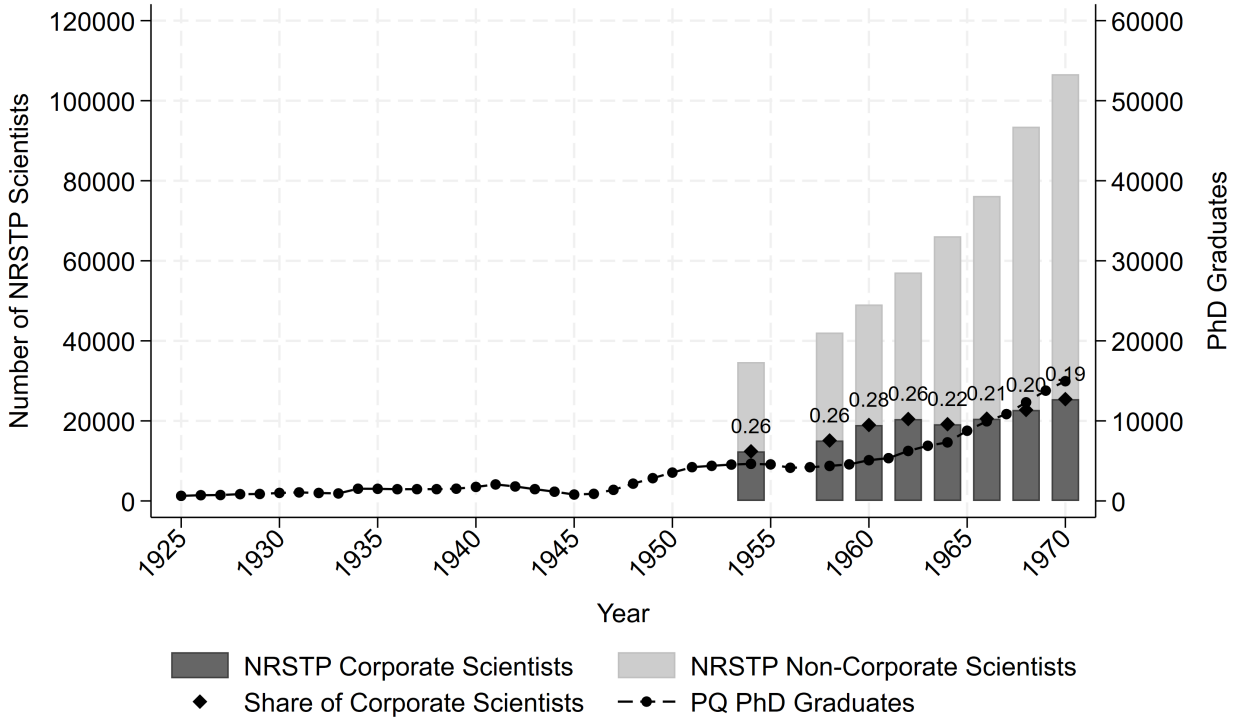
Notes: This figure presents two panels examining the growth of industrial research in our sample of patenting firms from 1925 to 1980. Panel (a) shows three metrics related to industrial laboratories, with each dot representing a different edition of the Industrial Research Laboratory (IRL) directory: the share of firms operating labs (black line); the share of labs with more than 100 scientists (dark gray line); and the share of labs with PhD scientists (light gray line), available for select years after 1965. Panel (b) tracks corporate publication: the overall share of firms publishing (black line); the share of firms publishing 50 or more papers (dark gray line); and the share of firms producing high-impact publications (light gray line), defined as publications receiving forward citations in the top 10th percentile of their OECD subfield.

Figure 2: Four-Firm Concentration Ratio (C4), by Proximity to Science



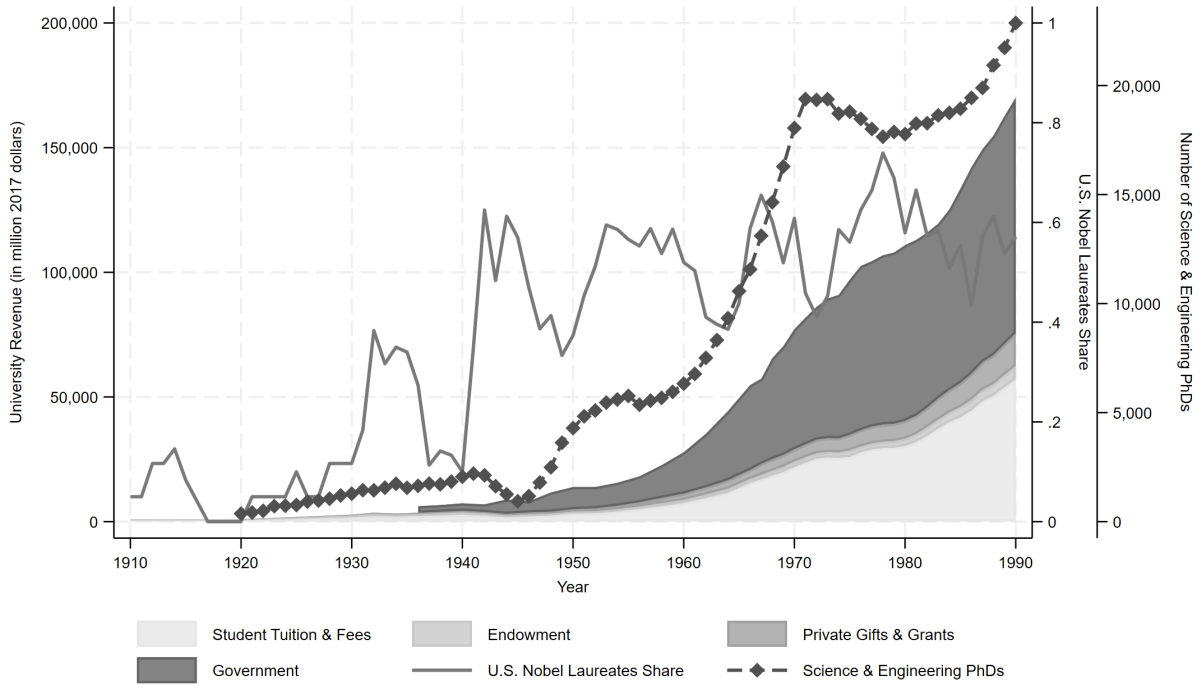
Notes: This figure presents the average C4 concentration ratios (share of top 4 firms) across different scientific and economic metrics for 1945-1980. For each metric, the C4 ratio is first calculated annually by identifying the share of the top 4 firms in that year, then averaged across all years in the sample period. "Membership" refers to the number of corporate scientists that are members of scientific and engineering societies. "High-Impact Publications" refer to number of publications with forward citations in the top 10th percentile of their respective OECD subfield in a given year. "Awards" refer to scientific & engineering awards received by corporate scientists. Details on scientific societies and awards can be found in Appendix A.2.

Figure 3: Production and Employment of Science & Engineering Doctorates, 1925-1970



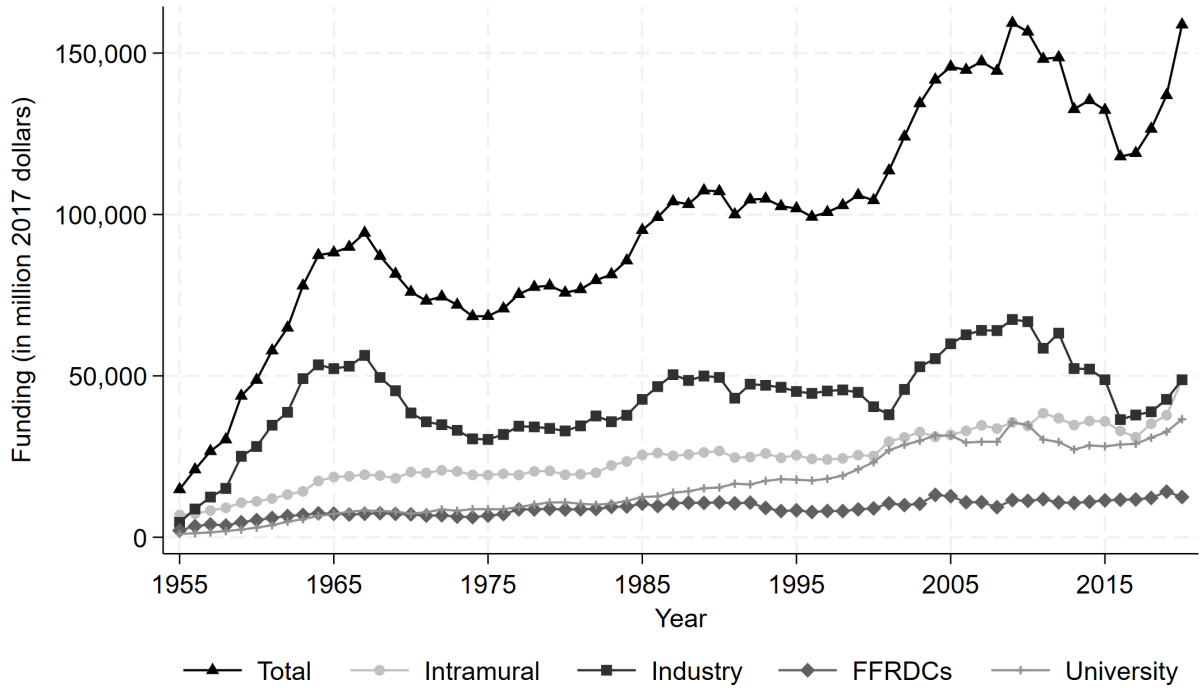
Notes: This figure tracks the production and employment of PhDs from 1925 to 1970 in the United States. The bars represent the total number of scientists in the National Register of Scientific and Technical Personnel (NRSTP) survey, split between those working in industry (dark gray) and non-corporate sectors (light gray). The black diamonds show the proportion of NRSTP survey scientists working in industry. The dashed line with circles shows the number of natural science and engineering PhD graduates from American universities (plotted on the right axis).

Figure 4: Expansion of the U.S. University Sector, 1910-1990)



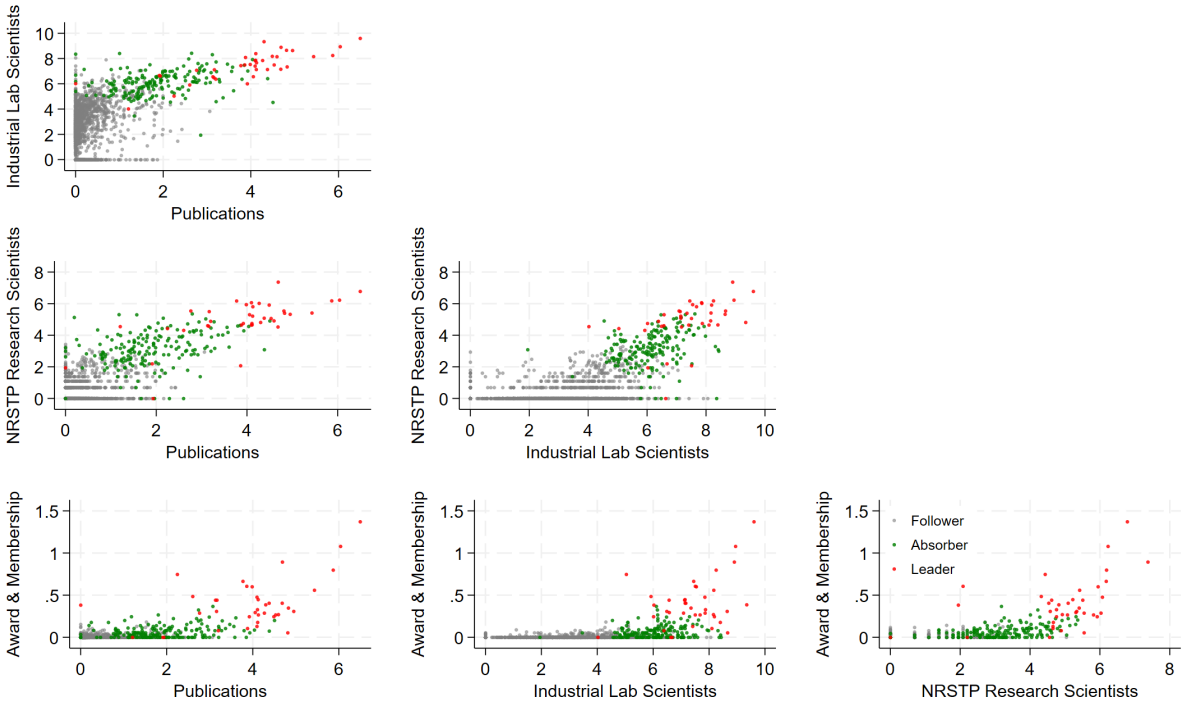
Notes: This figure illustrates primary indicators of U.S. higher education development from 1910 to 1990 by showing, first, the stacked area plots show revenues of higher education institutions (in millions of 2017 dollars) disaggregated into Student Tuition & Fees, Endowment Income, Private Gifts & Grants, and combined Government Support (federal, state, and local). Second, the gray line presents the U.S. share of Nobel Prize laureates in Physics, Chemistry, and Physiology or Medicine, represented as a five-year moving average of the proportion of laureates affiliated with U.S. institutions upon receiving the award; notably, the lower tallies from 1915–1919 and 1940–1944 reflect the Nobel Committee's suspension or limitation of awards during World Wars I and II. Third, the diamond-marked line depicts the number of science and engineering PhDs conferred annually by U.S. institutions. Revenue data are from the U.S. Department of Education, National Center for Education Statistics (Annual Report of the Commissioner, Biennial Survey of Education in the United States, Financial Statistics of Institutions of Higher Education, and Digest of Education Statistics), Nobel laureate information is sourced from the Nobel Prize Foundation's database and historical records, and PhD data come from the U.S. Department of Education (Biennial Survey of Education, Education Directory, Fall Enrollment in Colleges and Universities, Earned Degrees Conferred, IPEDS) and the National Science Foundation's Survey of Earned Doctorates. Dollar values are deflated to 2017 constant dollars using the Bureau of Economic Analysis (BEA) Deflator.

