Beyond Innovation: The Composition and Dynamics of Technology-Enabled Entrepreneurship*

Innessa Colaiacovo†
University of Oregon
Lundquist College of Business

Daniel P. Gross‡
Duke University
Fuqua School of Busness
and NBER

Jorge A. Guzman §
Columbia Business School
and NBER

November 2024

Abstract

Using data on nearly all U.S. business registrations since the turn of the twentieth century and 860 technologies which emerged over this period, we study entrepreneurial activity around new innovations. We find a large and varied entrepreneurial economy rooted in innovation, where the vast majority of startups are not high-growth technology entrepreneurs but rather local, low-growth firms engaged in complementary, value-adding activities. The composition of this firm creation varies over the technology lifecycle, growing more dispersed across sectors over time and shifting at the mean from upstream activities (e.g., production) to downstream (parts and service). We use this evidence to propose a formal conceptualization of technology-enabled entrepreneurship and discuss insights for entrepreneurial strategy, business ecosystems, and industry dynamics.

Keywords: entrepreneurship, innovation, ecosystems, industry dynamics, value chain

*We are grateful to the editor (Rajshree Agarwal) and two anonymous referees for comments that pushed us to sharpen the insights of this paper. We also thank Ashish Arora, Tim Simcoe, Daniel Keum, Naomi Lamoreaux, John Wallis, participants at the University of Utah Strategy seminar, the DRUID, Doriot, and Strategy Research Forum annual conferences, the AOM annual meeting, the NBER Innovation Information Initiative and DAE meetings, and especially discussants Oliver Alexy and Vera Rocha, for helpful comments. All errors are our own.
†Address: University of Oregon Lundquist College of Business, Eugene, OR, USA; email: innessa@uoregon.edu.
‡Address: Duke University Fuqua School of Business, Durham, NC, USA; email: daniel.gross@duke.edu.
§Address: Columbia Business School, New York, NY, USA; email: jag2367@columbia.edu.

1 Introduction

A central question of strategic management is where and when new market opportunities emerge (Nelson and Winter 1982). The most common attribution is to technological innovation, which is widely recognized as one of the principal catalysts of industrial and economic change (Schumpeter 1942, Tushman and Anderson 1986). Entrepreneurs have been found to be particularly adept at harnessing technological opportunities (e.g., Drucker 1985, Rumelt 1987), a pattern which has made innovative, high-growth startups a frequent focus of strategy research. Common questions in this literature include why some innovative startups grow large while others fail (e.g., Eisenmann 2021), and how high-tech startups impact industry evolution (e.g., Gort and Klepper 1982, Klepper 1996) and contribute to economic growth (e.g., Aghion and Howitt 1992).

Despite being the focus of most modern entrepreneurship research, high-tech startups comprise only a fraction of entrepreneurial activity (Aldrich and Ruef 2018). The vast majority of startups (in the U.S. and elsewhere) will never produce an invention, develop a new market category, or even seek to grow (Hurst and Pugsley 2011). Most startups are instead in the more conventional business of selling existing products or services to established customer segments. These "Main Street" entrepreneurs get less attention in strategy research (with exceptions; e.g., Bennett and Chatterji 2023), perhaps because they are perceived to be non-strategic or structurally unattractive. Yet smaller, local firms providing auto repair services, construction equipment rental, or IT consulting can be strategic, and their markets attractive: Smith et al. (2019) observe that many of the highest earners in the U.S. today are conventional entrepreneurs like these.

In this paper, we investigate the link between innovation and the full range of entrepreneurial opportunities it might present. Echoing the roadmap of Shane and Venkataraman (2000), who ask "why, when, and how opportunities for the creation of goods and services come into existence," we study the level, dynamics, and composition of firms created around new technologies. We call this phenomenon "technology-enabled entrepreneurship."[1] Though these firms are not technology inventors or producers they may nonetheless be significant sources of value. Our goals in this paper are twofold: to (i) develop systematic evidence on this population of firms and the market

[1] Technology-enabled entrepreneurship (TEE) is distinct from technology-*based* entrepreneurship (TBE; Hsu 2008) and innovation-driven entrepreneurship (IDE; Botelho et al. 2021). TBE and IDE characterize (often venture-backed) high growth-intentioned startups that seek to meaningfully redefine existing product categories and industry structures with technological innovation, opening up new markets. Technology-*enabled* entrepreneurs, in contrast, exist across a wider ecosystem—from suppliers, to distributors, to servicers—and support the diffusion and application of new technology through these complementary market functions (Adner 2006).

opportunities they reflect, and (ii) develop a structured approach to understanding these firms, how they compete, and how they relate to major themes in strategic management and related fields, such as innovation ecosystems, industry dynamics, and economic growth.

To systematically study technology-enabled entrepreneurship, we develop a new dataset linking technological innovation to business creation at scale. We first use patent data to identify nearly 1,000 major technologies of the past century, using keywords from U.S. patent publications.[2] We next use administrative business registration records from 47 states—which include all new corporations, limited partnerships, and limited liability companies—to measure over a century of firm creation, and a newly-developed procedure to classify these firms into economic sectors via words in firms' names. We link firms to innovation by identifying firms that include one of our focal technologies in their name.[3] The logic of this approach is that the invocation of a technology in a firm's name explicitly indicates it is closely related to, or even built upon, that technology.[4] We validate this inference by examining the 4-digit SIC industries that firms with these terms are most likely to be associated with in commercial firm directories like Dun & Bradstreet, where industries are usually revealing of firms' technological connections.

This approach has two caveats. First, our analysis is limited to a select set of technologies meeting two conditions: they were heavily developed over time, and they are observable in the patent record. We are thus unable to make statements about business creation around *all* new technologies, including those that are unsuccessful, short-lived, or not patented or patentable. Second, mentions of a technology in firm names incompletely measure its actual use. For example, the word "computer" does not appear in the name of prominent companies like Gateway or Microsoft. However, these are not the modal firms in the economy: far more common are firms like "Personal Computer Services Inc." or "Computer Rental Corp." (real firms in our data). These considerations mean that, even with comprehensive data on firm registrations, we can only partially characterize entrepreneurial activity related to individual technologies.

[2]For each U.S. patent, Google Patents provides "the top 10 salient terms extracted from the patent's title, abstract, claims, and description." For each term, we identify its first use, count up total subsequent uses, and obtain the top 200 new terms from each decade. These are often technological—words like "aircraft", "semiconductor", and "internet"—but to improve our precision, we manually curate this list to a subset which is unambiguously technological, dropping words such as "interface", "analog", "client" and more.

[3]This measurement approach is conceptually similar to (though independently conceived from) the approach taken by Kalyani et al. (2024) to link new technologies to job postings. The methods differ in two key ways: the emphasis on business creation (rather than job postings), and the long panel with a century of new technologies and technology-related firms (rather than only post-1976 technologies and 2010s job postings).

[4]Business names are intended to be a faithful description of the business and are often direct and transparent. Prior work has highlighted how the choices entrepreneurs make at founding, including the company's name, are indicative of their identity and market orientation (Belenzon et al. 2020, McDevitt 2014).

Using these data, we document three sets of facts on technology-enabled entrepreneurship. The first is an accounting. We find that new technologies on average trigger substantial entrepreneurial entry. Across the second half of the 20th century, 1-2% of all businesses registered each year were explicitly (via their name) linked to one of the 860 technologies in our sample, and more heavily developed technologies (i.e., those with more patents) are associated with more new firms. The vast majority of these firms are not manufacturing firms and do not show indicators of high growth intention (based on their legal jurisdiction of choice). At their peak in the 1920s and 1980s, respectively, radio- and computer-related firms (two examples in our data) comprised roughly 1 of every 200 new businesses in the U.S. In contrast, the peak for the median technology in our sample is 1 of every 200,000. Technologies also vary in whether they disproportionately draw entrepreneurial entry in manufacturing, retail trade, services, or other industrial sectors.

Our second set of facts examines dynamics. We begin by characterizing time patterns in business creation around new technologies. We find firm creation often follows an S-curve. Across our sample, we observe two patterns in how the composition of new businesses changes over time. First, on average, entrepreneurial activity disperses across economic sectors and value chain activities over a technology's lifecycle. Second, the distribution of new firms tends to shift downstream: words like *manufacturing* and *engineering* are more likely to appear in firm names early in the technology lifecycle, and words like *maintenance*, *parts*, and *service* become more common later, highlighting how the specific nature of entrepreneurial opportunity evolves from production to aftermarket service as industries develop or mature, and as technologies age.

Our third set of facts involves prediction. Which technologies are associated with follow-on entrepreneurial activity, and when does this activity take place? Our analysis relates firm creation to technologies' early-observable features. We consider three traditional constructs from the innovation literature, measured across a technology's first 100 associated patents: private value (to the patent owner; Kogan et al. 2017), "breakthrough" quality (i.e., impact on future innovation; Kelly et al. 2021), and generality (measured here as the number of distinct co-occurring keywords in these patents). Technologies with higher breakthrough quality and more early recombination are associated with more firm creation, and firm creation that is more dispersed across sectors. Those with higher private value trend in the opposite direction. We then examine a technology-year panel and ask whether recent trends in technological activity predict subsequent firm creation. Contemporaneous and recent (past 1-2 years) patenting activity has a large and precise relationship to the number of associated startups founded in a given technology and year.