Figure 5: Federal Obligations for R&D, by Performer, 1955-2022



Notes: The figure illustrates federal R&D obligations distributed across five performing sectors: Total, Intramural, Industry, FFRDCs, and University. Intramural activities encompass both the administration costs of intramural R&D programs and extramural R&D procurements by federal personnel, as well as actual intramural research performance. FFRDCs (Federally Funded Research and Development Centers) include facilities administered by industry, universities, and nonprofit organizations, with their funding levels reflecting changes due to the establishment of new centers, discontinuation of existing ones, and modifications in administrator classifications. Data is sourced from National Center for Science and Engineering Statistics, Survey of Federal Funds for Research and Development. Dollar values are deflated to millions of 2017 dollars using the Bureau of Economic Analysis (BEA) Deflator.

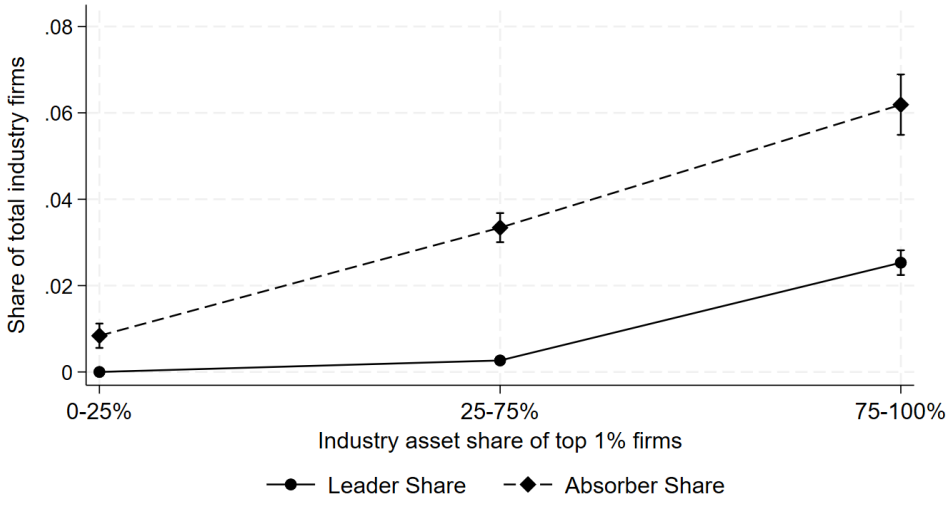
Figure 6: Ensemble Classification Scatter Plots, 1945-1980



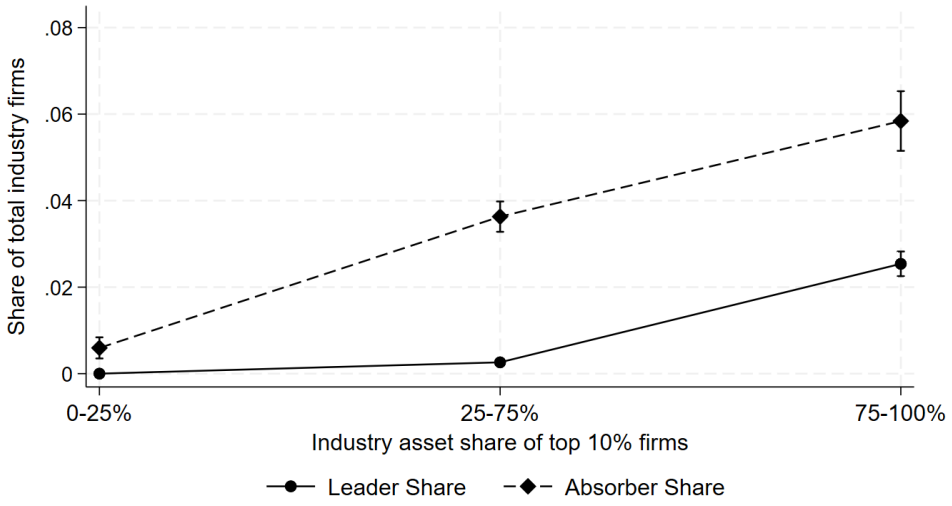
Notes: This figure presents scatter plots showing the relationships between different scientific metrics for firms during 1945-1980, using the ensemble classification method to distinguish between Leaders (red dots), Absorbers (green dots) and Followers (gray dots).

Figure 7: Comparison of Leader and Absorber Shares by Asset Concentration

(a) Leader and Absorber Shares by Top 1% Asset Concentration

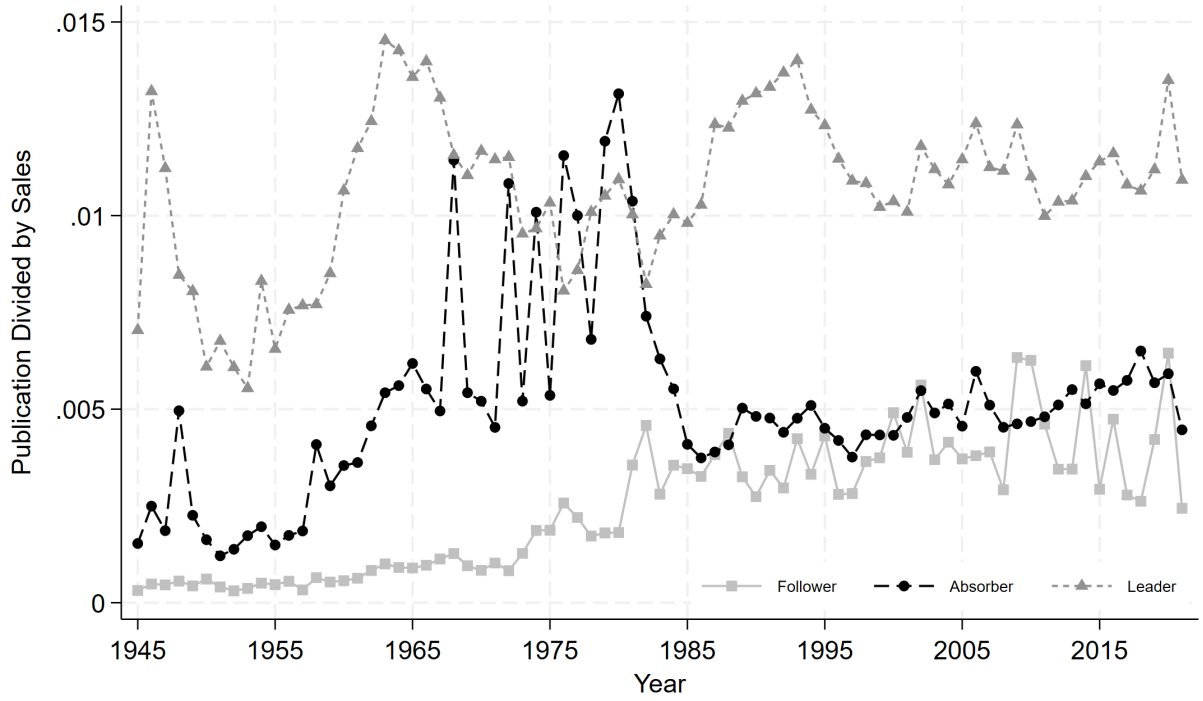


(b) Leader and Absorber Shares by Top 10% Asset Concentration



Notes: This figure examines the relationship between asset concentration and the prevalence of Leader and Absorber firms at the Bureau of Economic Analysis (BEA) Subsector level from 1945-1980. Panel (a) shows firm proportions for top 1% asset concentration, and panel (b) for top 10% asset concentration (from IRS Statistics of Income data (Kwon et al., 2024)). Industries are categorized by concentration percentiles (0-25%, 25-75%, 75-100%). Points represent the average share of Leaders (connected by solid line) and Absorbers (dashed line) out of total firms in each industry-year, with 95% confidence intervals.

Figure 8: Sales Normalized Corporate Publications by Ensemble Group, 1945-2020



Notes: This figure plots ratio of publications to sales for three distinct firm categories from 1945-2020, using the ensemble classification method to distinguish between Leaders (gray triangles with short dashed lines), Absorbers (gray circles with long dashed lines) and Followers (black squares with solid lines).

Table 1: Data Coverage (1945-1980)

Variable	1945-1949					1950-1954					1955-1959					1960-1964					1965-1969					1970-1974					1975-1980					
	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
Publication	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Patent	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Industrial Research Lab		✓				✓						✓				✓					✓					✓					✓				✓	
Lab (PhD Scientists)																										✓					✓				✓	
Lab (Function)																✓					✓					✓					✓				✓	
NER																				✓			✓		✓											
NRSTP																				✓		✓		✓		✓										
American Men & Women of Science	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Award & Membership	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: This table presents the temporal coverage of scientific indicators used in the classification and analysis of firms from 1945-1980. The data availability is shown at an annual frequency, with check marks (✓) indicating years where data is available for each variable.

Table 2: Summary Statistics for Main Variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
					Distribution		
	NumFirms	Obs.	Mean	Std. dev.	10th	50th	90th
Patents	2,899	51,369	12	51	0.0	0.0	23.0
Publications	2,899	51,369	4	31	0.0	0.0	3.0
Sales ($ mm)	2,781	40,398	2,134	7,038	64.4	504.4	4,724.1
Assets ($ mm)	2,782	40,563	2,203	9,954	48.3	379.1	4,439.8
Tobin's Q	2,744	39,160	1	1	0.2	0.6	1.7
Scientists from Survey	2,903	51,590	20	109	0.0	0.0	31.0
Research Scientists from Survey	2,903	51,590	11	63	0.0	0.0	14.0
Manager Scientists from Survey	2,903	51,590	2	10	0.0	0.0	3.0
Number of Lab Engineers/Scientists	2,815	11,279	174	980	0.0	0.0	257.0
Operate An Industrial Lab	1,758	35,764	1	0	1.0	1.0	1.0
Labs Working for Government	244	244	1	0	1.0	1.0	1.0
AMWS Scientists	2903	51,590	49	283	0.0	0.0	85.0

Notes: This table provides summary statistics for the main variables used in the econometric analyses at the firm-year level from 1945 to 1980. All monetary values are expressed in constant 2017 dollars using the Bureau of Economic Analysis (BEA) Deflator. Details on variable definition and data sources can be found in Table B.1.

Table 3: Ensemble Classification Results

	1945-1965	1965-1980	1945-1980
% of Leader	0.69%	1.87%	1.38%
% of Absorber	8.41%	14.20%	5.99%
% of Follower	90.90%	83.93%	92.63%

Notes: This table reports the proportion of firms classified into three categories — Leaders, Absorbers, and Followers — across different time periods from 1945 to 1980. The proportions are reported as percentages for each time period. All values are rounded to two decimal places.

Table 4: Corporate Production of Science and Utilization of Internal and External Scientific Knowledge

	Pr(Citing External Science)	Pr(Citing Internal Science)	Distance to External Science	Distance to Internal Science	FOR Main Similarity	FOR Sub Similarity	Science Citation Recency
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Absorber	0.031	0.002	-0.455	-0.460	0.095	0.144	-0.751
	(0.006)	(0.001)	(0.073)	(0.128)	(0.036)	(0.035)	(0.558)
Leader	0.035	0.006	-0.647	-1.072	0.178	0.255	-2.856
	(0.010)	(0.001)	(0.131)	(0.188)	(0.056)	(0.053)	(0.683)
Log(Assets)	-0.003	0.000	0.041	-0.017	0.030	0.002	0.396
	(0.002)	(0.000)	(0.029)	(0.035)	(0.015)	(0.015)	(0.173)
Average of Dep Var	0.034	0.000	1.527	3.330	-0.205	-0.208	11.807
2-Digit SIC Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.091	0.046	0.141	0.202	0.096	0.081	0.085
Number of Obs	2,703	2,703	1,219	589	1,317	1,274	858

Notes: This table examines how Science Leaders and Science Absorbers differ in their utilization of external versus internal scientific knowledge across four dimensions. The dependent variable (DV) for Columns 1 and 2 are the share of firm patents citing external scientific publications (publications by other organizations) and internal scientific publications (publications by the focal firm) respectively. The DV in Columns 3 and 4 measure average citation distance to external and internal science respectively (Ahmadpoor & Jones, 2017), where distance refers to the minimum number of citation links between patents and scientific publications (smaller distances indicate greater proximity). The DV in Columns 5 and 6 measure textual similarity between corporate and academic papers using normalized FOR (Field of Research) similarity scores at the maincategory level (Column 5) and subcategory level (Column 6). We first limit the comparison set to the top 10% most textually similar academic publications to each corporate publication. We then average the similarity metrics at the (corporate) publication level and then at the firm level. DV in Column 7 refers to the average lag between a patent's grant year and publication year of its cited articles. The patent level lags to cited papers are averaged at the firm-year level and then to the firm level. Lower values indicate that the firm cites more recent scientific publications. The key independent variables are indicators for Leader and Absorber firms, with Follower firms serving as the baseline category. All specifications include controls for logged Assets and 2-digit SIC industry fixed effects. Standard errors are clustered at the firm level.

Table 5: Corporate Science by Technological and Market Leadership

	Technological Leadership		Market Leadership	
	(1) First CPC Class Patents	(2) Breakthrough Patents	(3) ln(Assets)	(4) Market Share
Leader	3.539	21.996	2.235	0.006
	(7.905)	(56.891)	(0.087)	(0.002)
Absorber	289.747	2200.876	3.303	0.036
	(72.373)	(526.298)	(0.234)	(0.014)
Log(Assets)	1.857	10.398		0.005
	(0.749)	(5.254)		(0.001)
Log(Number of Patents + 1)	17.170	124.213		
	(3.171)	(22.874)		
Average of Dep Var	9.469	67.375	5.405	0.005
2-Digit SIC Dummies	Yes	Yes	Yes	Yes
R2	0.365	0.379	0.401	0.338
Number of Obs	2,774	2,774	2,778	2,775

Notes: Unit of analysis is at the firm level. Sample consists of patenting U.S. listed firms from 1945 to 1980. Dependent variables for Columns 1 and 2 are Technological Leadership measures: "First CPC Class Patent" counts the number of corporate patents that were the first patents ever granted in their respective CPC (Cooperative Patent Classification) classes. "Breakthrough Patents" are patents in the top 20% of forward-to-backward similarity ratios (Kelly et al., 2021). Forward similarity (FSim5) measures text similarity to future patents filed within 5 years, while backward similarity (BSim5) measures similarity to patents from the previous 5 years. DV for Columns 3 and 4 are Market Leadership measures. "Market Share" refers to firm sales from Compustat and Graham et al. (2015) divided by total business receipts in the firm's focal Bureau of Economic Analysis (BEA) subsector (approximately at the 2-digit SIC industry level, further details are available in Kwon et al. (2024)). "Leader" and "Absorber" are dummies indicating the firm's ensemble classification, with Follower firms serving as the baseline category. All specifications include 2-digit SIC industry fixed effects. Standard errors clustered at the firm level. All specifications except Column 3 include controls for firm size (Log(Assets), in 2017 dollars). Columns 1 and 2 include controls for logged number of patents.

Table 6: Corporate Science and Government Demand

	Government Patents	Industrial Lab for Government	Defense Contract Share
	(1)	(2)	(3)
Absorber	0.235	0.076	0.080
	(0.048)	(0.066)	(0.041)
Leader	1.014	0.296	0.111
	(0.177)	(0.117)	(0.040)
Log(Assets)	-0.006	-0.078	-0.007
	(0.004)	(0.022)	(0.002)
Log(Number of Patents + 1)	0.113	0.015	-0.014
	(0.012)	(0.025)	(0.007)
Log(Industrial Lab Scientists)	0.007	0.094	0.012
	(0.003)	(0.019)	(0.002)
Average of Dep Var	0.080	0.380	0.024
2-Digit SIC Dummies	Yes	Yes	Yes
R2	0.488	0.216	0.100
Number of Obs	2,733	621	2,504

Notes: Unit of analysis is at the firm level. Sample consists patenting U.S. listed firms from 1945 to 1980. "Government Patents" refer to firm patents licensed to or held by government entities. "Industrial Lab for Government" is a dummy equal to one if the research facilities that either fully or partially dedicated their research activities to government-sponsored projects, based on the 1965 Industrial Research Laboratories directory; and "Defense Contract Share" divides the dollar value of Department of Defense procurement contracts won by the focal firm by its sales. "Leader" and "Absorber" are dummies indicating the firm's ensemble classification. The Follower firms serve as the baseline category. All specifications control for the log number of patents, log number of industrial lab scientists, and firm size (Log(Assets)). Column 2 is restricted to the subsample of firms that operate industrial laboratories. All specifications include 2-digit SIC industry fixed effects, with standard errors clustered at the firm level reported in parentheses.

Table 7: Corporate Publications and Markets for Technology

	Publication (1980-2015)	Publication (1945-1980)
	(1)	(2)
Absorber	26.091	0.478
	(18.385)	(0.139)
Leader	323.668	4.076
	(132.908)	(0.699)
MFT Exposure	-82.420	
	(66.762)	
Absorber × MFT Exposure	-300.818	
	(165.051)	
Leader × MFT Exposure	199.219	
	(1562.580)	
University Patent		-0.018
		(0.054)
Absorber × University Patent		-0.079
		(0.032)
Leader × University Patent		-0.005
		(0.127)
Log(Assets)	16.259	0.505
	(3.683)	(0.143)
Average of Dep Var	46.605	0.839
Year FE	Yes	Yes
2-Digit SIC Dummies	No	Yes
OECD-Subfield FE	No	Yes
R2	0.301	0.100
Number of Obs	15,474	209,735

Notes: Column 1 presents the analysis for the modern period (1980–2015) at the firm-year level. (Firms founded after 1980 are excluded from this Column.) Column 2 reports the analysis for the main sample period (1945–1980) at the firm–scientific field–year level. "University Patent," defined at the field–year level, refers to a scientific field's exposure to university inventions each year. For each OECD subfield, we first calculate the share of NPL citations made from each USPC class between 1945 and 1980, then multiply the number of annual patents assigned to universities in each USPC class by this share and sum these weighted academic patents at the OECD subfield–year level. "MFT Exposure" captures a firm's exposure to an active market for technology. For each CPC class, we calculate the fraction of patents traded each year and multiply these shares by the firm's patent share in that CPC-year. We sum these weighted shares of traded patents at the firm–year level to produce MFT Exposure. All specifications control for the log number of patents and include year fixed effects and clustered standard errors at the firm level. The "Follower" firms form the omitted category.

References

Aghion, P., & Tirole, J. (1994). The management of innovation. *The Quarterly Journal of Economics*, *109*(4), 1185–1209.

Ahmadpoor, M., & Jones, B. F. (2017). The dual frontier: Patented inventions and prior scientific advance. *Science*, *357*(6351), 583–587.

Airoldi, A., & Moser, P. (2024). *Inequality in science: Who becomes a star?* (Working Paper No. 33063). National Bureau of Economic Research.

Alic, J. (2007). *Trillions for military technology: How the pentagon innovates and why it costs so much*. Springer.

American Association for the Advancement of Science. (2022). Historical Trends in Federal R&D: Defense and Nondefense R&D Outlays, 1953-2023. *American Association for the Advancement of Science*. Retrieved August 1, 2024, from https://www.aaas.org/programs/r-d-budget-and-policy/historical-trends-federal-rd

American cyanamid company technical bulletins. (1988). Retrieved August 7, 2024, from https://www.lib.ncsu.edu/findingaids/mc00315/access

Argonne National Laboratory. (2025). Our history. Retrieved January 16, 2025, from https://www.anl.gov/our-history

Arora, A., Belenzon, S., Cioaca, L. C., Sheer, L., & Zhang, H. (2023). *The effect of public science on corporate r&d* (Working Paper No. 31899). National Bureau of Economic Research.

Arora, A., Belenzon, S., & Dionisi, B. (2023). First-mover advantage and the private value of public science. *Research Policy*, *52*(9), 104867.

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., Patacconi, A., & 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., & Suh, J. (2022). Science and the market for technology. *Management Science*, *68*(10), 7176–7201.

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

Arora, A., Fosfuri, A., & Gambardella, A. (2004). *Markets for technology: The economics of innovation and corporate strategy*. MIT press.

Auerswald, P. E., & Branscomb, L. M. (2004). Reflections on mansfield, technological complexity, and the "golden age" of us corporate r&d. *The Journal of Technology Transfer*, *30*, 139–157.

Australian Bureau of Statistics. (2020). Australian and New Zealand standard research classification (ANZSRC), 2020. Retrieved February 13, 2025, from https://www.abs.gov.au/statistics/classifications/australian-and-new-zealand-standard-research-classification-anzsrc/2020

Autor, D., Dorn, D., Katz, L. F., Patterson, C., & Van Reenen, J. (2020). The fall of the labor share and the rise of superstar firms. *The Quarterly Journal of Economics*, *135*(2), 645–709.

Bardeen, J., & Brattain, W. H. (1948). The transistor, a semi-conductor triode. *Physical Review*, *74*(2), 230–231.

Barron, J. L. (1959). The integration of industrial sanitation. *The Sanitarian*, *21*(4), 210–214.

Belenzon, S., & Cioaca, L. C. (2021). *Guaranteed markets and corporate scientific research* (Working Paper No. 28644). National Bureau of Economic Research.

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, *35*(8), 1798–1828.

Bilstein, R. E. (1999). *Stages to saturn: A technological history of the apollo/saturn launch vehicle*. Diane Publishing.

Booch, Grady. (1977). John Backus. *Association for Computing Machinery*. Retrieved August 17, 2024, from https://amturing.acm.org/award_winners/backus_0703524.cfm

Boroush, Mark. (2021). *National patterns of r&d resources: 2018–19 data update*. Retrieved May 6, 2024, from %5Curl%7Bhttps://ncses.nsf.gov/pubs/nsf21325%7D

Brookhaven National Laboratory. (2025a). *Our history*. Retrieved February 21, 2025, from https://www.bnl.gov/about/history/

Brookhaven National Laboratory. (2025b). *Our History of Discovery*. Retrieved January 21, 2025, from https://www.bnl.gov/about/history/discoveries.php

Brookhaven National Laboratory. (2025c). *Reactors as research tools*. Retrieved January 21, 2025, from https://www.bnl.gov/about/history/reactors.php

Brooks, C. G., Grimwood, J., & Swenson, L. S. (1980). *Chariots for apollo: A history of manned lunar spacecraft*.

Cohen, W. M., & Levinthal, D. A. (1989). Innovation and learning: The two faces of r&d. *The economic journal*, *99*(397), 569–596.

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative science quarterly*, *35*(1), 128–152.

Cooley, J. W., & Tukey, J. W. (1965). An algorithm for the machine calculation of complex fourier series. *Mathematics of Computation*, *19*(90), 297–301.

De Loecker, J., Eeckhout, J., & Unger, G. (2020). The rise of market power and the macroeconomic implications. *The Quarterly Journal of Economics*, *135*(2), 561–644.

Dietterich, T. G. (2000). Ensemble methods in machine learning. *International workshop on multiple classifier systems*, 1–15.

Dong, X., Yu, Z., Cao, W., Shi, Y., & Ma, Q. (2020). A survey on ensemble learning. *Frontiers of Computer Science*, *14*, 241–258.

Dreier, C. (2022). An improved cost analysis of the apollo program. *Space Policy*, *60*, 101476.

Eichhorn, J. (1964). Synergism of free radical initiators with self-extinguishing additives in vinyl aromatic polymers. *Journal of Applied Polymer Science*, *8*(6), 2497–2524.

Fagen, M. D. (1978). *Technology in support of national security: Mit lincoln laboratory*. MIT Lincoln Laboratory. Retrieved January 9, 2025, from https://www.ll.mit.edu/sites/default/files/other/doc/2018-04/MIT_Lincoln_Laboratory_history_book.pdf

Feistel, H. (1973). *Cryptographic coding for data-bank privacy* (Research Report RC-2827). IBM Thomas J. Watson Research Center.

GE Vernova. (2024). Large boiling water reactors. *GE Hitachi Nuclear Energy*. Retrieved August 13, 2024, from https://www.gevernova.com/nuclear/carbon-free-power/large-reactors

Gertner, J. (2012). *The idea factory: Bell labs and the great age of american innovation*. Penguin Press.

Giaever, I. (1960). Energy gap in superconductors measured by electron tunneling. *Physical Review Letters*, *5*(4), 147–148.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

Graham, J. R., Leary, M. T., & Roberts, M. R. (2015). A Century of Capital Structure: The Leveraging of Corporate America. *Journal of Financial Economics*, *118*(3), 658–683.

Gross, D. P., & Sampat, B. N. (2023). America, jump-started: World war ii r&d and the takeoff of the us innovation system. *American Economic Review*, *113*(12), 3323–3356.

Gross, D. P., & Sampat, B. N. (2024). *The government patent register: A new resource for measuring us government-funded patenting* (tech. rep.). National Bureau of Economic Research.

Haigh, Thomas. (1973). Charles w. Bachman. *Association for Computing Machinery*. Retrieved August 17, 2024, from https://amturing.acm.org/award_winners/bachman_9385610.cfm

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural computation*, *18*(7), 1527–1554.

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *science*, *313*(5786), 504–507.

Jet Propulsion Laboratory. (2025). History of the jet propulsion laboratory. Retrieved January 7, 2025, from https://www.jpl.nasa.gov/who-we-are/history

Kelly, B., Papanikolaou, D., Seru, A., & Taddy, M. (2021). Measuring Technological Innovation over the Long Run. *American Economic Review: Insights*, *3*(3), 303–20.

Kim, S. D., & Moser, P. (2021). *Women in science. lessons from the baby boom* (Working Paper No. 29436). National Bureau of Economic Research.

Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

Kogan, L., Papanikolaou, D., Seru, A., & Stoffman, N. (2017). Technological Innovation, Resource Allocation, and Growth. *The Quarterly Journal of Economics*, *132*(2), 665–712.

Kohonen, T. (2013). Essentials of the self-organizing map. *Neural networks*, *37*, 52–65.

Koppers. (n.d.). Modern gas manufacturing/sng process description/koppers k-t coal-gasification process. *Pipeline Gas J.* Retrieved November 17, 2024, from https://www.osti.gov/biblio/7233455

Kwon, S. Y., Ma, Y., & Zimmermann, K. (2024). 100 years of rising corporate concentration. *American Economic Review*, *114*(7), 2111–2140.

Lawrence Livermore National Laboratory. (2025a). Our history. Retrieved January 16, 2025, from https://www.llnl.gov/about/history

Lawrence Livermore National Laboratory. (2025b). Our history. Retrieved January 16, 2025, from https://www.llnl.gov/purpose/history/1970s

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324.

Lee, J. (2003). Innovation and strategic divergence: An empirical study of the us pharmaceutical industry from 1920 to 1960. *Management Science*, *49*(2), 143–159.

Marx, M., & Fuegi, A. (2020). Reliance on Science: Worldwide Front-Page Patent Citations to Scientific Articles. *Strategic Management Journal*.

Maveety, M. (1945). Correlation of keeping properties of shortening with keeping quality of biscuits. *Quartermaster Corps Manual, Committee on Food Research, Subsistence Research and Development Laboratory, Deterioration of Fats and Oils*, 136.

Mccellan, P. (1940). Melamine preparation. *Industrial & Engineering Chemistry*, *32*(9), 1181–1186.

Mienye, I. D., & Sun, Y. (2022). A survey of ensemble learning: Concepts, algorithms, applications, and prospects. *IEEE Access*, *10*, 99129–99149.

MIT Lincoln Laboratory. (2025). History of mit lincoln laboratory. Retrieved March 13, 2025, from https://www.ll.mit.edu/about/history

Moser, P., & San, S. (2020). Immigration, science, and invention. lessons from the quota acts. *Lessons from the Quota Acts (March 21, 2020)*.

Moser, P., Voena, A., & Waldinger, F. (2014). German jewish émigrés and us invention. *American Economic Review*, *104*(10), 3222–3255.

Mowery, D. C. (2010). Military r&d and innovation. In *Handbook of the economics of innovation* (pp. 1219–1256). Elsevier.

Mowery, D. C. (2012). Defense-related r&d as a model for "grand challenges"' technology policies. *Research policy*, *41*(10), 1703–1715.

Murtagh, F. (1983). A survey of recent advances in hierarchical clustering algorithms. *The computer journal*, *26*(4), 354–359.