The implications of this evidence are potentially broad. Our first and most important insight is the simple fact that there are millions of local businesses that have been enabled by technological innovation, even when not in the business of producing it. Entrepreneurial opportunity introduced by new technology is thus broad and diffuse. However, relative to growth-oriented tech startups with high tendency for R&D and potential for outside investment and lucrative exits, there is less theory and evidence being developed for ventures in this wider set.

In the latter part of the paper, we give conceptual structure to technology-enabled entrepreneurship as a phenomenon, taking an abductive approach.[5] In brief, we arrive at a view of technology-enabled entrepreneurs as firms that help cultivate new market segments, and create and capture value, through activities that enhance the value of new technologies invented and produced by others. We assess possible drivers of long-term advantage for these firms, such as local market power, local relationships and reputation, and extreme specialization, which are distinct from traditional sources of advantage for innovating firms. We then explore gaps in existing strategy frameworks (e.g., entrepreneurial strategy and ecosystems strategy) in relation to technology-enabled startups, and potential extensions that can accommodate them, such as "choosing your activity" before your technology (versus choosing technology first), and dynamic ecosystems strategy (as complementary value creation evolves over the technology lifecycle).

Beyond research investigating the relationship between entrepreneurship and innovation, our paper is related to the literatures on innovation ecosystems (e.g., Moore 1993, Iansiti and Levien 2004a,b, Santos and Eisenhardt 2005, Adner and Kapoor 2010) and industry dynamics (e.g., Abernathy and Utterback 1978, Gort and Klepper 1982, Suarez and Utterback 1995, Klepper 1996, Agarwal and Tripsas 2008, Agarwal et al. 2017). Our analysis brings an industry evolution lens to the ecosystems literature: by examining how entrepreneurial entry upstream of, downstream of, and complementary to new technologies varies over time, we provide a systematic analysis of entrepreneurial dynamics that considers not only technology producers but other technology-related firms as well. The evidence indicates that there is a much broader entrepreneurial ecosystem around new technologies that can flourish even as entry into manufacturing starts to decline, and that this ecosystem evolves over time—insights which raise opportunities to extend existing frameworks for understanding both industry evolution and business ecosystems alike.

The implications of this paper also extend beyond strategy: the scale and scope of technology-

[5] Since our analysis is an empirical accounting of a new (or at least underexamined) phenomenon, we do not perform a contrast of multiple theories as in Kim et al. (2024)—though we still aim to explain our results in a useful (and appropriately general) way that provides theoretical consilience (Pillai et al. 2024).

enabled entrepreneurship suggest these firms may be important to aggregate economic growth. Though endogenous growth literature emphasizes entrepreneurs as innovators, non-innovating startups can also play a crucial, complementary role in converting technological change into economic change—as Hvide and Meling (2023) show for broadband-enabled businesses. Our results suggest they may do so in a variety of ways, including making, selling, installing, and/or servicing new technology, or advising users on how to use it productively. It may be as much through these value-adding activities as through R&D that innovation diffuses into the economy and is transformed into productivity growth. The potential extensions of this paper are thus in our view considerably broader than strategy. We leave these questions to future research.

We proceed as follows. Section 2 elaborates on established connections between innovation and entrepreneurship, why non-innovating firms have received less attention in strategy research, and what might be learned from them now. In Section 3 we describe how we measure innovation and firm creation, and in Section 4 we present a first look at the data, including features of entrepreneurial activity around the top ten technologies of the past century and other motivating descriptive patterns. Section 5 is the core of the paper, where we present our three sets of facts: (i) an accounting of technology-enabled entrepreneurship, (ii) its dynamics, and (iii) its predictors. Section 6 more deeply explores who technology-enabled entrepreneurs are, what they do, and some extensions to existing strategy frameworks that the evidence of this paper suggests. Section 7 reflects on where research might go from here and concludes.

2 Technological vs. Local Entrepreneurship: A False Dichotomy?

Research on the relationship between entrepreneurship and innovation largely centers around a "Schumpeterian" perspective, where startups driven by financial incentives compete through innovation. These technology-based (or innovation-driven) entrepreneurs (TBE/IDE; Hsu et al. 2007, Botelho et al. 2021) are thought to be a major driver of technological progress, industry evolution, and economic growth. In strategic management, the Schumpeterian perspective also emphasizes the importance of appropriability—the ability to capture value from innovation—as a driver of entrepreneurial success (Teece 1986, Gans and Stern 2003). Schumpeterian entrepreneurs are typically high-growth, highly innovative, and venture-backed.

At the other end of the spectrum, a parallel literature has studied self-employment and the choices individuals make to work for themselves vs. others. Unlike Schumpeterian entrepreneurs, these small and medium-sized enterprises (SMEs) typically do not intend to scale their businesses or innovate (Hurst and Pugsley 2011), and they make lower average (but higher variance) returns than comparable salaried workers (Hamilton 2000, Levine and Rubinstein 2017)—a marked contrast to high-tech entrepreneurs, who tend to earn higher returns (Braguinsky et al. 2012). This is not to say that these businesses do not provide important services: local medical and legal professional practices, skilled craftspeople, and personal care services create value for their customers and their communities. But these entrepreneurs are generally thought to have little relation to transformative innovation and have received less attention in strategy research.

The inset below presents a schema of the innovation-based entrepreneurship literature on three dimensions: whether a specific technology is in the core of a firm's strategy, whether the firm makes vs. uses technology, and whether the firm is growth-oriented. The existing strands of research (innovation-driven startups and traditional SMEs) seem to present complementary perspectives, suggesting entrepreneurship exists at the extremes of the distribution, with: (i) hard-driving disruptors who create scalable companies with the potential for lucrative exits and (ii) more passive owner-operators who run small, stable businesses, sometimes at an opportunity cost. Accordingly, the shape of this literature suggests entrepreneurship can be broadly categorized into two groups: high-growth and local (e.g., Schoar 2010).

Categories of startups in relation to technology and growth orientation

Technology critical vs. non-critical	Innovating vs. using innovation	Growth orientation	Description	Prior focus
Critical	Innovating	High	Innovation-driven (TBE/IDE)	X
		Low	Niche products	X
	Using innovation	High Low	Technology-enabled	
Not critical	Using innovation	High	Conventional growth	X
		Low	Traditional SMEs	X

Notes: Table outlines categories of research in entrepreneurial strategy in relation to whether technology is integral to firms' value proposition, whether firms are making versus using technology, and whether they are high- or low-growth oriented. TBE and IDE are acronyms for technology-based entrepreneurship and innovation-driven entrepreneurship (e.g., Hsu 2008, Botelho et al. 2021), which are past conceptualizations of high-growth, innovative startups. Table highlights the segment which is the focus of our attention in this paper.

By our reading, however, the existing literature seems to have made two important oversights. First, strategy research often treats local entrepreneurship as uninteresting—effectively sidelining entrepreneurship that is either small or low-tech. The relative neglect of these firms may reflect a view of them as low-upside and low-impact: neither very strategic nor very relevant to strategy.

However, this view is inconsistent with some of these firms' long-term survival and with features of the U.S. income distribution, where some top earners draw income from small and medium-sized, privately-owned, technology-related firms (Smith et al. 2019).

Second, existing research has focused almost exclusively on firm size as a performance indicator. Strategy research, for example, often focuses on scaling (vis-à-vis sales, headcount, or other metrics). Economic and policy perspectives tend to emphasize job creation. However, a focus on firms' own performance overlooks the broader impacts they might have on industries and economies, up to and including productivity growth: firms may create substantial value even if they fail to capture most of that value themselves. Shifting focus to value created (beyond value captured) broadens the lens to the rest of the value chain (Adner 2012), and uncovers a more nuanced insight. Even if firms remain small and fragmented, these firms can still play an important role in transforming innovation into impact (for example, by facilitating diffusion).

In order for these technology-enabled firms to enter and operate in the myriad ways that they create value for others, they need pathways to (at least some) value capture. Better understanding what the entrepreneurial opportunity landscape is, and how it evolves over a technology's lifecycle, is one goal of this paper. Our data will support the investigation of these questions via patterns of firm creation. In later discussion, we will (more conceptually) explore the role that technology-enabled startups play in value creation, the features of technology-enabled market segments and startups that support firm performance, and how existing frameworks might be extended to incorporate or address these firms' opportunities, activities, and constraints.

3 Data and Measurement

To undertake this analysis, we develop a new dataset to measure entrepreneurial activity that emerges around technological innovation. This dataset, in turn, requires identifying specific technologies, measuring firm creation, and drawing links between them.