National biscuit company (nabisco) annual reports. (1960). Retrieved December 21, 2024, from https://archive.org/details/nationalbiscuitcompanynabiscoannualreports/nationalbiscuit1960/

National biscuit company (nabisco) annual reports. (1965). Retrieved December 21, 2024, from https://archive.org/details/nationalbiscuitcompanynabiscoannualreports/nationalbiscuit1965/

National Science Foundation. (2019). *Business research and development and innovation: 2019*. Retrieved January 21, 2025, from https://ncses.nsf.gov/pubs/nsf22343

Nelson, L. S. (1969). Editorial policies and perspectives.

Nelson, R. R. (1962). The link between science and invention: The case of the transistor. *The Rate and Direction of Inventive Activity: Economic and Social Factors*, 549–583.

Nelson, R. R., & Langlois, R. N. (1983). Industrial innovation policy: Lessons from american history. *Science*, *219*(4586), 814–818.

Ng, A., Jordan, M., & Weiss, Y. (2001). On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, *14*.

Noyes, T., & Dickinson, W. E. (1956). Engineering design of a magnetic-disk random-access memory. *Proceedings of the Western Joint Computer Conference*, 42–44.

Nuclear Engineering Division, Argonne National Laboratory. (2013). *Argonne's nuclear science and technology legacy* (tech. rep.). Argonne National Laboratory. Retrieved January 16, 2025, from https://www.ne.anl.gov/About/hn/news960121.shtml

Office of NIH History and Stetten Museum. (2025). A short history of the national institutes of health. *National Institutes of Health*. Retrieved January 7, 2025, from https://history.nih.gov/display/history/A+Short+History+of+the+National+Institutes+of+Health

Penzias, A. A., & Wilson, R. W. (1965). A measurement of excess antenna temperature at 4080 mc/s. *The Astrophysical Journal*, *142*, 419–421.

Pinaya, W. H. L., Vieira, S., Garcia-Dias, R., & Mechelli, A. (2020). Autoencoders. In *Machine learning* (pp. 193–208). Elsevier.

Png, I. P. (2019). Us r&d, 1975–1998: A new dataset. *Strategic Management Journal*, *40*(5), 715–735.

Pugh, E. W. (2009). *Building ibm: Shaping an industry and its technology*. MIT press.

Redmond, K. C., & Smith, T. M. (1983). *Jpl and the american space program: A history of the jet propulsion laboratory*. Yale University Press.

Redmond, K. C., & Smith, T. M. (2000). *From whirlwind to mitre: The r&d story of the sage air defense computer*. MIT Press.

Reynolds, D. A., et al. (2009). Gaussian mixture models. *Encyclopedia of biometrics*, *741*(659-663).

Rinner, T. E. (2003). Protecting minor improvements on core patents: Complementing traditional patent protection with strategic disclosure, 2 j. marshall rev. intell. prop. l. 398 (2003). *UIC Review of Intellectual Property Law*, *2*(2), 10.

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

Salakhutdinov, R., & Hinton, G. (2009). Deep boltzmann machines. *Artificial intelligence and statistics*, 448–455.

Schawlow, A. L., & Townes, C. H. (1958). Infrared and optical masers. *Physical review*, *112*(6), 1940.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423.

Simons, J., Randall, D., & Archer, S. (1939). Hydrogen fluoride as a condensing agent. vii. 1 the acylation of aromatic compounds. *Journal of the American Chemical Society*, *61*(7), 1795–1796.

Smillie, Keith. (1979). Kenneth E. Iverson. *Association for Computing Machinery*. Retrieved August 17, 2024, from https://amturing.acm.org/award_winners/iverson_9147499.cfm

Smolensky, P., et al. (1986). Information processing in dynamical systems: Foundations of harmony theory.

Stambaugh, J. (2024). Overview of high-power and wideband radar technology development at mit lincoln laboratory. *Remote Sensing*, *16*(9), 1530.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

White, E., Murray, M., & Maveety, D. (1946). The change in thiamin content of breakfast cereals during cooking.

Willis A. Jensen, F. T., Douglas C. Montgomery, & Vining, G. G. (2018). 50 years of the journal of quality technology. *Journal of Quality Technology*, *50*(1), 2–16.

Zhang, T., Ramakrishnan, R., & Livny, M. (1996). Birch: An efficient data clustering method for very large databases. *ACM sigmod record*, *25*(2), 103–114.

For Online Publication

Appendix A Data Appendix

A.1 Measures of Scientific Inputs

A.1.1 Industrial Research Laboratories

The Industrial Research Laboratories (IRL) directory is a key reference resource documenting corporate R&D laboratories throughout the United States. Originally published by the National Academy of Sciences, the directory encompasses information on laboratories' research focus areas, principal scientists and engineers, and facility locations. Data are gathered via surveys sent to research directors, corporate communications officers, and administrative staff, who provide direct input on their organizations' research activities, personnel, and facilities. This data-collection process allows the IRL directory to present an up-to-date portrait of corporate research laboratories, while oversight by the National Academy of Sciences — an independent institution — further reinforces the directory's reliability and credibility.

Table A1: Industrial Research Laboratories of the United States

Year	Edition	Number of industrial laboratories	Labs being matched to CRSP firms
1920	1	297	
1921	2	526	
1927	3	999	
1931	4	1,620	
1933	5	1,562	
1938	6	1,769	
1940	7	2,264	
1946	8	2,443	621
1950	9	2,845	688
1956	10	4,060	835
1960	11	4,145	974
1965	12	3,115	877
1970	13	3,122	1082
1975	14	3,216	1005
1979	16	6,293	1123

Notes: This table showcases the editions of IRL being collected.

Building upon the work of Png (2019), who digitized the Industrial Research Laboratories (IRL) directory for the period from 1975 to 1998, creating a dataset of 2,805 companies and 8,525 associated facilities, we expanded the digitization effort to cover earlier editions of the IRL directory. Our dataset focuses on the IRL data ranging from 1920 to 1979, providing a complementary perspective on the evolution of industrial research in the United States over a longer time frame.

We digitized the following IRL editions: 1920 (1st edition), 1921 (2nd edition), 1927 (3rd edition), 1931 (4th edition), 1933 (5th edition), 1938 (6th edition), 1940 (7th edition), 1946 (8th edition), 1950 (9th edition), 1956 (10th edition), 1960 (11th edition), 1965 (12th edition), 1970 (13th edition), 1975 (14th edition), and 1979 (16th edition). Each edition contains a varying number of industrial research laboratories and their subdivisions, as shown in TableA1. The 1920 edition lists 297 laboratories, while the 1979 edition includes an impressive 6,293 entries, reflecting the substantial growth and development of industrial research facilities over the six decades covered by our dataset.

GENERAL CONNECTORS CORPORATION
233 W. Orange Grove Ave.
Burbank, California 91503

Chairman of the Board: Harold Shire

Fields of Interest: Design and development of pneumatic ducting, hydraulic connectors, high temperature/pressure materials for aircraft, including jets, missiles, and submarines.
Major Activity of Company: Mfg. 85%, R&D 10%, T&E 5%.
R&D performed for: Owners 50%, Others 50%.
Facilities: Complete lab for rubber and metal used in expansion joints and ducting.
Research Personnel:
Jack Kramer, Pres. (Design Engrg.)
Harold Shire, Ch. of Board (Math.)
T. E. Ellison, Engr. (Pneumatics)
R. Stanley Allison, Engr. (Pneumatics)
David Garcia, Engr. (Rubber)
Robert Larsen, Engr. (Welding)
Laboratory Staff: Chemists 1, mech. engr. 2, welding engr. 1, design engr. 1, math. 1, techn. 6, other lab staff 3.
Recruiting Contact: Harold Shire, Chairman of Board

GENERAL DEVICES, INCORPORATED
Princeton, New Jersey

President: John F. Brinster

Fields of Interest: Research, design and development of aerospace telemetry, multichannel instrumentation, and data acquisition equipment.
Major Activity of Company: Mfg. 70%, R&D 25%, T&E 5%.
R&D performed for: Owners 80%, U.S. Govt. 20%.
Facilities: Complete equipment for electromechanical and electronic product development.
Research Personnel:
John F. Brinster, Pres.
Howard K. Schoenwetler, V.P. Electronics & Syst.
William B. Ely, Head, Adv. Dev.
Laboratory Staff: Electronic engr. 8, electro-mech. 2, physicists 2, techn. 10, other lab staff 5.
Recruiting Contact: R. Winship, Personnel Mgr.

GENERAL DYNAMICS CORPORATION
One Rockefeller Plaza
New York, New York 10020

President and Chairman of the Board: Roger Lewis

Fields of Interest: Earth and aerospace sciences, including but not limited to: hydrodynamics, aerodynamics, hypersonics, thermodynamics, geophysics, mechanics, solid state physics, acoustics, space physics, organic chemistry, mathematics, structural methods and techniques, structural materials, fuels, refractories, composites, alloys, coatings, bonds, seals, human factors, life support, psychology, space medicine, and operations/systems research and techniques; energetic sciences, including but not limited to: fission and controlled thermonuclear energy, nuclear physics, metallurgy, radiochemistry, nuclear, electromechanical, and chemical power and propulsion, ion and electric propulsion, direct conversion, plasma and theoretical physics; information sciences, including but not limited to: data and information acquisition, transmission, processing, display and read-out, techniques and processes, and instrumentation.
Major Activity of Company: Mfg. 81%, R&D 19%.
R&D performed for: Owners 13%, U.S. Govt. 87%.
Facilities: Complete and modern to support the R&D activity and interests.
Research Personnel:
E. H. Heinemann, V.P. Engrg. & Program Dev.
W. J. Eding, Staff Asst. (Rsch. & Dev. Programs)
H. T. Loeser, Staff Asst. (Marine Supt.)
J. L. Bowers, Staff Asst. (Electronic Syst.)
G. L. Armstrong, Staff Asst. (Missles & Space Supt.)
K. Van Every, Staff Asst. (Aeronautical Syst.) General
Laboratory Staff: Chemists 159, M.D.'s 9, engr. 8,572, math. 723, metallurgists 105, physicists 479, psychologists 18, other professionals 257, techn. & tech. support 5,957, other lab staff 3,218.

-GD/Astronautics
P.O. Box 1128
San Diego, California 92112
Fields of Interest: Flight sciences, including complete vehicle and ground systems research, design, and development; life sciences, including design and development of manned space systems; space sciences, including applications of basic sciences to space environments and space systems.
Major Activity of Parent Company: Mfg. - of Lab Div: Mfg. 56%, R&D 36%, T&E 8%.
R&D performed for: Owners 6%, U.S. Govt. 94%.
Facilities: Over 70 different completely equipped labs covering all fields of aerospace R&D.
Research Personnel:
R. C. Sebold, V.P. Rsch, Dev. & Engrg.
H. F. Dunholtar, Dir. Rsch. & Advanced Tech.
W. W. Withee, V.P. Engrg.
K. J. Bossart, Tech. Dir.
K. A. Shricke, Dir, Advanced Studies
Laboratory Staff: Biologists 7, chemists 19, M.D.'s 4, engr. 804, aero. engr. 134, elec/electronic engr. 366, mech. engr. 260, math. 104, metallurgists 12, physicists 68, psychologists 11, other professionals 100, techn. 554, other lab staff 127.
Recruiting Contact: M. V. Wisdom, Dr, Industrial Relations

-GD/Convair
P.O. Box 1950
San Diego, California 92112
President: J. H. Famme
Fields of Interest: Aerodynamic and hydrodynamic sciences; thermodynamics; structures and materials; electromagnetic propagation, electronic systems; nuclear and solid state physics; space flight mechanics, rocket and jet propulsion.
Major Activity of Parent Company: Mfg. - of Lab Div: Mfg. 85%, R&D 12%, T&E 3%.
R&D performed for: Owners 65%, U.S. Govt. 20%, Others 15%.
Facilities: Well equipped labs for aerodynamics, material and processes, structures, hydrodynamics, thermodynamics, manufacturing research; standard as well as supporting design, instrumentation, and fabrication functions.
Research Personnel:
C. W. Frick, V.P. Engrg. (Aerodynamics)
W. T. MacCarthy, Mgr. of Rsch. Labs (Aerodynamics)
E. F. Strong, Chief of Material & Performance Labs (Structures)
B. R. Parkin, Mgr. of Aerospace Tech. (Hydrodynamics)
W. W. Fox, Dir. of Engrg.
Laboratory Staff: Chemists 2, Aero. engr. 13, test engr. 17, electronics engr. 8, math. 4, metallurgists 3, design engr. 12, thermo. engr. 4, techn. 33, other lab staff 14.
Recruiting Contact: S. H. Imdike, Adm-Engr. Personnel

-GD/Electro Dynamic
150 Avenel St.
Avenel, New Jersey 07001
Fields of Interest: Research, development, testing and analysis on rotating electrical machinery including development of special purpose machines and machines for unusual environmental conditions such as flooded motors or very high speed motors; fans and blowers, including pumping of gasses with vaneaxial type fans; noise reduction, improvement of design and performance; lift fans for GEM machines; dynapak HERF machines and actuators including hot and cold forging of metal parts by HERF methods; basic forging relationships as well as application to specific parts; high energy release actuators for siesmic exploration in the range of 10^4 to 10^6 ft. lbs.
Major Activity of Parent Company: Mfg. - of Lab Div: Mfg. 85%, R&D 10%, T&E 5%.
R&D performed for: Owners 60%, U.S. Govt. 20%, Others 20%.
Facilities: Digital computer plant including test department.
Research Personnel:
W. A. LaPierre, Mgr. R&D
H. Shapiro, V.P. Engrg.
Laboratory Staff: Elec. engr. 1, mech. engr. 1, other professionals 1, techn. 6,
Recruiting Contact: W. D. McCarthy, Mgr. Industrial Relations

-GD/Electronics
1400 N. Goodman St.
Rochester, New York 14601

To facilitate the analysis and processing of the Industrial Research Laboratories directory, we adopted the digitization methodology outlined by Png (2019) through the following steps:

1. Digitization — This process involves transforming the physical books into a machine-readable format by creating digital text files from the pages. The resulting digital files enable computers to read, search, and analyze the data for further processing, making it easier to extract relevant information and conduct systematic analyses.

2. Subsidiary Standardization — One of the most critical challenges in working with the IRL data is the inconsistent representation of subsidiaries and divisions across different editions. The IRL directory denotes subsidiaries and parent firms using subtle indicators, such as letters, underscore and indentation, which can be overlooked or misinterpreted. To address this issue, we devoted effort to standardizing the subsidiaries and divisions for each parent company and grouping them accordingly.

By reviewing the indicators in the IRL directory and harmonizing entries across multiple editions, we developed a standardized classification system that clearly identifies relationships between parent firms and their subsidiaries. This process allows us to differentiate among research facilities operating under the same corporate umbrella, ensuring that each lab is accurately tied to its corporate parent.

Such standardization is vital for two reasons. First, it enables analysis at multiple levels of aggregation — ranging from individual research labs to entire corporate entities — thereby providing a nuanced view of industrial research activities both within and across firms. Second, distinguishing between subsidiaries and parent companies sheds light on how research responsibilities are distributed within a corporation. By systematically capturing these hierarchical relationships, our dataset permits an in-depth examination of whether R&D is managed centrally or delegated to specific business units, as well as how organizational structure influences innovation outcomes. This level of detail is essential for understanding the strategic allocation of research resources and the evolution of industrial research in the United States.

3. Data preparation and Parsing

To ensure the accuracy and reliability of the data extracted from the IRL directory, we performed manual checks and corrections to address issues stemming from document quality, scanning artifacts, and OCR errors. This process involved reviewing digitized files, rectifying misplaced page numbers, and fixing incorrect spacing. We then standardized company names across multiple IRL editions to create a consistent panel of organizations over time. For example, we unified variations of "Du Pont de Nemours, E. I., and Co., Inc." and "E. I. Dupont Co." under the standardized name "E.I. Dupont." Next, we employed data parsing techniques to extract key variables — such as the number of scientists, research focus, and location — from the corrected IRL directory entries, converting the unstructured text into structured data suitable for analysis.

To integrate these IRL data with our comprehensive panel of firms, we used a fuzzy string-matching algorithm that paired each parent company in the IRL directory with a corresponding entry in our firm dataset. This algorithm relied on the Levenshtein distance, which quantifies similarity by measuring the minimal number of single-character edits needed to transform one string into another. By automatically identifying exact and near matches, we significantly reduced manual matching effort. Finally, we verified all matches through individual inspections, ensuring that only correct pairings remained in the final dataset. This multi-step process preserves both the breadth and integrity of the information from the IRL directory while supporting systematic analysis of corporate research activities.

A.1.2 American Men & Women of Science

The American Men and Women of Science (AMWS) is a longstanding biographical directory that profiles prominent, living scientists in the United States across multiple disciplines. Established in 1906, AMWS offers in-depth information on each scientist's educational background, research interests, professional experience, and institutional affiliations, capturing the trajectory of their career rather than solely relying on point-in-time survey data. Over its 119-year history, AMWS has documented the careers of more than 300,000 scientists, selected based on their notable contributions, with entries proposed by peers, academic leaders, and heads of research organizations.

To collect scientists' data from the historical AMWS volumes, we undertook the following steps:

Digitization. We digitized the AMWS editions at 5-year intervals between 1906 and 1998, converting the physical books into a machine-readable format. The digitized editions include 1949 (3 volumes), 1955 (2 volumes), 1961 (5 volumes), and 1986 (8 volumes), with each volume containing approximately 600 to 900 pages of information about the scientists. We then converted the scanned images into editable and searchable text documents to facilitate data extraction and analysis. In total, our dataset covers 48,040 scientists from the 1949 edition, 47,875 from the 1955 edition, 97,702 from the 1961 edition, and 113,913 from the 1986 edition.

Matching with firms. To link scientists to corporations, we searched each scientist's profile for mentions of company names, including variations such as "AT&T," "Bell Labs," and "American Tel & Tel Corp." This process enabled us to identify 12,865 scientists in the 1949 edition, 17,684 in the 1955 edition, 25,232 in the 1961 edition, and 31,238 in the 1986 edition who had worked for at least one corporation.

Parsing with GPT-4 Turbo. To efficiently process the vast amount of unstructured information in the AMWS dataset, we employed OpenAI's GPT-4 Turbo, a state-of-the-art language model, to extract and structure the relevant data. GPT-4 Turbo allowed us to analyze the unstructured text from each scientist's profile and convert it into a standardized format suitable for further analysis.

Here's an example of how we used GPT-4 Turbo's API to process a scientist's information:

FILLMAN, DR. LOUISE A(NNA). 3202 Princeton Ave, Midland, Texas. GEOLOGY. Gardner, Ill. Sept. 3, 98. B.A. Iowa, 20, M.S. 21, Ph.D.(geol), 24; summers, Clark, 28, 34. Washington (Seattle), 36. Asst. geol. & geog, Iowa, 21-24; prof. geol, Simpson Col, 25-41; asst. geologist, U.S. Geol. Surv, Mo, 42-43; stratigrapher, Shell Oil Co. Inc, 43-58; CONSULT, 58- Chmn. Cmt. for pre-Pennsylvanian stratigraphic terms of West Texas and Eastern N. Mex. Fel. Geol. Soc: Asn. Petrol. Geol. Stratigraphy: subsurface stratigraphy for Permian basin; Cenozoic deposits of the Black Hills of South Dakota.

FILLEY, DR. JOHN P(ATON), Dept. of Mental Health, University of North Carolina, Chapel Hill, N.C. PSYCHIATRY. New Haven, Conn, Sept. 25, 22; m. 48; c. 4. B.S, Yale, 48, M.D, 52. Intern pediat, hosps, Minnesota, 52-53; res. psychiat, N.C. Mem. Hosp, 53-55, fel. child psychiat, 55-57, instr, 57-58; asst. prof. MENTAL HEALTH, SCH. PUB. HEALTH, NORTH CAROLINA, 58-60, ASSOC. PROF, 60- U.S.N.R, 43-46. Am. Med. Asn; Psychiat. Asn; Pub. Health Asn. Personality development; preventive psychiatry; child psychiatry.

FILLINGER, PROF. HARRIETT H(ULDAH), R.D. 1, Hollins, Va. CHEMISTRY. Bristol, Va, June 27, 90. S.B, Chicago, 20, M.S, 21, summers, 24, 35, 41,46. Head dept. sci, Sullins Jr. Col, 11-15, 16-19; Miss. Synodical Jr. Col, 15-16; prof. CHEM. & head dept, HOLLINS COL, 21-56, EMER. PROF, 56- Lectr, Clinch Valley Col, Virginia, 56- A.A; Chem. Soc. Reactions in liquid ammonia and silica gels; visual education; chromatographic adsorption analysis-in organic-metal ions; preparation of ammonia derivatives of sulfuric acid and electrolysis in ammonia systems; inorganic chromatography; systematic qualitative analysis for metal ions.

FILLION, STANLEY HERBERT, 28 Tudor Lane, Scarsdale, N.Y. ENGINEERING. Springfield, Mass, Oct. 27, 07; m. 40; c. 3. B.S, Worcester Polytech, 30; M.S, Michigan, 34. Instr. civil eng. Worcester Polytech, 30-37, asst. prof. 37-43; asst. chief engr, WAUGH EQUIPMENT CO, NEW YORK, 44-57, CHIEF ENGR, 57- Soc. Civil Eng; Soc. Exp. Stress Anal. Measurement of impact forces in railway car structures; design of car structures and cushioning therefore; mathematical studies of impacts.

FILLIOS, DR. LOUIS CHARLES, 19 Eliot Rd, Lexington 73, Mass. BIOCHEMISTRY. Boston, Mass, July 1, 23; m. 47; c. 3. A.B, Harvard, 48, M.S, 53, U.S. Pub. Health Serv. fel, 54-56, Sc.D, 56. Asst. NUTRIT, SCH. PUB. HEALTH, HARVARD, 53-54, res. fel, 56, res. assoc, 57-58, ASSOC, 58- U.S.A.A.F, 43-45, 1st Lt. A.A; N.Y. Acad. Experimental atherosclerosis; lipid metabolism; purine and pyrimidine metabolism; endocrinology.

FILLIS, DR. BENJAMIN EARLE, Dept. of Urology, Stritch School of Medicine, Loyola University, Chicago, Ill. UROLOGY. Bancroft, Iowa, Dec. 4, 89; m. 15; c. 3. Iowa State Col, 07-09; M.D, Northwestern, 14. ASSOC. PROF. UROL, STRITCH SCH. MED, LOYOLA (ILL), 37- Diplomate, Am. Bd. Urol. Med.C, 18-19, 1st Lt. Urol. Asn; fel. Col. Surg. General urology.

FILLMAN, DR. LOUISE A(NNA), 3202 Princeton Ave, Midland, Texas. GEOLOGY. Gardner, Ill, Sept. 3, 98. B.A, Iowa, 20, M.S, 21, Ph.D.(geol), 24; summers, Clark, 28, 34, Washington (Seattle), 36. Asst. geol. & geog, Iowa, 21-24; prof. geol, Simpson Col, 25-41; asst. geologist, U.S. Geol. Surv, Mo, 42-43; stratigrapher, Shell Oil Co, Inc, 43-58; CONSULT, 58- Chmn. Cmt. for pre-Pennsylvanian stratigraphic terms of West Texas and Eastern N. Mex. Fel. Geol. Soc; Asn. Petrol. Geol. Stratigraphy; subsurface stratigraphy for Permian basin; Cenozoic deposits of the Black Hills of South Dakota.

FILLMORE, DR. FRANKLIN L(aBLANCHE), North American Aviation, Inc, P.O. Box 309, Canoga Park, Calif. PHYSICS, MATHEMATICS. Hoquiam, Wash, June 23, 16. A.B, California, 38, M.A, 39, Ph.D.(physics), 50; Wisconsin, 39-40. Asst. math, Wisconsin, 39-40; teacher, high sch, Calif, 40-41; asst. physics, California, 41, asst. & electronics coordinator, radiation lab, 41-45, physicist, 45-50, instr. elec. eng. univ. exten, 45-50; RES. ENGR, N. AM. AVIATION, INC, 50- With Office Sci. Res. & Develop, 44. A.A; Phys. Soc; Nuclear Soc. Electrostatic generators; nuclear physics; reactor physics.

FILLMORE, RICHARD H(AROLD), Sarah P. Duke Gardens, Duke University, Durham, N.C. HORTICULTURE. Springhill, N.S, Can, March 19, 12; m. 42; c. 1. B.A, Acadia, 45; fel. Massachusetts, 45-56; M.S, Cornell, 47. Curator of grounds, Acadia, 39-45; propagator, Arnold Arboretum, Harvard, 47-52; mgr. propagation dept, Shenandoah Nurseries, Iowa, 52-55; asst. prof. hort, Ohio State, 56; HORTICULTURIST IN CHARGE SARAH P. DUKE GARDENS, DUKE, 56- Soc. Hort. Sci; Soc. Plant Physiol; Int. Dendrol. Union. Plant propagation; water relations of plants; plant growth substances; plant breeding; ornamental horticulture.

FILLNOW, RONALD H(ENRY), Bettis Atomic Power Lab, Westinghouse Electric Corp, Box 1468, Pittsburgh 30, Pa. METALLURGY. Clintonville, Wis, Nov. 29, 21; m. 46; c. 5. B.S, Wisconsin, 49, M.S, 50; Pittsburgh. From assoc. engr. to sr. engr, ATOMIC POWER DIV, WESTINGHOUSE ELEC. CORP, 50-54, supvry engr, 55-57, MGR. IRRADIATION EFFECTS, 57- U.S.A.A.F, 42-45. Soc. Metals; Solid state studies of nuclear induced property changes in reactor fuels; structural and control materials.

FILMER, PROF. ROBERT S(TERLING), New Jersey State Agricultural Experiment Station, New Brunswick, N.J. APICULTURE. Stratford, Conn, May, 23, 02; m. 29. B.S, Conn. Agr. Col, 26; M.S, Rutgers, 30. Jr. entomologist, bur. entom, U.S. Dept. Agr, Wash, 27-29; asst, RUTGERS, 29-30, instr. ENTOMOL, 30-36, ASST. PROF, 36-, assoc. entomologist in charge apicult. res, N.J. STATE AGR. EXP. STA, 30-44, assoc. res. specialist, 44-52, RES. SPECIALIST, 52- Entom. Soc. Insects affecting field crops; ecology; breeding bees for resistance to disease; arsenic substitutes.

FILSON, DR. MALCOLM HAROLD, Central Michigan University, Mt. Pleasant, Mich. CHEMISTRY. Chattanooga, Tenn, Oct. 19, 07; m. 32; wid; c. 3. B.S, Kentucky, 29, M.S, 31; fel, Michigan, 31-34, Ph.D.(anal. chem), 36; Mich. State, 52, 53. Asst. anal. chem, Kentucky, 29-31, asst. res. chemist, Exp. Sta, 29; instr. inorg. chem, Ohio Northern, 34-35; prof. chem. & head dept. chem. & physics, Miss. Woman's Col, 35-36; PROF. CHEM. & PHYSICS, CENTRAL MICHIGAN, 36-, HEAD DEPT, 53- Summer fel, Nat. Sci. Found, Wisconsin, 59. Consult, U.S.A. A.A; Chem. Soc; Research Soc. Analytical chemistry; chromyl chloride; chromium plating baths; chromyl fluoride; periodates; perchloric acid in analytical chemistry; solubilities and viscosity as applied to silver, lead and mercury compounds.

FILTEAU, PROF. GABRIEL, Faculty of Science, Laval University, Quebec, Que, Can. ZOOLOGY. Quebec, Que, Oct. 16, 18; m. 46; c. 2. B.A, Laval, 39, B.Sc.A, 44, Ph.D.(biol), 51; Queen's (Can), 40-42. Asst. biologist, biol. sta, St. Laurent, Que, 44-46; lectr. ZOOL, LAVAL, 46-50, asst. prof, 50-55, PROF, 55- Asst. biologist, biol. sta, St. Laurent, Que, 46-50. Fel, Nuffield Found, England, 56. R.C.A.F, 58-, Wing Comdr. French-Can. Asn. Adv. Sci; Freshwater Biol. Asn, England. Freshwater biology, especially plankton; invertebrate zoology.

FILZ, FRANK R(ALPH), 105 Electric Ave, Fitchburg, Mass. ANALYTICAL CHEMISTRY. Fitchburg, Mass, Oct. 25, 06; m. 28; c. 2. Grad, Int. Correspondence Schs, 27. Asst. chemist, Wheelwright Paper Co, 27-30, chief chemist, 30-34; WHEELWRIGHT DIV, MEAD CORP, 34-40, tech. supvr, 40-52, TECH. DIR, 52- A.A; Chem. Soc; assoc. Tech. Asn. Pulp & Paper Indust. Development of control testing methods; water purification; paper machine coating; papersizing.