3.1 Identifying new technologies through patents

Measuring technologies is less straightforward than might initially appear. There is no single authoritative list of technologies for us to study, in part because a technology, as commonly understood, is not unitary: "technology" often represents a bundle of innovations. Though often used in research, patent data present challenges: patents are written on inventions which are often narrow and comprise parts, devices, or processes which are embodied in products, rather than final products themselves. Patent classes are broader than any one patent alone, yet they can be hard to map to individual technologies (as commonly understood).

Instead of patents or patent classes, we use patent keywords to identify discrete technologies that each patent is associated with, in plain-language—following a growing body of work in management and other fields which takes similar approaches (e.g., Packalen and Bhattacharya 2012, Arts et al. 2021, Gross 2023, or Kalyani et al. 2024). These keywords were obtained from Google Patents public data (using BigQuery) and represent each patent's "top 10 salient terms extracted from the patent's title, abstract, claims, and description."[6] Keywords are typically technological, and sufficiently general that they appear across multiple patents and often clearly identify the technology(ies) to which a given patent relates: for example, some frequently-recurring keywords include *television*, *radio*, *semiconductor*, and *internet*. Other keywords are simply terms that co-occur in patents with these technologies, such as *receiver*, *signal*, *substrate*, or *data*.

We retrieve these keywords for all patents issued by the U.S. Patent and Trademark Office (USPTO) since 1836 (when the patent record begins) and use them to identify new technologies as follows. We first identify the first use of each keyword in the patent record and measure the total number of patents with that keyword (through 2009). We then filter to the 200 most heavily used keywords introduced each decade from the 1880s (thus capturing influential 20th century technologies such as the automobile) to the 2000s (covering modern technologies such as the internet)—drawing uniformly across decades to create a representative sample.[7]

As the above examples illustrate, not all keywords are technologies as typically conceived. We therefore manually review all keywords and sanitize this set of words that are not, by our reading, obviously or unambiguously identifying of a technology in non-patent contexts. For instance, the term "programming" appears in our list, but in addition to software engineering, it also has non-technological uses (e.g., community programming). We also drop biological or chemical entities (e.g., "alkylene"), units of measure (e.g., "pixel"), and terms with less than three characters, as well as a handful of terms related to nuclear physics (e.g., "atomic"), which we have found are often used colloquially in firm names. The end result is a set of 860 technologies. Of these, 658

[6] No formal documentation is available on Google's methodology, but the data developers shared with us over email that these terms are obtained from a procedure similar to TF-IDF (a common method for measuring words' importance in textual documents), with corrections for very common terms and bigrams. This method aims to find terms that are common in a patent, but rare across the population of patents.

[7] We use natural language processing tools to singularize plural nouns before performing this aggregation.

technologies can be linked to firms (discussed below). These comprise the sample for our analysis (see Appendix Table B.2 for a sample list of these technologies).[8]

This approach comes with a few limitations. One is that our sample is conditioned on the total number of patents filed on a technology, which implies we will be studying technologies that with hindsight were influential. Since we do not know how selection creates differences between these technologies and others, we are not able to make statements about the population of all new technologies an entrepreneur may observe at any given moment in time, nor about entrepreneurship around ultimately unsuccessful technologies. Additionally, we sample on patents, and will therefore miss innovations that are not patentable or patented, including many organizational innovations and most business process innovation.[9] Finally, our use of text-based measures raises the question of whether, and how, language has evolved over time. Language associated with novel ideas may develop alongside them, and initial patents may not fully capture the innovation in contemporary language. We believe this is unlikely to be problematic in our case, as most of the keywords in our sample balance specificity with genericity and entered the English lexicon when they began to grow. However, care in interpretation is nevertheless advisable.

Additional innovation characteristics

Having identified a set of technologies via patent keywords, we measure a number of their characteristics from the patent data. We begin with the USPTO historical master file (Marco et al. 2015), which provides a master list of granted patents with grant dates and patent classes. We merge in data on the economic value of patents from Kogan et al. (2017) (henceforth, KPSS), which estimate the stock market response to patent grants since 1926. This measure represents investors' perceived economic value of a patent to the patent holder, and is only available for patents issued to public companies (by construction). We also merge in data on "breakthrough" inventions, as developed by Kelly et al. (2021) (henceforth, KPST), who calculate the textual similarity of each patent to others filed in the years immediately before and after, and measure the ratio of forward to backward similarity—defining patents that were novel for their time, but similar to later patents, as "breakthroughs". We use the ratio of patents' 5-year forward similarity to 5-year backward similarity (as provided by KPST) as our preferred measure. Finally, we use patent data to measure

[8]Though we explored other refinements such as grouping seemingly-related terms (e.g., automotive and automobile), ambiguity over when to group terms—particularly when they could be considered synonyms but are not perfect substitutes—led us to err towards caution by keeping these terms independent.

[9]The U.S. Federal Court of Appeals established the patentability of business methods in a 1998 decision in the case of State Street Bank v. Signature Financial Group (149 F.3d 1368 (Fed. Cir. 1998)).

technologies' combinatoric features, motivated by work in management and economics suggesting recombinant search can produce technological breakthroughs (e.g., Weitzman 1998, Fleming 2001). We use co-occurring patent keywords to measure recombination.

Rather than evaluating technology characteristics generally, we are interested to understand what can be learned specifically at the earliest stages of a technology's lifecycle about entrepreneurial opportunities it may introduce. We therefore focus on early patents. Using the first 100 patents filed on each technology, we measure: (i) the highest KPSS value of any of these patents, (ii) the highest KPST breakthrough value of any of these patents, and (iii) the number of distinct keywords appearing in these patents. Table 2 describes these variables.

[Table 2 about here]

3.2 Measuring firm creation through business registration records

We measure firm creation using state-level business registration records. Business registration is the act of creating a new limited partnership, a corporation, or—since 1993, in most of the U.S.—limited liability companies (LLCs). As such, it represents the legal founding of a company and, in most cases, an anchoring moment of entrepreneurship. This provides us a way to measure business creation over long horizons, as this paper requires, with sufficiently consistent definitions across space and over time. We explain these measures next.

While U.S. corporations have been created since at least the 1600s, it was traditionally a slow process that was tightly controlled by the state (pre-American Revolution, by England; later, by state governments). Since creating new corporations often required state legislatures to pass an act providing it a charter, there was historically a considerable difference between overall entrepreneurship and legal firm formation. The incorporation process was gradually opened up in the late 1800s, beginning with New Jersey in 1896, and quickly expanding to other states. By roughly 1915, two features characterized business incorporation across the U.S. First, all states allowed general (i.e., open) incorporation. Second, firms in every state could choose a legal jurisdiction (a sort of statutory domicile) different from their physical headquarters location. Incorporating firms in practice typically chose between two options: their local state law, which advantages new firms operating in the state, and Delaware corporate law, which is more beneficial for larger firms that engage in business across states, have a large operation, or intend to list on the stock market. In a more recent, post-1988 sample, being registered in Delaware is associated with a 23 times higher probability of reaching a high growth outcome (Guzman and Stern 2020).

We collected data on all business registrations across 50 U.S. states through the Startup Cartography Project (Andrews et al. 2022), of which 47 included historical data.[10] Each record provides the company's name, physical location, state of registration, jurisdiction of choice, corporate form (i.e., corporation versus LLC), and registration date. These data allow us to measure all new legal entities registered in the U.S. since the late 1800s, providing us a rich view of entrepreneurship over time. The major limitations, in our view, are twofold. First, the data do not include firms' industry or business characteristics, which motivates our developing a new way to categorize businesses into sectors. Second, they specifically measure the creation of legal entities, and do not include unincorporated sole proprietors or general partnerships.

Focusing on new legal entities allows us to study an anchoring moment in firm formation that maps naturally to a definition of entrepreneurship: it is the moment a business idea gets translated into an independent organization tasked with executing on it. This is a consistent definition that our data allow us to apply over our entire sample. One implication of this approach, however, is that our measures of firm creation will be sensitive to changes over time in the propensity for incorporation over other organizational forms, including unregistered partnership and sole proprietorship.[11] To limit these concerns, we include technology fixed effects in our analysis where we can, and condition on invention half-decades (and, in robustness checks, invention year; see Appendix C) to compare technologies born into similar institutional environments.

Industry classification

Registration data do not include a firm's line of business. In administrative datasets such as those of the U.S. Census Bureau, firms' industry is categorized into a standardized industry classification by professional coders and proprietary recommendation algorithms, using data collected by mandatory-response census instruments. State offices that register businesses, however, do not classify new registrations into industries, nor do they consistently collect information to support a direct ex-post industry classification by administrators or researchers.

For our aims in this paper, however, we wish to know what business a firm is in. To do so,

[10]The Startup Cartography Project primarily documents entrepreneurship since 1988, but we are able to retrieve the full history of firm registrations, both active and deceased, from nearly all states.

[11]As Guinnane et al. (2007) point out, incorporation patterns have varied across U.S. history, especially in the postwar period as corporate and personal income tax rates diverged, preferencing incorporation.

we build on recent research showing that firm names predict performance and growth potential (McDevitt 2014, Belenzon et al. 2017), and consider that firm names might also provide useful information about industry. We develop a procedure that associates words in firm names with each of ten economic sectors, which we can use to classify firms to sectors. Concretely, we use Dun & Bradstreet (D&B) data, which include both firm names and Standard Industrial Classification (SIC) codes, to train an algorithm that measures the frequency with which individual words in firm names appear in 10 high-level sectors (into which the U.S. SIC classification is organized) and apply this to impute firms' sectors based on their names.[12] This approach mimics methods used by the U.S. Census Bureau to classify firms into industries, which uses a combination of firm names and one-line business descriptions (Kearney and Kornbau 2005), and by D&B itself, which at times uses firm names to impute their industry (Cramer 2017).

As we show in Appendix A, this procedure does a reliably good job of identifying firms' economic sectors in multiple validation samples. It comes with the added advantage that it supports a more flexible classification than SICS or NAICS: firms can be fractionally associated to multiple sectors and thus span sectoral boundaries, as many firms do in practice.

Measuring value chain activities

Following a similar logic, we associate firms to specific value-adding activities related to technological innovation using words in firm names. We focus on five categories which can be viewed as sequential steps in the value chain—research, manufacturing, sales, distribution, and service, omitting other activities that might be more nebulously defined or measured (e.g., marketing). In contrast to our sector classification, which takes a data-driven approach to fractionally associate firms to each of ten economic sectors, we connect the firms in our sample to these five value-adding activities via hand-chosen words and substrings (such as *manufact*, *store*, *service*, or *repair*), which prioritizes precision over completeness. When this procedure associates a firm to multiple activities (because, e.g., the firm's name has both "manufacturing" and "supply" in the name), we discretely measure it as engaged in its most upstream associated activity.

3.3 Linking technologies and firms

The last step in our data collection is to draw links between innovations and firm creation. To do so, we take our initial set of 860 technologies, and search for these technologies in firm names.

[12] These sectors are: Agriculture/Forestry, Construction, Finance/Insurance/Real Estate, Manufacturing, Mining, Public Administration, Retail Trade, Services, Transportation/Public Utilities, and Wholesale Trade.

In doing so, we measure firms which explicitly self-identify with each technology, which we will presume indicates that the firm is closely related to that technology. Within technologies, we then aggregate registrations to count firms for all years, by year, and of various types (e.g., manufacturing firms vs. others, or Delaware vs. local jurisdiction). We link a total of 658 technologies to new firms in this way. Table 2 describes their characteristics.

3.4 Limitations and validation

Our approach, while systematic, also has limits. There are almost certain to be false negatives: firm names will not fully measure firms' relation to new technologies, both because some names do not explicitly convey the business a firm is in and because some firms may be engaging with multiple technologies or a single technology may not be sufficiently characteristic of the firm to be in its name. There is also a risk of false positives: firms that mention a technology in their name without any business relationship to it. Though it seems uncommon prima facie that firms would misrepresent their line of business, particularly given prior evidence that names are used to signal firm types (McDevitt 2014), it is possible that major innovations may be faddish and their use represents marketable naming conventions. To evaluate whether the presence of a technology's keyword in a firm's name is a meaningful signal, we examine the industries these firms are associated with in our Dun & Bradstreet data. Appendix Table B.1 lists all technologies in our data with at least 500 associated firms in D&B, and the top three 4-digit SIC industries they are in. Across the table, these industries reflect the technology in question. A corollary question is whether co-occurring words in firm names are reliable indicators of the sectors and value chain activities that firms are engaged in. The validation of our sector classification method (Appendix A)—which is based on these co-occurring words—shows that this content is revealing.

A second limitation is that there is a small share of firms that we are unable to classify to sectors because the words in their names are either generic stop words or were not in our D&B training data. However, even under our most restrictive procedure, we classify 90% of firms, and this rate is stable over time. Some words included in our technologies list can also be contended, particularly non-nouns, and in robustness checks, we further limit the sample in key analyses to only nouns (Appendix C). As additional references, Appendix Tables B.2 and B.3 show the firm registration counts for technologies with the most patents and most associated firms (respectively), the most-common co-occurring words in firm names (Appendix Table B.2), and three randomly-chosen associated firms as examples (Appendix Table B.3).

4 Descriptive Evidence and Case Studies

We begin our analysis by providing descriptive information on the technologies and firms in our sample, and basic relationships between them. Table 1 lists the technologies with the most associated startups by decade, from the 1900s to the 1990s. For each technology we list the number of associated firms, the share we classify to manufacturing versus other sectors, example firms, and co-occurring words in firm names. The terms that emerge reflect some of the most transformative technologies of the century, including wireless communications, automobiles, video, semiconductors, software, and the internet.[13] Our evidence identifying these technologies as the important technologies driving entrepreneurship over the twentieth century has not, to our knowledge, been seen before. Notably, little of this entrepreneurship is in technology production, albeit with some clear exceptions, such as with semiconductors. For most technologies, associated firm creation covers a broad set of (mostly non-manufacturing) business activities.

[Table 1 around here]

Table 2 presents summary statistics on our full technology sample. The table provides evidence of large variation across technologies in the number of associated firms (which ranges from zero to nearly 100,000, with a mean of ∼1,000 and large skew). We also observe a low share of firms registered in Delaware (a signal of growth intention) or which we classify to manufacturing. The latter results reveal that creating and capturing value around new technology is not simply the strategic domain of innovative, high-growth startups, as is often assumed.

[Table 2 around here]

Example technologies

A number of patterns in our data—and themes of this paper—can be previewed through a few prominent examples. The first of these is the "entrepreneurial lifecycle".

A long literature has examined technology and industry lifecycles (e.g., Abernathy and Utterback 1978, Gort and Klepper 1982, Klepper and Graddy 1990, Klepper 1996, Agarwal et al. 2017), at times with an emphasis on startups (Agarwal and Moeen 2015). In Figure 1 we document the

[13] Computers are notably absent from this table, due to the term emerging in the patent data in the late 19th century (in reference to mechanical calculating machines), but are included in our broader dataset.

evolution of firm creation over time for ten example technologies in our data. For each one, we measure the annual share of all U.S. firm registrations that are associated to this technology (scaled by 100,000 for readability), and smooth sharp fluctuations by calculating a 5-year moving average. The vertical dashed lines mark the technology's "emergence" year, which we define as the first year when it has at least 10 cumulative associated firms and enters a five-year period with more than 10% average annual growth in firm creation—effectively, when firm creation first takes off (akin to the approach taken by Kalyani et al. 2024).

[Figure 1 about here]

In all cases, firm creation grows, peaks, and (in all but the most recent cases) declines. But there is also substantial heterogeneity. As Table 3 shows, some of these technologies at their peak generate an order of magnitude more entrepreneurial entry than others. In some cases, entry is slow to build; in others, fast. Some technologies are multi-peaked, while others stimulate short but sharp bursts of firm creation. Embedded in these figures is an implicit, approximate 'S' curve, with discernible phases in the lifecycle (Gort and Klepper 1982).[14] Though much of the same could be said about previously documented industry lifecycles, this evidence is distinctive in one key way: rather than characterizing technology producers alone, it reflects the full range of entrepreneurial activity, most of which consists of firms that are not producing the technology but are nevertheless building new enterprises around it. When we transition to systematic analysis across the full sample in Section 5, we will also expand the sampling frame to a larger set of technologies than has been feasible in most prior scholarship on industry evolution, which typically takes the form of single-industry studies or a samples with at most a few dozen products or technologies.

[Table 3 about here]

Figure 2 examines the sectoral composition of these startups.[15] Though a few technologies have a large share of entry in manufacturing (e.g., semiconductors, and to a lesser degree electronics and

[14] In Appendix B we also undertake an empirical exercise to compare our sample and methods to those of Klepper and Graddy (1990) (KG), who characterize 46 products (some technological, some not) into three stages vis-à-vis net entry into manufacturing: growth, shakeout, and stabilization. For the subset of the products in KG we could reliably search for in our data, in Appendix Figure B.1 we plot the smoothed entrepreneurial lifecycle. Intriguingly—yet reassuringly—we find that KG's Stage 1 (take-off) generally coincides with the take-off in firm creation and Stage 2 (shakeout) coincides with the beginning of a permanent decline.

[15] In preparing these figures, we present technologies' sector shares for contiguous years with at least 10 firms, to reduce noise and aid visualization. Note that this chart is most easily read in full color.

aircraft), others concentrate in construction (HVAC), transportation and utilities (aircraft, radio, television), or services (automotive, software, internet). Research studying only one sector will thus tend to omit most technology-driven entrepreneurial activity.