FIMIAN, DR. WALTER J(OSEPH), JR, Dept. of Zoology, Boston College, Boston, Mass. MORPHOLOGY, RADIOBIOLOGY. New York, N.Y, May 22, 26; m. 50; c. 2. A.B, Vermont, 50; fel, Notre Dame, 51-52, 53-54, M.S, 52, Ph.D, 55. Instr. & res. investigator, Marquette, 54-55; asst. prof. ZOOL, BOSTON COL, 55-60, ASSOC. PROF, 60- U.S.N, 44-46. A.A; Research Soc; Soc. Zool. Tissue culture studies; amphibian regeneration; melanin syntheses; experimental morphogenesis; radiation biology.

FINA, DR. LOUIS R, Dept. of Bacteriology, Kansas State University, Manhattan, Kans. BACTERIOLOGY. Cleveland, Ohio, Dec. 13, 18; m. 46; c. 2. A.B, Illinois, 42, M.S, 48, Ph.D.(bacter, chem), 50. Asst. microbiologist, Wyoming, 50-54; ASST. PROF. BACTER. & MICROBIOLOGIST, EXP. STA, KANSAS STATE, 54- Med.C, U.S.N, 42-46. Soc. Bact. Methane fermentation; industrial problems; Rumen microbiology.

FINAMORE, DR. FRANK J(OSEPH), R.D. 2, Murphysboro, Ill. BIOCHEMISTRY. Paterson, N.J, Aug. 3, 26; m. 48; c. 3. B.A, Virginia, 49, M.S, 51; Ph.D.(physiol), Fla. State, 54. Asst. physiol, Fla. State, 51-54; instr. biochem, med. col, 54, Illinois res. assoc, Inst. Tuberc. res, 54-55; asst. prof. PHYSIOL. SOUTHERN ILLINOIS, 55-59, ASSOC. PROF, 59- Res. participant, Oak Ridge Nat. Lab, 57-59. U.S.N, 45-46. A.A. Chemical embryology; nucleic acid chemistry and physiology.

FINAN, PROF. E(DWARD) J(EROME), see 9th ed, deceased 60.

FINBERG, DR. LAURENCE, Baltimore City Hospital, Baltimore, Md. PEDIATRICS, PHYSIOLOGY. Chicago, Ill, May 20, 23; m. 45; c. 3. B.S, Chicago, 44, M.D, 46. Instr. PEDIAT, SCH. MED, HOPKINS, 51-56, ASST. PROF, 56-, PEDIATRICIAN, HOSP, 51- Asst. chief pediatrician, Baltimore City Hosp, 51- U.S. Pub. Health Serv, 47-49. A.A; Soc. Pediat. Research; Fedn. Clin. Research; Acad. Pediat. Electrolyte physiology, especially relating to disturbances of sodium and water and osmotic equilibrium· metabolic and infectious diseases.

FINBY, DR. NATHANIEL, Dept. of Radiology, Cornell University Medical College, New York 21, N.Y. RADIOLOGY. New York, N.Y, May 31, 17; m. 47. A.B, Hopkins, 38, M.D, 42. Intern, Grasslands Hosp, N.Y, 42-43; private practice, 46-52; res. RADIOL, N.Y. Hosp, 52-55; asst. prof, MED. COL, CORNELL, 56-58, ASSOC. PROF, 58- Consult, Rockefeller Inst, 56-60. U.S.A, 43-46, Capt. Radiol. Soc. N.Am. Diagnostic radiology.

FINCH, DR. C(HARLES) R(ICHARD), 205 Buchanan Dr, Midland, Mich. CHEMICAL ENGINEERING. Memphis, Tenn, Nov. 30, 28; m. 56; c. 3. B.S, Maryland, 50, Ph.D.(chem. eng), 55. Asst. chem. eng, Maryland, 51-53, instr, 53-54; PROJ. LEADER, PILOT PLANT PROCESS DEVELOP, DOW CHEM. CO, 56- U.S.A.F, 54-56, 1st Lt. Chem. Soc; Inst. Chem. Eng. Process development in field of organic chemistry.

FINCH, DR. CLEMENT ALFRED, University of Washington Medical School, Seattle, Wash. MEDICINE. Broadalbin, N.Y, July 4, 15; m. 41; c. 2. B.A, Union Col, 36; fel, Rochester, 38-39, M.D, 41. Intern MED, Peter Bent Brigham Hosp, Boston, 41-42, asst. res, 42-43, res, 44-46, jr. assoc, 46-48, assoc, 48-49; res. fel. hematol, Evans Mem. Hosp, 43-44; assoc. prof, MED. SCH, WASHINGTON (SEATTLE), 49-55, PROF, 55- Instr, Harvard Med. Sch, 46-48, assoc, 48-49. Physiol. Soc; Soc. Clin. Invest; Am. Med. Asn; Am. Physicians; Fedn. Clin. Research; Int. Soc. Hematol. Hematology and iron metabolism; biophysics; clinical pathology.

FINCH, PROF. DAN M, Division of Electrical Engineering, University of California, Berkeley 4, Calif. ELECTRICAL ENGINEERING. San Francisco, Calif, June 4, 15; m. 38; c. 2. B.S, California, 37; fel, Harvard, 37-38. Asst, Technicolor Motion Pictures, 38-40; asst. chief engr, Paramount Motion Pictures, 40-41; equipment engr, Dept. Motor Vehicles, Calif, 41-45; asst. prof. ELEC. ENG. CALIFORNIA, 45-50, assoc. prof, 50-55, PROF, 55- Civilian with Office Sci. Res. & Develop; Naval Res. Cmt, 44. Inst. Elec. Eng; Illum. Eng. Soc; Soc. Automotive Eng. Motor vehicle and aircraft visibility under normal and adverse weather; fog visibility and transmission; illumination and radiation instruments.

FINCH, DONALD I(RA), Leeds & Northrup Co, Research & Development Center, North Wales, Pa. METALLURGY. Grand Rapids, Mich, July 20, 07; m. 30. B.S, Michigan, 29, 29-30. Res. & develop. metallurgist, METALL. SECT, RES. & DEVELOP. DEPT, LEEDS & NORTHRUP CO, 30-33, 35-44, acting div. chief, 44-45, chief, 45-57, HEAD, 57- With Office Sci. Res. & Develop; Naval Ord. Lab, 44. Soc. Metals; Inst. Min, Metall. & Petrol. Eng; Soc. Test. Mat; Soc. Mech. Eng; Nat. Soc. Prfnl. Eng; London Iron & Steel Inst; British Inst. Metals. Pyrolitic breakdown of hydrocarbons and their relationship to gas carburizing; automatic measurement and control of industrial furnace atmospheres; design and development of gas generators, ammonia dissociators; and gas carburizing furnaces; diffusion in metals and metal impregnation; metallurgy of resistance alloys; metals and alloys for thermo-electric purposes and of special solders; special contact materials; diaphragms as prime elements.

FINCH, DR. G(AYLORD) K(IRKWOOD), 717 Yadkin St, Kingsport, Tenn. ORGANIC CHEMISTRY. Owosso, Mich, Nov. 16, 23; m. 45; c. 4. B.S.Ch.E, Michigan, 45, fel, 46-50, M.S, 48, Ph.D.(chem), 54. Ord. engr, U.S. Govt.

Using the prompt shown in Listing 1, we input the text of Dr. Louise Anna Fillman's profile into GPT-4 Turbo. The model then analyzed the unstructured text and generated the following structured information:

Listing 1: GPT 4 Turbo prompt for extracting scientists' professional experience information

```
### GPT 4 Turbo:
How can I help you today?

### User:
f‘‘‘Given the following information about a scientist: {entry_text}.
   Please organize the data into two tab-separated tables. Split the
   tables and the table title with table content by using double newline
   characters (\n\n). If any information is missing, use ‘N/A’ in its
   place and proceed to the next information.
        1) Professional Experience: Include the scientist’s full name,
           position, institution and year using the following headers: ‘
           Full Name’, ‘Position’, ‘Institution’, ‘Year’.

        2) Education: Include the scientist’s full name, institution,
           degree, and year using the following headers: ‘Full Name’, ‘
           Institution’, ‘Degree’, ‘Year’.

        3) Personal Information birth city, birth day, marital status and
           children (E.g. Copenhagen, Denmark, Apr 26, 22; m 50; c 4.
           represents Birth city, Birthday, martial year(m), number of
           children(c) )

For example:
        Professional Experience\n\nFull Name\tPosition\tInstitution\tYear\
           n
        Education\n\nFull Name\tInstitution\tDegree\tYear\n
        Personal Information\n\nFull Name\tBirth city\tBirthday\tmarrige
           date\tchildren\n
        ’’’

        ‘‘Do not include headers or additional explanatory text. Only the
           structured data is required.’’

### GPT4 Turbo:
 ...
 ...
 ...
```

Table A2: A sample entry of AMWS: Fillman, Dr. Louise Anna

Job Title	Insitution	Years
Asst. geol. & geog	Iowa	1921-1924
prof. geol	Simpson Col	1925-1941
asst. geologist	U.S. Geol. Surv, Mo	1942-1943
stratigrapher	Shell Oil Co. Inc	1943-1958
CONSULT		1958-

Notes: This table demonstrates how GPT-4 Turbo can analyze unstructured text about a scientist and generate structured information from it.

By applying this process to each scientist's profile, we obtained a structured dataset containing rows of professional experiences, with each row including the job title, institution, and years of employment. This structured data allowed us to identify which scientists worked for specific firms and during what time periods.

To assess the accuracy of the GPT-4 Turbo model's output, we conducted a manual evaluation on a randomly selected sample of 1,000 processed profiles for each edition of AMWS. We carefully reviewed the model's output and compared it to the original unstructured text. We found that the model achieved an impressive accuracy rate of over 95%, demonstrating its effectiveness in extracting and organizing information from the AMWS dataset.

In addition to the manual evaluation, we also implemented a human-in-the-loop approach to further ensure the quality of the extracted information. We manually reviewed and corrected any errors or inconsistencies identified during the evaluation process. This combination of automated processing using GPT-4 Turbo and human oversight resulted in an accurate structured dataset.

Our experience with using GPT-4 Turbo for extracting information from the AMWS dataset highlights the potential of large language models in cleaning and organizing unstructured data. By fine-tuning the model prompts on a specific dataset and combining its outputs with human verification, researchers can achieve high accuracy rates and efficiently process large volumes of unstructured text. This approach can be particularly valuable for historical datasets, where the information is often presented in inconsistent formats and styles.

Name Disambiguation. To ensure the accuracy of our dataset and avoid double-counting, we disambiguated scientists appearing in the 1949, 1955, 1961, and 1986 editions of the AMWS directory. Because later editions typically feature more comprehensive information on a scientist's career, we retained only the most recent entry for each individual. For instance, a researcher listed in both 1949 and 1961 would be captured once in our dataset, with the 1961 entry providing details that encompass the intervening years. By cross-referencing data across these editions and removing duplicates, we arrived at final counts of 48,040 scientists for 1949, 47,875 for 1955, 97,702 for 1961, and 113,913 for 1986. This approach preserves the most up-to-date information for each scientist — covering educational history, career trajectory, and professional affiliations — while ensuring the integrity of our longitudinal analysis of corporate R&D over time.

A.1.3 National Register of Scientific and Technical Personnel

The National Register of Scientific and Technical Personnel (NRSTP) is a comprehensive dataset compiled by the National Science Foundation (NSF) during the early Cold War. Collected in eight waves — 1954, 1958, 1960, 1962, 1964, 1966, 1968, and 1970 — the NRSTP contains detailed information on scientists and engineers, including their skills, locations, and employers. Designed to

inform government policy and strategic planning in science and technology, the data were gathered through surveys distributed to professional societies such as the American Chemical Society and the American Physical Society, whose members provided responses on their educational backgrounds, research activities, and employment histories.

To make the NRSTP data more useful for research, we focused on the period from 1964 to 1970, when the NRSTP recorded employment information for each respondent. We restricted our sample to scientists holding natural science Ph.D. degrees, ensuring a consistent level of expertise across the dataset. Drawing on the surveys' employment fields, we matched 26,245 of these PhD scientists to 855 distinct firms, generating 50,960 total scientist-firm observations. Because some individuals worked at multiple organizations during the survey period, we created separate observations for each scientist-firm pairing. This approach allows for a finer-grained analysis of employment patterns, job mobility, and the relationship between scientists' career trajectories and corporate research efforts.[27]

A.1.4 National Engineers Register

The National Engineers Register (NER) is a comprehensive database compiled by the Engineers Joint Council under contract with the National Science Foundation (NSF), forming part of the broader National Register of Scientific and Technical Personnel (NRSTP). Its chief purpose was to ensure that engineers could be readily identified and mobilized during national emergencies. To this end, the NER contains detailed records on engineers' names, locations, birth dates and places, educational backgrounds, employment histories (including job titles and previous employers), and areas of expertise. Data were collected across three waves — 1964, 1967, and 1969 — via a random sample of members from various professional engineering societies. To be included in the survey, individuals had to hold at least a bachelor's degree in engineering, be registered as a professional engineer or engineer-in-training, and maintain full membership in a recognized engineering society.

We focused our analysis on PhD scientists in the natural sciences who appear in the NER. Using the registry's employment information, we matched 3,603 of these scientists to 475 firms, generating 5,996 person-firm observations. When a scientist was employed by multiple firms, each unique employer-employee pairing was recorded as a separate observation, allowing for a more nuanced examination of employment patterns and mobility within the engineering workforce.

A.2 Measures of Scientific Outputs

A.2.1 Prestigious Science & Engineering Awards

Nobel Laureates and Nobel Prize-Winning Publications — We hand-collected data on Nobel publications. Nobel lectures are the primary resource for identifying the laureates' prize-winning publications. We identified all references cited in these lectures (Li et al., 2019) and selected the papers specifically mentioned as being integral to the award. For references not explicitly acknowledged in the lectures, we applied a set of criteria to determine whether a paper should be qualified as award-winning: (1) The Nobel laureate must be listed as an author; (2) The publication date must fall within the relevant time frame for the award; (3) The institution and co-author details must align with other public information; (4) The paper's topic must be consistent with the theme and motivation behind the Nobel Prize.

In cases where the Nobel lectures lacked explicit citations–often true for awards prior to 1940–and notably for those whose influential works in languages other than English — we implemented

[27] We are grateful to Dan Gross for sharing the NRSTP data with us. Gross converted the original survey files from microfilm into machine-readable CSV files, greatly enhancing their accessibility for research.

additional steps to authenticate our identification process. These involved an extensive search across a variety of supplementary sources such as: (1) Wikipedia pages of Nobel laureates; (2) Review articles and books that provide in-depth accounts of Nobel laureates' lifetime contributions, such as the authoritative work by Katsumori and Katsumori, 2011; (3) Personal or laboratory websites maintained by Nobel laureates themselves, which often showcase their most significant works and provide invaluable aid in identifying their award-winning papers; (4) Previously conducted research that has compiled lists of Nobel prize-winning papers, such as the study by Li et al., 2019.

We cross-referenced all identified prize-winning papers against the Dimensions database. Through this process, we compiled a total of 865 award-winning publications authored by 585 Nobel laureates, spanning the entire history of the Nobel Prize from its inception in 1901 to 2023. This collection spans across multiple scientific disciplines, including 285 papers in Physics by 209 laureates, 265 in Chemistry by 197 laureates, and 316 in Medicine by 224 laureates. Subsequently, we fuzzy matched the affiliations of Nobel laureates at the time of publication of their prize-winning papers with a list of companies. This process was conducted to establish connections between scientists and the corporations for which they have worked.

The President's National Medal of Science — The National Medal of Science (NMS) is a prestigious award conferred by the President of the United States on scientists and engineers who have made extraordinary contributions to advancing knowledge in fields such as behavioral and social sciences, biology, chemistry, engineering, mathematics, and physics. Established in 1959, the NMS is bestowed annually to honor the nation's most distinguished researchers and innovators.

To build a comprehensive dataset of NMS recipients, we scraped the "President's National Medal of Science: Recipient Search Results" page on the National Science Foundation (NSF) website.[28] From these records, we extracted information on each honoree's name, year of award, field of contribution, and affiliated institutions. We then linked these affiliations to a list of companies, enabling us to study the connection between NMS recipients and corporate organizations that may have supported or benefited from their research.

National Medal of Technology and Innovation — The National Medal of Technology and Innovation (NMTI) is the highest distinction conferred by the President of the United States on American inventors and innovators who have substantially advanced the development and commercialization of pivotal new technologies. This award honors achievements that bolster the nation's economic, environmental, and social well-being through technological innovation.

To assemble a comprehensive dataset of NMTI recipients, we scraped the "National Medal of Technology and Innovation (NMTI) Recipient" page on the United States Patent and Trademark Office (USPTO) website,[29] extracting each laureate's name, year of award, field of contribution, and institutional affiliations. We then examined the committee's summary of recipients' accomplishments, cross-referencing this information with additional sources — including biographical data, patent records, and published research — to determine the organizations where these innovators conducted their most influential work. By matching these affiliations to a list of companies, we established a link between NMTI recipients and the corporate entities that supported or benefited from their breakthroughs, thereby providing a clearer picture of the role businesses play in fostering technological advancement.

[28] https://www.nsf.gov/od/nms/results.jsp

[29] https://www.uspto.gov/learning-and-resources/ip-programs-and-awards/national-medal-technology-and-innovation/recipients/1985

A.2.2 Membership in Science & Engineering Societies

American Association for the Advancement of Science — The American Association for the Advancement of Science (AAAS) Fellows program honors distinguished scientists, engineers, and innovators whose work has significantly advanced their respective fields. Elected annually by the AAAS Council, Fellows represent a broad range of disciplines — spanning research, teaching, technology, administration in academia, industry, and government — as well as excellence in communicating and interpreting science to the public. Established in 1874, the AAAS Fellows program is regarded as a prestigious mark of scientific excellence, and notable inductees have included Thomas Edison, W.E.B. DuBois, Maria Mitchell, Steven Chu, Ellen Ochoa, and Irwin M. Jacobs.

To compile a comprehensive dataset of AAAS Fellows, we scraped the relevant data from the AAAS website,[30] extracting each Fellow's name, year of election, geographic location, disciplinary section, institutional affiliation, and professional status. We then linked these Fellows' primary affiliations to a list of companies, enabling us to explore connections between AAAS-recognized scientists and the corporate sector.

National Academy of Science — The National Academy of Sciences (NAS) is a distinguished organization that elects prominent scientists in recognition of their outstanding contributions to original research. Membership in the NAS is widely regarded as a pinnacle of scientific achievement. Currently, the NAS includes approximately 2,400 active members and 500 international members, nearly 190 of whom have been awarded Nobel Prizes.

Membership is conferred solely through election rather than application. While many names are informally suggested, only existing NAS members can submit formal nominations. Each nomination undergoes rigorous evaluation, culminating in a final ballot at the Academy's annual meeting every April. Up to 120 new members may be elected each year; nominees must be U.S. citizens, whereas a maximum of 30 non-citizens per year are chosen as international members.

NAS members belong to specific sections — defined by scientific discipline — organized into six overarching classes: Physical and Mathematical Sciences, Biological Sciences, Engineering and Applied Sciences, Biomedical Sciences, Behavioral and Social Sciences, and Applied Biological, Agricultural, and Environmental Sciences.

To assemble a comprehensive dataset of NAS members, we scraped the relevant data from the NAS website.[31] Information collected includes each member's name, election year, primary institution, membership type, and disciplinary section(s), along with geographic details. We then matched these members' primary affiliations to a list of companies, enabling us to assess the connections between NAS-recognized scientific expertise and corporate organizations.

National Academy of Engineering — The National Academy of Engineering (NAE) is a nonprofit, non-governmental organization within the broader National Academies of Sciences, Engineering, and Medicine, which also comprise the National Academy of Sciences (NAS), the National Academy of Medicine, and the National Research Council. The NAE recognizes and elects distinguished engineers who have made notable contributions to research, practice, or education, as well as those who have pioneered new or emerging fields of technology or significantly advanced more traditional areas of engineering.

Membership in the NAE is conferred solely through peer election and is among the highest honors in the engineering profession. More than 5,000 individuals currently serve as NAE members or international members, representing top figures in industry, academia, and government. To construct a comprehensive dataset of NAE members, we scraped data from the NAE's official

[30] https://www.aaas.org/fellows/historic

[31] https://nasonline.org/member-directory/

website.[32] This information included each member's name, year of election, institutional affiliation, membership type, and engineering section. The NAE divides its membership into twelve sections, including Aerospace, Bioengineering, Civil & Environmental, Computer Science & Engineering, and others. We then matched these members' institutional affiliations to a list of companies, providing insight into the involvement of leading engineering experts in corporate innovation.

National Academy of Medicine — Formerly known as the Institute of Medicine (IOM), the National Academy of Medicine (NAM) is one of the three academies forming the National Academies of Sciences, Engineering, and Medicine in the United States. Founded in 1970, the NAM operates under the 1863 congressional charter of the National Academy of Sciences. It is a private, nonprofit institution that provides objective advice on matters of science, technology, and health, acting independently from government.

Election to the NAM is recognized as one of the highest professional honors in health and medicine, acknowledging individuals who have achieved outstanding professional merit and demonstrated a sustained commitment to service. To develop a comprehensive dataset of NAM members, we scraped the official NAM website.[33] This process involved collecting information on each member's name, institutional affiliation, membership type, and section. We then mapped these affiliations to a list of companies, enabling us to explore ties between high-level medical expertise and corporate organizations.

American Academy of Arts and Science — The American Academy of Arts and Sciences (AAA&S) — founded in 1780 during the American Revolution by figures such as John Adams, John Hancock, James Bowdoin, and Andrew Oliver — is among the oldest and most prestigious learned societies in the United States. Election to the AAA&S is a sought-after distinction for scholars, recognizing exceptional achievement in the arts and sciences.

We constructed a comprehensive list of AAA&S members by scraping the Academy's online directory.[34] From these entries, we collected each member's name, institutional affiliation, area of expertise, and year of election. Although the Academy's membership encompasses five broad areas — Mathematical and Physical Sciences, Biological Sciences, Social and Behavioral Sciences, Humanities and Arts, and leadership, Policy, and Communications — our analysis focused on the Mathematical and Physical Sciences and Biological Sciences categories. We subsequently matched these members to a list of companies, linking corporate affiliations with AAA&S fellows who have demonstrated excellence in scientific research and leadership.

American Philosophical Society — Founded in 1743 by Benjamin Franklin, the American Philosophical Society (APS) is the oldest learned society in the United States. Dedicated to promoting "useful knowledge," the APS honors leading scholars, scientists, and professionals through elected membership. It supports research and discovery with grants, fellowships, lectures, and prize programs, and maintains a valuable research library of manuscripts and historical collections. As of April 2024, the APS has around 1,000 active elected members, approximately 840 of whom are Resident Members (based in the United States) and 160 are International Members from over two dozen countries; since 1900, more than 260 APS members have received the Nobel Prize.

To create a dataset of APS members, we scraped information from the APS Member Directory,[35] which offers a list of individuals elected from the society's founding through the present day. This directory includes essential details such as each member's name, election year, and institutional affiliation or field of study. The APS organizes its membership into five classes — Mathematical and Physical Sciences, Biological Sciences, Social Sciences, Humanities and the Arts, and Pro-

[32] https://www.nae.edu/20412/MemberDirectory
[33] https://nam.edu/about-the-nam/
[34] https://www.amacad.org/directory
[35] https://www.amphilsoc.org/elected-members

fessions and leaders in Public & Private Affairs. Similar to our approach for the AAA&S, we focused on members from the Mathematical and Physical Sciences and Biological Sciences classes, then matched their institutional affiliations to a list of companies to map the connection between APS-recognized expertise and corporate innovation.

Appendix B Ensemble Learning

B.1 Ensemble Model Selection

We employ a two-stage selection procedure to identify which of the sixteen models to include in our ensemble. Table B.1 adds details on how the input variables are calculated. In the first stage, we rank each model's classification outcomes against a ground truth set of firms independently identified as either Science Leaders or development-only. For Science Leaders, we select twenty companies that exhibit non-zero values across all of our scientific indicators. These firms include AT&T, GE, IBM, RCA, DuPont, Xerox, Monsanto, Dow Chemical, Merck, Standard Oil, Hercules, Lockheed, TRW, Sperry, United Aircraft, Varian Associates, Texas Instruments, American Science and Engineering, Bolt Beranek & Newman (BBN Technologies), and Intel.[36] We define 871 "D" firms (Followers) that score zero on all of these metrics, indicating no engagement with scientific research.

[36]They employ Nobel laureates or National Medal of Science recipients, have corporate scientists in prestigious scientific societies, cite their own science in patents, reference Nobel Prize–winning research in publications, employ NRSTP-surveyed research scientists, maintain PhD-staffed laboratories, and have researchers listed in *American Men and Women of Science*.

Table B3: Ensemble Model Variable Definitions

Variable	Description
A. Inputs of Corporate Science	
Number of National Register of Scientific and Technical Personnel (NRSTP) and National Engineers Register (NER) scientists employed	These variables reflect the broader human capital input to a firm's scientific activities. Firms with a higher number of NRSTP and NER scientists are more likely to be actively engaged in scientific research across various disciplines.
Presence and size of industrial research laboratories	The existence and scale of dedicated research facilities within a firm serve as indicators of its commitment to scientific exploration and knowledge generation. Firms with larger and more established research laboratories are more likely to be at the forefront of scientific advancements.
Number of American Men and Women of Science (AMWS) scientists employed	This variable represents the human capital input of the most prominent scientists within a firm. AMWS scientists are recognized for their significant contributions to scientific knowledge, and their presence in a firm indicates a strong commitment to advanced scientific research.
Corporate scientists' memberships, fellowships, and awards	This variable signifies the presence of highly accomplished scientists within a firm. Scientists who hold prestigious memberships, fellowships, and awards, such as membership in the National Academy of Sciences or the National Academy of Engineering, are recognized for their exceptional contributions to their respective fields. The presence of such scientists in a firm indicates its ability to attract and retain top scientific talent and its commitment to cutting-edge research.
A. Outputs of Corporate Science	
Scientific publications	The number and impact of scientific publications authored by a firm's employees provide a direct output measure of its scientific activities. Firms that actively produce new scientific knowledge are expected to have a higher volume and quality of scientific publications in peer-reviewed journals.
Publications citing Nobel Prize-winning research	This variable highlights a firm's engagement with groundbreaking scientific discoveries. By examining the extent to which a firm's publications cite Nobel Prize-winning research, we can gauge its involvement in advancing the scientific frontier and building upon the most influential scientific contributions.
Patents	Patents represent another key output of a firm's R&D efforts, capturing its ability to translate scientific knowledge into novel and commercially valuable innovations. Firms with a strong science orientation are expected to have a higher volume and quality of patents, reflecting their capacity to bridge the gap between scientific research and practical applications.

Because our ground truth includes only two categories — Leaders and Followers — we measure each model's performance using the rows "Leaders Correctly Identified" and "Followers Correctly Identified" in Table B4. "Leaders Correctly Identified" captures the fraction of actual Leaders that remain classified as Leaders rather than being mislabeled as Absorbers or Followers, while "Followers Correctly Identified" reflects how many true Followers avoid being upgraded to Leaders or Absorbers. We also report a "Weighted Accuracy" that accounts for the imbalance of 20 Leaders versus 871 Followers. In addition, the table indicates whether "Leaders" or "Followers" are predicted outside their ground-truth class (for instance, "Leaders Predicted as Absorbers or Followers"). Models that satisfy minimum thresholds on these metrics progress to the second stage of ensemble voting, where they yield a three-class designation (Leader, Absorber, Follower) for the full sample.

BIRCH, RBM, and SOM fail to meet these thresholds by mislabeling nearly all "ground-truth" Leaders as Absorbers or Followers, while Spectral erroneously upgrades hundreds of true Followers to Leader status. The Attention model also struggles to keep "ground-truth" Leaders separate from the other categories. In addition, BIRCH and Attention push more than 98 percent of firms into the Follower group, effectively defaulting to a near–one-class solution. Consequently, these methods are excluded from further consideration.