[Figure 2 about here]

5 Systematic Patterns

5.1 The quantity and composition of technology-enabled startups

We now provide a more systematic accounting of technology-enabled entrepreneurship. We begin by measuring how many such firms there are, and what they do. Figure 3 provides a first accounting, showing the share of annual U.S. business registrations associated with the technologies in our sample by year, from 1920 to 2000, separately reporting firms registered in Delaware versus other jurisdictions and firms we classify into manufacturing versus other sectors. When we look across the twentieth century, we find that a substantial share of entrepreneurship is related to new technology. Between 1950 and 2000, roughly 1-2% of new firms each year explicitly include one of the technologies in our sample in their name—a total which almost certainly undercounts technology-enabled entry due to both our limited technology sample and our reliance on firm names, which only partially measure the phenomenon.[16]

The vast majority (∼85%) of these firms are local, non-manufacturing firms, with most of the remainder evenly split between local manufacturing and Delaware-registered non-manufacturing firms. Rather than fast-moving, high-growth innovators, most firms linked to new technologies thus have founding characteristics consistent with small, local enterprises that are creating and capturing value by providing ancillary goods and services, finding and exploiting the market opportunities that innovation creates—such as local equipment sales, rental, installation, maintenance, or advisory services. We also see variation in this aggregate accounting over time, with two inflection points: a permanent increase after World War II, and a temporary surge in the 1980s with the growth of personal computer-, electronics-, and software-related businesses.

[Figure 3 about here]

[16] This estimate is much larger than the number of venture-backed firms, which effectively did not exist prior to 1970 and even today only comprises less than 0.1% of U.S. startups in any given year (Kerr et al. 2014).

Figure 4 explores whether more heavily-developed technologies also have more associated firms, presenting a binned scatterplot of log firms against log patents for our sampled technologies, conditional on the year of each technology's first patent. We do so for all firms, manufacturing vs. non-manufacturing firms, and Delaware vs. non-Delaware registered firms. We find a large correlation between these two measures: a doubling of patents is associated with 53% more firm creation in a given technology. Strikingly, this relationship is statistically and quantitatively very similar across all firm types, revealing a second insight: not only are more developed technologies more likely to spawn new business, but they appear to do so for businesses of all flavors, including small, local, non-technology producing firms that are unlikely to grow.

[Figure 4 around here]

5.2 Entrepreneurial dynamics over the lifecycle

5.2.1 The dispersion of entrepreneurial entry

We next examine the evolution of technology-enabled startups over the technology lifecycle. Despite a long history of research on technology industry dynamics in strategy and other fields, relatively little attention has been given to understanding the decentralized emergence of a wider ecosystem around new technologies or considering how it evolves alongside the more frequently studied technology-producing sector, despite the number of firms that are demonstrably pursuing other technology-enabled opportunities (as seen in Figure 3).

Our starting point is to examine the dispersion of entrepreneurial activity across value chain positions and economic sectors over the technology lifecycle. Because there are many ways to measure dispersion, we take several approaches. One is to measure the top activity's (or sector's) share of firm creation. Another is to construct a Herfindahl index across activities (sectors). A third is to measure the coefficient of variation (C.V.), calculating the variance of each activity's (sector's) share of firm creation in a given technology divided by the mean (as Kalyani et al. 2024 do in studying the geographic spread of jobs related to new technologies). All three are concentration measures that, when they decline, will signal entrepreneurial entry growing more heterogeneous and diffuse, and (for the C.V.) more balanced across sectors.

In Table 4 we estimate a regression of each measure, calculated at the technology-year level, on indicators for each decade of a technology's lifecycle, controlling for technology and calendar year fixed effects. Concretely, we estimate the following equation:

$$Y_{it} = \sum_{s} \beta_s \cdot \mathbb{1}(\text{Technology is } s \text{ years into lifecycle}) + \alpha_i + \delta_t + \varepsilon_{it} \qquad (1)$$

where i and t index technologies and calendar years, α_i and δ_t are fixed effects, and β_s are the parameters of interest. Columns (1) to (3) of the table examine activity-based concentration measures, and Columns (4) to (6) sector-based measures. All of these measures decline monotonically across technologies' lifecycles. This indicates a substantial dispersing of entrepreneurial activity as technology-based economies evolve. The estimated declines are large in magnitude for activities relative to sectors, which wash out more of the variation, but still large for the latter. This evidence reinforces our thesis that new technologies present highly varied opportunities for new enterprises, and that these opportunities grow increasingly varied as technologies age.

[Table 4 about here]

5.2.2 Migration of entrepreneurship down the value chain

As entry disperses, how does the positioning of these firms change? Table 5 uses co-occurring words and our sector and activity classifications to evaluate how economic activities undertaken by technology-related startups vary over the lifecycle. Panel (A) reports the relative frequency of the manufacturing, services, and 'other' sectors among the first 100 firms in a given technology, the 101-200th, the 201-500th, and so on. Panel (B) does so for specific value chain activities, and Panel (C) for specific co-occurring words in firm names, which we partition into three sets: words that are relatively common early in the technology lifecycle but which decrease across it, those that increase, and those that follow an inverted-U pattern.

[Table 5 around here]

We observe three patterns. First, entry related to the production and development of a technology tends to peak early in its lifecycle. This is apparent for manufacturing sector firms (Panel A), firms we associate to research and production activities (Panel B), and paired words like "manufacturing" or "engineering" (Panel C). Second, as technologies mature, downstream entry appears to grow: Table 5 shows an increasing relative frequency for service sector firms (Panel A), firms we associate to service activities (Panel B), and for terms like "maintenance", "parts", and "service"

(Panel C). Third, firms more closely related to distribution follow an inverted-U pattern, peaking along the middle of the innovation's lifecycle: in the middle of the lifecycle is when paired words such as "equipment" and "supply" reach their peak.

We interpret this evidence as indicating that, as technologies mature, entry shifts from production to servicing, which increasingly adds value by maximizing the longevity of technologies that have already diffused. In between, entry also gravitates towards complementary activities, including those that support its adoption, like sales and distribution. Whereas prior work has often emphasized a decline in entry as technological industries mature and consolidate (Klepper and Graddy 1990), our broader firm sample instead shows a downstream migration of entrepreneurial entry—and by revealed preference, commercial opportunity.

We more formally examine these transitional dynamics in Table 6, which estimates changes in the share of firms associated to manufacturing vs. services, at the technology-year level, on indicators for each decade of the technology lifecycle (Equation 1). Columns (1) and (2) of Table 6 examine activity shares, and Columns (3) and (4) sector shares. The results document a similar, monotonic shift on balance as Table 5, from production-oriented firms to service-oriented firms, with the decline in the former offset by growth in the latter. [17]

[Table 6 about here]

5.3 What predicts technology-enabled startups?

Which technologies create more startups and when? Evaluating which technologies are likely to present high-potential commercial opportunities is a problem at the center of the entrepreneur's decision to start a technology-related firm (Gans et al. 2021).

We first ask whether characteristics which can be measured early in a technology's lifecycle predict the total number of firms created around that technology. Table 7, Panel (A) estimates a cross-technology regression of (log) firm creation against the three characteristics described in Section 3: the maximal private value (to the patent holder) of the first 100 related patents (estimated from stock price reactions, for patents issued to public firms; Kogan et al. 2017), the maximal "breakthrough" value (which captures the departure from prior art and impact on subsequent

[17] Our results also speak to Cusumano et al. (2015), who study how individual companies (such as Oracle) have moved into services as industries mature, but do not assess differences in entry rates or across other technology-associated industries.

patenting; Kelly et al. 2021), and the level of recombination (the number of distinct, co-occurring keywords in the first 100 patents). We standardize values to simplify interpretations and comparisons of effects. Formally, we estimate the following specification:

$$Ln(\text{Total firms})_i = \beta_1 \cdot Std(\text{Private value})_i + \beta_2 \cdot Std(\text{Breakthrough value})_i + \beta_3 \cdot Std(\text{Recombination})_i + \zeta_i + \varepsilon_i \tag{2}$$

where i indexes technologies, and we include fixed effects for the decade of a technology's emergence (ζ_i). Column (1) estimates this relationship for all firms, and Columns (2) to (5) do so for: Delaware-registered (i.e., growth-oriented) versus other firms, and manufacturing versus non-manufacturing firms. Due to the construction of the dependent variable, the sample in each column is limited to technologies that have non-zero firm creation of each type.

[Table 7 about here]

We find that firm creation does not correlate strongly with the private value of early patents—a result which reflects that the independent variable measures investor expectations of value capture by patent owners rather than spillovers. In contrast, the existence of early "breakthrough" patents correlates positively with firm creation of all types—with a one standard deviation increase associated with roughly 35-40% greater entry. This is also the case for the level of recombination, with a one standard deviation increase related to 20-25% greater entry.

In Panel (B) we consider instead the dispersion of entry across value chain activities and economic sectors, using the same outcome measures as in Table 4. We do not observe a strong correlation between private value or breakthrough value and dispersion in our data. We do observe recombination leading to lower concentration (higher dispersion), though only statistically significant for our activity-based measure. Estimates for sector-based concentration are directionally similar, but smaller in magnitude and not statistically significant.

Taken together, these results reveal several comparative statics relevant to technology selection in the formulation of entrepreneurial strategy. The first is technologies that depart from recent trends in patenting but shape the path of future patenting—the Kelly et al. (2021) approach—also shape the path of future entrepreneurial opportunity. Moreover, the statistical null relationship of firm creation to private patent value is itself interesting: one might expect the commercial potential of a technology to be reflected in both high market value to the patent holder and firm creation (generating a positive relationship) or that incumbent firms which hold valuable patents may work to foreclose entry around their technology (generating a negative relationship). The null result may potentially reflect the intersection of these countervailing forces.

The most striking finding across both panels, in our view, is the relationship between recombination and firm creation. Technologies with wide and rapid experimentation in their earliest years generate significantly more new firms, across a wider range of activities. In other words, an early signal that an innovation is likely to generate a large number of market opportunities, and perhaps diffuse widely, is to observe it developing with a wide set of distinct varieties, complements, and applications—the objects that co-occurring, recombinant patent keywords represent. This insight connects closely with notions of general-purpose technologies (Bresnahan and Trajtenberg 1995), which have been useful for framing thought but difficult to empirically operationalize. Here we have an intuitive approach to measuring technological generality as a choice-revealed characteristic and relating it to firm creation. For the entrepreneur, this blossoming might be observed in the patent record as a signal of emerging entrepreneurial opportunity.

Finally, we examine whether the level and timing of technology-enabled entrepreneurship can be predicted by contemporaneous or recent patenting activity. In Table 8, we estimate relationship between firm creation and patenting at the technology-year level:

$$Ln(Firms)_{it} = \beta \cdot Ln(Patents)_{it} + \alpha_i + \delta_t + \varepsilon_{it} \tag{3}$$

and

$$Ln(Firms)_{it} = \sum_{s=1}^{4} \beta_s \cdot Ln(Patents)_{i,t-s} + \alpha_i + \delta_t + \varepsilon_{it} \tag{4}$$

where i and t index technologies and calendar years, and α_i and δ_t are fixed effects. Panel (A) presents results for Equation (3) and Panel (B) for Equation (4).

[Table 8 about here]

The associations we find are positive and precisely estimated. Across all types of firms (including manufacturing, non-manufacturing, Delaware, non-Delaware; Columns 1 to 5), we find in Panel (A) that patenting and firm creation coincide, with a doubling in patents in a given year associated with 25-30% higher firm creation in that year. Invention and entrepreneurship are, of course, co-determined in equilibrium, and the pattern likely reflects the coincidence of technological and market opportunity. In Panel (B), we find that patenting over the prior two years, and especially for the prior year, correlates with current (year t) entrepreneurship: a doubling of patents on a given technology in the prior year is associated with 20-25% higher firm creation of all types in that technology in the current year, and a doubling two years ago, with 5-10% higher firm creation in the current year. This result may reflect the same underlying fundamentals as Panel (A), due to serial correlation in patenting, but it hints at a predictive rubric for assessing imminent changes in opportunity and a potential structural relationship.

To be clear, Tables 7 and 8 are a prediction exercise and do not estimate or represent causal relationships. Even so, the evidence presents several new facts on the correlates of technology-enabled entrepreneurship, such as the elevated level and breadth of entrepreneurial activity around more technologically-important and general-purpose innovation, and that it tends to coincide in time and space with technological developments.

6 Implications for Research and Practice

These results collectively reinforce that the universe of technology-related entrepreneurship is much larger than the subpopulation of innovating startups usually studied by research in strategy and adjacent fields. In this section we position these firms within strategy research and infer features that may be informative for managerial practice. Due to data limitations—including lack of direct measures of firms' positioning or performance—our approach is primarily abductive, as we seek to make sense of the evidence of the large population of firms engaged in activities across technology value chains and examine what its implications are likely to be.

We consider two sets of questions. The first set concerns managerial choices and the strategic logic of technology-enabled entrepreneurship. When in a technology's lifecycle do opportunities emerge? What are potential ways new firms can attack these opportunities, and what challenges stand in their way? Finally, how does this strategic logic compare to that of innovation-producing firms? The second set of questions considers how our findings relate to prior research, and whether extensions to existing intellectual frameworks are needed.

6.1 Strategy for the technology-enabled entrepreneur

Figure 5 summarizes our conceptualization of technology-enabled entrepreneurship and major differences in comparison to ecosystem innovators, which we define as inventing firms that create value through collaborative arrangements with others. We focus on this comparison to improve precision,harnessing the fact that ecosystem innovators exist symbiotically with technology-enabled startups—e.g., an innovation-based company like Kodak and small, local photo developers—enabling a parsimonious analysis of common differences between these two categories.

[Figure 5 about here]

We start by considering the nature of the market opportunity. Ecosystem innovators' primary activities are developing and commercializing innovation, and (to the extent possible or necessary) orchestrating the value chain. In contrast, technology-enabled firms are arms-length participants in the value chain, often in a localized and/or specialized niche.

Existing research on product lifecycles suggests that opportunities for innovative startups will tend to be greatest earlier in technology lifecycles, before a dominant design emerges (e.g., Suarez and Utterback 1995, Klepper 1996). Technology-enabled startups, however, benefit from products being sufficiently standardized that there is a well-defined value chain. As a result, entrepreneurial opportunity will tend to be higher later in technology lifecycles, once a technology has matured. This logic is consistent with Stigler's (1951) observation that firms in young industries will tend to be vertically integrated, but that as industries grow, they can increasingly disintegrate—in turn creating opportunities for other entrants up and downstream.

The potential drivers of long-term advantage are also distinct. Technology-enabled firms can create value by finding a niche. They may also obtain pricing power through hyper-local natural monopolies (e.g., in specific neighborhoods) or by serving a specific clientele. Those intending to be suppliers can be hyperspecialized or have exclusive partnerships with innovation producers—though the latter also presents risks of holdup. Specialized local market knowledge, local relationships and reputation, and narrow focus may help these firms outperform on cost. These potential sources of advantage are reflexively mirrored by ecosystem innovators: common drivers of sustained advantage are intellectual property over the core technology, ownership over the means of production and/or its efficient use, economies of scope and scale, and bargaining power vis-à-vis the rest of the technology-supporting ecosystem, due to its unique position at the center.

Common challenges facing technology-enabled firms are typically structural. Limited ownership of proprietary assets, low entry barriers, labor intensity, and difficulty scaling can all put pressure on profitability. Local small business categories like equipment rental or installation, for example, can be heavily competed. Innovating firms, on the other hand, are exposed by their dependence on a wider technology-enabled ecosystem with local market knowledge and reach, particularly because penetrating hyperlocal markets often requires assets or activities which are sufficiently incongruent with innovation producers' core business to pursue directly.

The bottom of our table provides examples of firms of each type which we find in our data. Print and copy equipment manufacturers interdepend with hundreds (or thousands) of business service centers, auto manufacturers interdepend with thousands of dealers and tens of thousands of mechanics, HVAC equipment-makers interdepend with thousands of heating and cooling installation and repair services, and so on. One of the more salient differences between the innovative, technology-producing firms in this list and their technology-enabled counterparts is in their returns to scale: whereas technology-producing firms may have increasing—or only slowly decreasing—returns to scale, technology-enabled firms are more likely to have rapidly diminishing returns. This has more fundamental implications for industry structure and is a reason why firms may remain small and specialized, and why competition may proliferate.

More broadly, these examples can help us better understand the roles technology-enabled firms play in the high-tech economy. Though the set of such firms is large and heterogeneous, there are a few canonical categories. One is firms which support the diffusion of new technology through rental markets: firms like copy centers, video rental stores, Internet cafes, or 3D printers make indivisible capital divisible, acquiring high fixed cost technology and exploiting knowledge of local demand to find a market and rent to local consumers. Equipment dealers (like auto dealers) represent a related but distinct category: exclusive distributors of specific technologies. Repair services are another, and their availability influences the rate at which capital depreciates. Each of these functions can be modeled formally—such that there are specific, natural ways to incorporate technology-enabled entrepreneurship into research on the impacts of innovation.

6.2 Extensions to existing models and frameworks

The evidence of a large entrepreneurial economy around new technologies has bearing on multiple literatures in strategic management and opens the door to several new questions. In Figure 6 we examine implications for high-tech industries broadly (top) and for technology-enabled startups (left) and innovation-driven startups (right) specifically.

[Figure 6 about here]

The evidence adds first to the classical industry studies literature, which typically focuses on entry and competition among technology producers alone (e.g., Gort and Klepper 1982, Klepper and Graddy 1990, Klepper 1996, Agarwal and Gort 1996). Our work expands this analysis to technology-related firm creation in other sectors—a population of firms which is an order of magnitude larger and serves complementary market functions. Considering the traditional industry evolution paradigm in isolation, one may conclude that the entrepreneurial opportunities around a technological innovation plateau as a technology matures. Our work highlights that many entrepreneurial opportunities can persist into (or even emerge in) the later stages of the technology lifecycle, including after the shakeout of producers—and on average the nature of these opportunities migrate downstream, from manufacturing to maintenance.