After removing these underperformers, eleven models remain: Agglomerative Clustering, Autoencoder, CNN, Deep Autoencoder, Deep Belief Network, Deep Boltzmann Machine, Gaussian Mixture Model, Hierarchical Clustering, K-means, ResNet, and Variational Autoencoder. These models demonstrated weighted accuracy scores ranging from 57.5 to 99.4 percent. Most also produce a nontrivial share of Absorber assignments, avoiding the near-one-class outcomes seen in BIRCH and Attention. These patterns suggest that the retained methods can detect a reasonable fraction of high-intensity science firms without grossly mislabeling the large Follower group, providing a sufficiently balanced basis for the subsequent ensemble procedure.

We combine the validated models' predictions via majority voting (Zhou, 2012), assigning each firm to the category — Science Leader, Science Absorber, or Follower — that receives the most votes across all 11 classifiers (Kuncheva, 2014; Lam & Suen, 1997; Ruta & Gabrys, 2005). To resolve ties systematically, we adopt a hierarchical decision rule. When Leader and Absorber votes are tied for the majority, the firm is assigned to the Leader category, reflecting the higher intensity of scientific engagement.[37] When Absorber and Follower votes are tied, the firm is labeled as an Absorber, indicating a moderate level of scientific involvement.[38] Finally, when Leader and Follower votes are tied, we also classify the firm as an Absorber.[39]

This approach resulted in 38 Science Leaders (1.3%), 176 Science Absorbers (6.1%), and 2,690 Followers (92.6%). Since our ground truth dataset contains only two labeled categories — Science Leaders and Followers — we validate the model by checking its accuracy against these known labels. Among the 871 ground-truth Followers, all were correctly identified, demonstrating the model's robustness in recognizing firms with no measurable scientific engagement. Similarly, 14 out of 20 ground-truth Leaders were accurately classified, confirming the model's ability to detect firms with high scientific engagement.

While we lack explicit ground-truth labels for the Science Absorber category, the ensemble voting process provides a structured means of identifying firms that exhibit intermediate levels of

[37] This tie occurred in six cases, including General Dynamics Corporation, which received five votes for Leader, five for Absorber, and one for Follower.

[38] This rule affected nineteen firms, including Air Products and Chemicals, Inc., which received four votes each for Absorber and Follower, along with three for Leader.

[39] In five cases, including Computer Sciences Corporation, four votes each went to Leaders and Followers, while three went to Absorbers.

Table B4: Ensemble Model Selection

	AggClust	Attention	Autoencoder	BIRCH	CNN	DAE	DBN	Deep Boltzmann	GMM	Hierarchical	KMeans	ResNet	RBM	SOM	Spectral	VAE
Panel A: Full Sample Classification Distribution																
Leaders	32	2	77	1	6	31	564	559	3	32	9	101	2	9	2832	621
Absorbers	151	7	514	31	166	247	38	822	272	151	186	121	682	177	16	511
Followers	2721	2895	2313	2872	2732	2626	2302	1523	2629	2721	2709	2682	2220	2718	56	1772
Panel B: Validation Performance																
Leaders Correctly Identified	0.65	0.10	0.85	0.05	0.30	0.70	1.00	1.00	0.15	0.65	0.30	0.45	0.00	0.00	1.00	1.00
Followers Correctly Identified	1.00	1.00	0.99	1.00	1.00	1.00	0.99	0.59	1.00	1.00	1.00	1.00	0.97	1.00	0.05	0.98
Weighted Accuracy[a]	82.5%	55.0%	91.9%	52.5%	64.9%	85.0%	99.4%	79.3%	57.5%	82.5%	65.0%	72.5%	48.4%	49.9%	52.4%	99.0%
Panel C: Confusion Matrix (Ground-Truth: Leader vs. Follower)																
Leaders Correctly Identified	13	2	17	1	6	14	20	20	3	13	6	9	0	0	20	20
Followers Correctly Identified	871	871	861	871	870	871	861	511	871	871	871	871	843	870	41	853
Leaders Predicted as Absorbers or Followers	7	18	3	19	14	6	0	0	17	7	14	11	20	20	0	0
Followers Predicted as Leaders or Absorbers	0	0	10	0	1	0	10	360	0	0	0	0	28	1	830	18
Panel D: Cross-Category Misclassification (Ground-Truth vs. Predicted)																
Leader → Absorber	6	4	3	12	12	6	0	0	16	6	13	3	20	17	0	0
Leader → Follower	1	14	0	7	2	0	0	0	1	1	1	8	0	3	0	0
Follower → Leader	0	0	0	0	0	0	9	9	0	0	0	0	1	1	816	16
Follower → Absorber	0	0	10	0	1	0	1	351	0	0	0	0	27	0	14	2

Notes: This table presents the classification performance and model selection results for our 16 deep learning methods across four panels. Panel A shows each model's classification of the full sample (n=2,904) into three predicted categories: Leader, Absorber, and Follower. Panel B reports the validation metrics for distinguishing "ground-truth" Leaders (20 firms) from Followers (871 firms), namely Leaders correctly identified, Followers correctly identified, and a weighted accuracy measure. "Leaders Correctly Identified" corresponds to the fraction of "ground-truth" Leaders not mislabeled as Absorbers or Followers, while "Followers Correctly Identified" is the fraction of "ground-truth" Followers not elevated to Leader or Absorber status. We apply a class-weighted accuracy to avoid trivial solutions in which nearly all firms are lumped into the larger Follower group.[a] Panel C provides a breakdown of how many "ground-truth" Leaders and Followers each method correctly predicts (rows labeled "Leaders Correctly Identified" and "Followers Correctly Identified"), as well as how many are mistakenly placed in other categories (rows labeled "Leaders Predicted as Absorbers or Followers" and "Followers Predicted as Leaders or Absorbers"). Panel D further details the specific cross-category misclassifications, showing how often a Leader is predicted as Absorber or Follower ("Leader → Absorber," "Leader → Follower") and how often a Follower is misclassified as Leader or Absorber ("Follower → Leader," "Follower → Absorber"). The highlighted methods are those selected for our ensemble based on their superior performance in separating Leaders from Followers according to these validation metrics, while also avoiding near-one-class classifications. The highlighted models are those selected for our ensemble based on their superior performance in classifying benchmark cases, achieving either "Leaders Correctly Identified" or "Followers Correctly Identified" above 0.95. These eleven models include: Agglomerative Clustering, Autoencoder, Convolutional Neural Networks (CNN), Deep Autoencoder (DAE), Deep Belief Network (DBN), Deep Boltzmann Machine, Gaussian Mixture Model (GMM), Hierarchical Clustering, K-means, Residual Network (ResNet), and Variational Autoencoder (VAE). Several models, such as Attention Mechanism, BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies), Restricted Boltzmann Machine (RBM), Self-organizing Map (SOM), and Spectral Clustering, were excluded due to poor performance in "Leaders Correctly Identified" or "Followers Correctly Identified" metrics, with each showing severe deficiencies in at least one of these measures.

[a]Rather than dividing by the total sample size, we weight Leader and Follower performance equally, giving each class 50% of the accuracy score. This approach ensures that identifying Leaders well is not overshadowed by the far more numerous Followers.

scientific activity. These firms do not fit neatly into either the Leader or Follower category but show moderate engagement with scientific research. To ensure the validity of this classification, we conduct additional checks on Science Absorbers, examining indicators such as patent citations to scientific research, corporate research affiliations, and the presence of PhD-trained scientists. These supplementary analyses confirm that Science Absorbers exhibit characteristics consistent with firms positioned between Science Leaders and Followers.

B.2 Ensemble Model Feature Importance

Table B5 presents the SHAP (SHapley Additive exPlanations) values[40] indicating feature importance in classifying firms across three categories (Science Leaders, Absorbers, and Followers) over different time periods. The top panel shows that industrial laboratory scientists (0.047-0.067), patents (0.047-0.054), and American Men & Women of Science listings (0.037-0.058) consistently emerge as the most influential features for classification. Publications also maintain stable importance (0.032-0.040) across periods, while NRSTP research scientist counts show moderate influence (0.017-0.019). Less influential features include citations of Nobel Prize-winning research (0.003-0.004), patent self-citations (0.001-0.002), and scientist society membership (0.002-0.004).

Table B5: Feature Importance

Feature	**Ensemble** (1945-1965)	**Ensemble** (1965-1980)	**Ensemble** (1945-1980)
Publication	0.040	0.039	0.032
Patent	0.054	0.047	0.050
Industrial Laboratory Scientist	0.067	0.051	0.047
NRSTP Research Scientist	0.019	0.017	0.018
American Men & Women Science	0.058	0.037	0.039
Publication Citing Nobel Prize Winning Pub	0.003	0.004	0.003
Patent Citing Same Firm's Pub	0.001	0.002	0.001
Award	0.000	0.000	0.001
Scientist Membership	0.003	0.002	0.004
Leader	11	30	38
Absorber	135	228	176
Follower	1,459	1,347	2,690
% of Leader	0.69%	1.87%	1.38%
% of Absorber	8.41%	14.20%	5.99%
% of Follower	90.90%	83.93%	92.63%

Notes: This analysis examines the classification of firms as Science Leaders, Absorbers, and Followers across different time periods from 1945-1980. The table presents feature importance scores using SHAP (SHapley Additive exPlanations) values across three time periods: the full sample (1945-1980), early period (1945-1965), and later period (1965-1980). The SHAP values measure each feature's contribution to the model's predictions, with higher values indicating greater influence. The bottom panel shows the resulting classification counts and proportions, with Science Leaders making up between 0.69% and 1.87% of firms depending on the model and time period. All SHAP values are rounded to three decimal places.

While certain variables generate relatively low SHAP (Shapley additive explanation) scores,

[40]SHAP values measure the contribution of each feature to the model's predictions. They quantify how much each feature increases or decreases the likelihood of a given classification, providing an interpretable metric of feature importance in machine learning models.

they remain important for three reasons. First, these features may capture distinct aspects of firm behavior that are not fully represented by more influential variables. For instance, citations to Nobel Prize–winning research yield small individual SHAP scores (0.003–0.004), yet they can help differentiate high-publishing firms that genuinely produce new science from those that merely absorb external discoveries. In this sense, low SHAP scores do not imply irrelevance but rather reflect context-dependent contributions. Second, even modest indicators can enhance model performance through interactions with other features. For example, patent citations to the same firm's publications (0.001–0.002) assume greater importance when combined with high patent counts (0.047–0.054), thereby identifying firms that build upon their own research. Lastly, including a diverse array of variables ensures the model's robustness across varied contexts and time periods. This is crucial given the significant temporal fluctuations in our sample, where the proportion of Science Leaders ranges from 0.69 percent to 1.87 percent. Relying on a narrow set of indicators might yield less reliable classifications over different eras.

B.3 Ensemble Model Cluster Comparison

Table B6 presents the mean values of various scientific metrics across the three firm classifications: Science Leaders, Science Absorbers, and Followers. Science Leaders exhibit the highest levels of scientific engagement, as reflected in their significantly greater output of publications (3.888) and patents (4.415) compared to Science Absorbers (1.891 and 3.331, respectively) and Followers (0.151 and 0.805). This pattern extends to firms' research infrastructure, with Science Leaders employing the largest number of industrial laboratory scientists (7.403) and NRSTP research scientists (4.929), followed by Science Absorbers (6.106 and 3.126), and Followers (1.673 and 0.185), highlighting their superior investment in in-house research capabilities.

Table B6: Firm Characteristics Across Ensemble Groups

	Ensemble Classification		
	Follower	Absorber	Leader
Publications	0.151	1.891	3.888
Patents	0.805	3.331	4.415
Industrial Laboratory Scientist	1.673	6.106	7.403
NRSTP Research Scientist	0.185	3.126	4.929
American Men & Women Science	0.746	5.225	6.731
Publication Citing Nobel Prize Winning Pub	0.002	0.079	0.927
Patent Citing Same Firm's pub	0.001	0.066	0.613
Award	0.000	0.003	0.033
Scientist Membership	0.003	0.061	0.382
Number of Firms	2690	176	38

Notes: This table presents the mean values of scientific metrics for predicted Science Leaders, Science Absorbers, and Followers. All variables are expressed in natural logarithms.

Further differentiating these groups is their connection to the broader scientific community. Science Leaders are more frequently listed in American Men & Women of Science directories (6.731) and cite Nobel Prize-winning research at a substantially higher rate (0.927) than Science Absorbers (5.225 and 0.079) and Followers (0.746 and 0.002). Similarly, Science Leaders demonstrate a stronger integration of scientific knowledge into their innovation processes, as indicated by their

patents' self-citation rate (0.613), which far exceeds that of Science Absorbers (0.066) and Followers (0.001).

Moreover, Science Leaders are more likely to receive scientific awards (0.033) and maintain membership in scientific societies (0.382), reinforcing their active participation in the scientific ecosystem. Despite these clear differences, Science Leaders represent a small fraction of the total sample, accounting for only 38 firms, whereas Science Absorbers and Followers comprise 176 and 2,690 firms, respectively. These findings underscore the sharp stratification of firms in terms of their scientific engagement, with Science Leaders emerging as the most research-intensive and scientifically integrated firms, Science Absorbers occupying an intermediary position, and Followers demonstrating minimal involvement in formal scientific activities.

References for Appendices

Katsumori, M., & Katsumori, M. (2011). Bohr and the development of quantum theory: A brief review. *Niels Bohr's Complementarity: Its Structure, History, and Intersections with Hermeneutics and Deconstruction*, 1–10.

Kuncheva, L. I. (2014). *Combining pattern classifiers: Methods and algorithms*. John Wiley & Sons.

Lam, L., & Suen, S. (1997). Application of majority voting to pattern recognition: An analysis of its behavior and performance. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *27*(5), 553–568.

Li, J., Yin, Y., Fortunato, S., & Wang, D. (2019). A dataset of publication records for nobel laureates. *Scientific data*, *6*(1), 33.

Png, I. P. (2019). Us r&d, 1975–1998: A new dataset. *Strategic Management Journal*, *40*(5), 715–735.

Ruta, D., & Gabrys, B. (2005). Classifier selection for majority voting. *Information fusion*, *6*(1), 63–81.

Zhou, Z.-H. (2012). *Ensemble methods: Foundations and algorithms*. CRC press.