Our evidence also links to research in entrepreneurial strategy, which seeks to understand how new firms can identify and exploit new opportunities to create and capture value, and why some new firms do so better than others. These opportunities are often thought to be found in the introduction of new technology—yet the more "workaday" technology-enabled businesses engaged in distributing, installing, and maintaining these technologies can also be profitable, as we have previously noted. Whereas entrepreneurial strategy research proposes entrepreneurs "choose their technology" (Gans et al. 2021) before making new investments, our evidence suggests entrepreneurs might also consider first "choosing their activity" (a target position in the value chain). Indeed, these same entrepreneurs may later find themselves pivoting around that activity as technologies cycle, whereas those which choose technologies may obsolesce.

The evidence of a large population of innovation-supporting startups inevitably implicates innovation producers themselves, reinforcing the value to the innovator of having an ecosystems strategy, as well as the value of ecosystems as a construct and unit of analysis in strategic management (Adner 2006, Adner and Kapoor 2010). However, whereas existing literature largely characterizes innovation ecosystems as a collection of intentional actors undertaking coordinated investments—Adner (2017, p. 39) defines ecosystems as "the alignment structure of a multilateral set of partners"—the ecosystem we document in this paper comprises a much broader collection of firms, most of which have no relationship (formal or informal) to the innovation producer nor to each other. These firms can play a crucial role in enhancing the value of the innovation itself, thus suggesting a broader conceptualization of "arms-length ecosystems" and a broadening of ecosystems strategy to consider the innovation-supporting economy. The evidence that specific complementors become more or less common at different stages of a technology's lifecycle also suggests ecosystems strategy may need to be dynamically adjusted over time, as technologies mature.

The implications of technology-enabled entrepreneurship potentially extend further. Scholarship in the management and economics of innovation considers how firms can profit from innovation (e.g., Teece 1986), and the role of innovation in propelling industrial change and economic growth (Schumpeter 1942, Aghion and Howitt 1992). The size of the technology-supporting economy, and the wide range of firms that comprise it, suggest new possibilities for these lines of work. One is extending the Teecian paradigm to consider not only how firms may profit from their own innovation, but also how they may profit from *other firms'* innovation, especially up and downstream of the innovation itself. Another is in studying Schumpeterian growth: although the Schumpeterian paradigm traditionally emphasizes innovative entrepreneurs as engines of "creative destruction," our evidence suggests there is a second entrepreneurial economy supporting the productive use of new technology that innovative startups create. These firms are the "nuts and bolts" of the high-tech economy—what Russell and Vinsel (2018) dub "maintainers"—and without them, Schumpeterian growth might be impossible altogether. With its focus on innovative startups, prior work may thus be underestimating the economic impact of entrepreneurship.

7 Concluding Remarks

The opportunities, challenges, and impacts of innovation are the subject of decades of scholarship in strategic management and other fields. Despite this attention, there has been little broad, systematic evaluation of the diverse ways innovation affects the entry of new firms, the characteristics of these firms, and the technological features that predict entry. Using 860 major technologies identified through patent data, linked to a century of firm creation via firm names, we characterize the level, composition, and dynamics of what we label "technology-enabled entrepreneurship". Many firms (1-2% of newly-registered U.S. businesses) are explicitly linked to technologies in our sample, and most are non-manufacturing and low growth-intentioned—representing a large economy of small businesses pursuing opportunities around the creation, diffusion, and maintenance of these technologies. We then document patterns in the entrepreneurial lifecycle, finding that startup activity grows increasingly diverse as technologies mature, and show that the level and variety of firm creation can be predicted from technologies' early features.

Our results complement existing research in several literatures, including industry evolution, entrepreneurial strategy, and innovation management. At its essence, the core findings of this paper are that entrepreneurial dynamics and industry dynamics are distinct phenomena, and the nature of entrepreneurial opportunity can change as technologies, and industries, evolve. Our results suggest that by focusing on the creators and manufacturers of new products, existing research on industry dynamics may be missing a significant fraction of the market opportunity that innovation engenders over time. The evidence of this "hidden economy" of technology-enabled entrepreneurship points to substantial opportunities for future research in strategy and beyond.

Several of these questions were enumerated earlier. First, there is the competitive strategy lens: how structurally attractive is technology-enabled entrepreneurship, what makes some firms more profitable than others, and what new conceptual tools (if any) are needed to create and sustain successful technology-enabled businesses? There is also the ecosystems frame: how do these firms relate to innovating firms—and what is the role of this broader, arms-length, economy-wide ecosystem in comparison to the narrow set of firms which collaborate with innovation producers? The Teecian view leads us to consider the possibility of profiting from others' technology, and ask: how do new firms establish advantaged positions in these complementary markets? And as we have noted, the implications potentially roll up to aggregate outcomes, as we ponder the role of these often-overlooked small businesses in providing a crucial bridge between innovation and its use in production: what role do they play in the process of economic growth?

The limitations of our work also leave open important questions. For example, we lack financial performance data to evaluate the profitability of the firms in our sample, and to examine sources of performance differences within and across segments or over time. We similarly lack survival data to examine determinants of survival or characterize net entry. These, and more, remain opportunities for future work. Yet the audience for research on technology-enabled entrepreneurship is large, and includes scholars, managers, and policymakers. We will thus end by encouraging future research—including our own—to continue to engage with these questions.

References

Abernathy, William J and James M Utterback. 1978. "Patterns of industrial innovation," *Technology Review*, Vol. 80, No. 7, pp. 40–47.

Adner, Ron. 2006. "Match your innovation strategy to your innovation ecosystem," *Harvard Business Review*, Vol. 84, No. 4, p. 98.

——— . 2012. *The Wide Lens: What successful innovators see that others miss*. New York: Portfolio/Penguin.

——— . 2017. "Ecosystem as structure: An actionable construct for strategy," *Journal of Management*, Vol. 43, No. 1, pp. 39–58.

Adner, Ron and Rahul Kapoor. 2010. "Value creation in innovation ecosystems: How the structure of technological interdependence affects firm performance in new technology generations," *Strategic Management Journal*, Vol. 31, No. 3, pp. 306–333.

Agarwal, Rajshree and Michael Gort. 1996. "The evolution of markets and entry, exit and survival of firms," *The Review of Economics and Statistics*, pp. 489–498.

Agarwal, Rajshree and Mahka Moeen. 2015. "Entrepreneurial startups (de novo), diversifying entrants (de alio) and incumbent firms," *The Palgrave Encyclopedia of Strategic Management. Palgrave Macmillan: London*, pp. 504–507.

Agarwal, Rajshree, Mahka Moeen, and Sonali K Shah. 2017. "Athena's birth: Triggers, actors, and actions preceding industry inception," *Strategic Entrepreneurship Journal*, Vol. 11, No. 3, pp. 287–305.

Agarwal, Rajshree and Mary Tripsas. 2008. "Technology and industry evolution," *The handbook of technology and innovation management*, Vol. 1, pp. 1–55.

Aghion, Philippe and Peter Howitt. 1992. "A model of growth through creative destruction," *Econometrica*, Vol. 60, pp. 323–351.

Aldrich, Howard E and Martin Ruef. 2018. "Unicorns, gazelles, and other distractions on the way to understanding real entrepreneurship in the United States," *Academy of Management Perspectives*, Vol. 32, No. 4, pp. 458–472.

Andrews, Raymond J, Catherine Fazio, Jorge Guzman, Yupeng Liu, and Scott Stern. 2022. "The Startup Cartography Project: Measuring and mapping entrepreneurial ecosystems," *Research Policy*, Vol. 51, No. 2, p. 104437.

Arts, Sam, Jianan Hou, and Juan Carlos Gomez. 2021. "Natural language processing to identify the creation and impact of new technologies in patent text: Code, data, and new measures," *Research Policy*, Vol. 50, No. 2, pp. 104–144.

Belenzon, Sharon, Aaron K Chatterji, and Brendan Daley. 2017. "Eponymous entrepreneurs," *American Economic Review*, Vol. 107, No. 6, pp. 1638–55.

——— . 2020. "Choosing between growth and glory," *Management Science*, Vol. 66, No. 5, pp. 2050–2074.

Bennett, Victor M and Aaron K Chatterji. 2023. "The entrepreneurial process: Evidence from a nationally representative survey," *Strategic Management Journal*, Vol. 44, No. 1, pp. 86–116.

Botelho, Tristan L, Daniel Fehder, and Yael Hochberg. 2021. *Innovation-driven entrepreneurship*. NBER Working Paper No. 28990.

Braguinsky, Serguey, Steven Klepper, and Atsushi Ohyama. 2012. "High-tech entrepreneurship," *The Journal of Law and Economics*, Vol. 55, No. 4, pp. 869–900.

Bresnahan, Timothy F and Manuel Trajtenberg. 1995. "General purpose technologies 'Engines of growth'?" *Journal of Econometrics*, Vol. 65, No. 1, pp. 83–108.

Cramer, Jonathan James. 2017. "Dun & Bradstreet enlists the help of AI to solve a decades-old problem," Available at https://www.dnb.com/perspectives/master-data/how-dnb-improves-sic-classification-ai.html.

Cusumano, Michael A, Steven J Kahl, and Fernando F Suarez. 2015. "Services, industry evolution, and the competitive strategies of product firms," *Strategic Management Journal*, Vol. 36, No. 4, pp. 559–575.

Drucker, Peter F. 1985. *Innovation and entrepreneurship*. New York: HarperCollins.

Eisenmann, Tom. 2021. *Why startups fail: A new roadmap for entrepreneurial success.* New York: Currency.

Fleming, Lee. 2001. "Recombinant uncertainty in technological search," *Management Science*, Vol. 47, No. 1, pp. 117–132.

Gans, Joshua S, Michael Kearney, Erin L Scott, and Scott Stern. 2021. "Choosing technology: An entrepreneurial strategy approach," *Strategy Science*, Vol. 6, No. 1, pp. 39–53.

Gans, Joshua S and Scott Stern. 2003. "The product market and the market for "ideas": Commercialization strategies for technology entrepreneurs," *Research Policy*, Vol. 32, No. 2, pp. 333–350.

Gort, Michael and Steven Klepper. 1982. "Time paths in the diffusion of product innovations," *The Economic Journal*, Vol. 92, No. 367, pp. 630–653.

Gross, Daniel P. 2023. "The hidden costs of securing innovation: The manifold impacts of compulsory invention secrecy," *Management Science*, Vol. 69, pp. 2318–2338.

Guinnane, Timothy, Ron Harris, Naomi R Lamoreaux, and Jean-Laurent Rosenthal. 2007. "Putting the corporation in its place," *Enterprise & Society*, Vol. 8, No. 3, pp. 687–729.

Guzman, Jorge and Scott Stern. 2020. "The state of American entrepreneurship: New estimates of the quantity and quality of entrepreneurship for 32 US states, 1988-2014," *American Economic Journal: Economic Policy*, Vol. 12, No. 4, pp. 212–43.

Hamilton, Barton H. 2000. "Does entrepreneurship pay? An empirical analysis of the returns to self-employment," *Journal of Political Economy*, Vol. 108, No. 3, pp. 604–631.

Hsu, David H. 2008. "Technology-based entrepreneurship," *Handbook of Technology and Innovation Management*, Vol. 3, pp. 367–388.

Hsu, David H, Edward B Roberts, and Charles E Eesley. 2007. "Entrepreneurs from technology-based universities: Evidence from MIT," *Research Policy*, Vol. 36, No. 5, pp. 768–788.

Hurst, Erik and Benjamin Wild Pugsley. 2011. "What do small businesses do?" *Brookings Papers on Economic Activity*, Vol. Fall 2011, pp. 73–142.

Hvide, Hans K. and Tom G. Meling. 2023. *New technology and business dynamics.* Working paper.

Iansiti, Marco and Roy Levien. 2004a. *The keystone advantage: What the new dynamics of business ecosystems mean for strategy, innovation, and sustainability.* Boston: Harvard Business Press.

——— . 2004b. "Strategy as ecology," *Harvard Business Review*, Vol. 82, No. 3, pp. 68–78.

Kalyani, Aakash, Nichlas Bloom, Marcela Carvalho, Tarek Alexander Hassan, Josh Lerner, and Ahmed Tahoun. 2024. *The diffusion of new technologies.* NBER Working Paper No. 28999.

Kearney, Anne T and Michael E Kornbau. 2005. "An automated industry coding application for new US business establishments," in *Proceedings of the American Statistical Association.*

Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy. 2021. "Measuring technological innovation over the long run," *American Economic Review: Insights*, Vol. 3, No. 3, pp. 303–20.

Kerr, William R, Ramana Nanda, and Matthew Rhodes-Kropf. 2014. "Entrepreneurship as experimentation," *Journal of Economic Perspectives*, Vol. 28, No. 3, pp. 25–48.

Kim, Seojin, Rajshree Agarwal, and Brent D Goldfarb. 2024. *Creating competencies for radical technologies: Revisiting incumbent-entrant dynamics in the bionic prosthetic industry.* SSRN 4024471.

Klepper, Steven. 1996. "Entry, exit, growth, and innovation over the product life cycle," *American Economic Review*, pp. 562–583.

Klepper, Steven and Elizabeth Graddy. 1990. "The evolution of new industries and the determinants of market structure," *RAND Journal of Economics*, pp. 27–44.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman. 2017. "Technological innovation, resource allocation, and growth," *Quarterly Journal of Economics*, Vol. 132, No. 2, pp. 665–712.

Levine, Ross and Yona Rubinstein. 2017. "Smart and illicit: Who becomes an entrepreneur and do they earn more?" *Quarterly Journal of Economics*, Vol. 132, No. 2, pp. 963–1018.

Marco, Alan C, Michael Carley, Steven Jackson, and Amanda Myers. 2015. *The USPTO historical patent data files: Two centuries of innovation.* SSRN working paper, available at https://ssrn.com/abstract=2616724.

McDevitt, Ryan C. 2014. ""A" business by any other name: Firm name choice as a signal of firm quality," *Journal of Political Economy*, Vol. 122, No. 4, pp. 909–944.

Moore, James F. 1993. "Predators and prey: A new ecology of competition," *Harvard Business Review*, Vol. 71, No. 3, pp. 75–86.

Nelson, Richard R and Sidney G Winter. 1982. *An evolutionary theory of economic change*. Cambridge: Harvard University Press.

Packalen, Mikko and Jay Bhattacharya. 2012. *Words in patents: Research inputs and the value of innovativeness in invention*. NBER Working Paper No. 18494.

Pillai, Sandeep Devanatha, Brent Goldfarb, and David Kirsch. 2024. *Lovely and likely: Using historical methods to improve inference to the best explanation in strategy*. Forthcoming at Strategic Management Journal.

Rumelt, Richard P. 1987. "Theory, strategy, and entrepreneurship," in Teece, David J ed. *Competitive Challenge: Strategies for Industrial Innovation and Renewal*, pp. 137–158.

Russell, Andrew L and Lee Vinsel. 2018. "After innovation, turn to maintenance," *Technology and Culture*, Vol. 59, No. 1, pp. 1–25.

Santos, Filipe M and Kathleen M Eisenhardt. 2005. "Organizational boundaries and theories of organization," *Organization science*, Vol. 16, No. 5, pp. 491–508.

Schoar, Antoinette. 2010. "The divide between subsistence and transformational entrepreneurship," *Innovation Policy and the Economy*, Vol. 10, No. 1, pp. 57–81.

Schumpeter, Joseph. 1942. *Capitalism, Socialism and Democracy*. New York: Harper & Brothers.

Shane, Scott and Sankaran Venkataraman. 2000. "The promise of entrepreneurship as a field of research," *Academy of Management Review*, Vol. 25, No. 1, pp. 217–226.

Smith, Matthew, Danny Yagan, Owen Zidar, and Eric Zwick. 2019. "Capitalists in the twenty-first century," *The Quarterly Journal of Economics*, Vol. 134, No. 4, pp. 1675–1745.

Suarez, Fernando F and James M Utterback. 1995. "Dominant designs and the survival of firms," *Strategic Management Journal*, Vol. 16, No. 6, pp. 415–430.

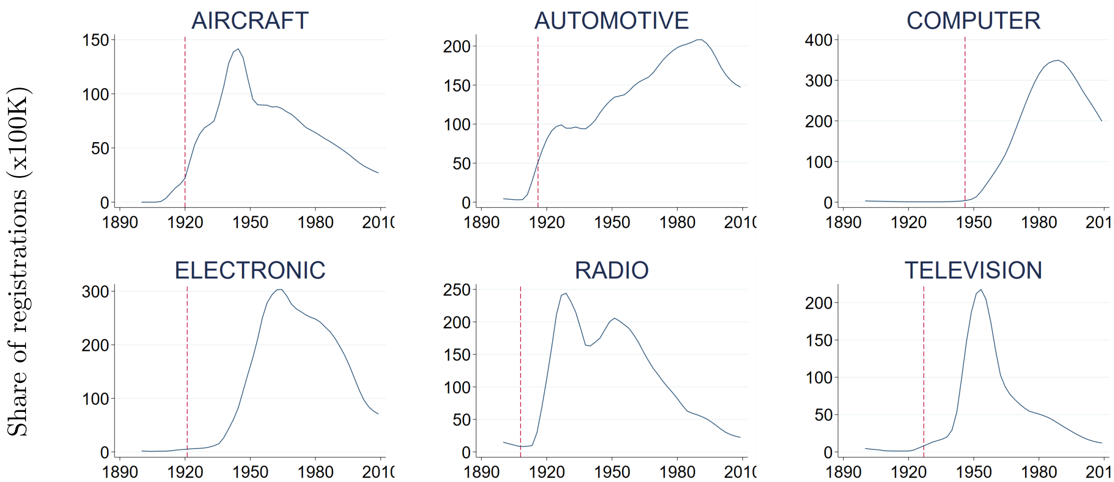
Teece, David J. 1986. "Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy," *Research Policy*, Vol. 15, No. 6, pp. 285–305.

Tushman, Michael L and Philip Anderson. 1986. "Technological discontinuities and organizational environments," *Administrative Science Quarterly*, pp. 439–465.

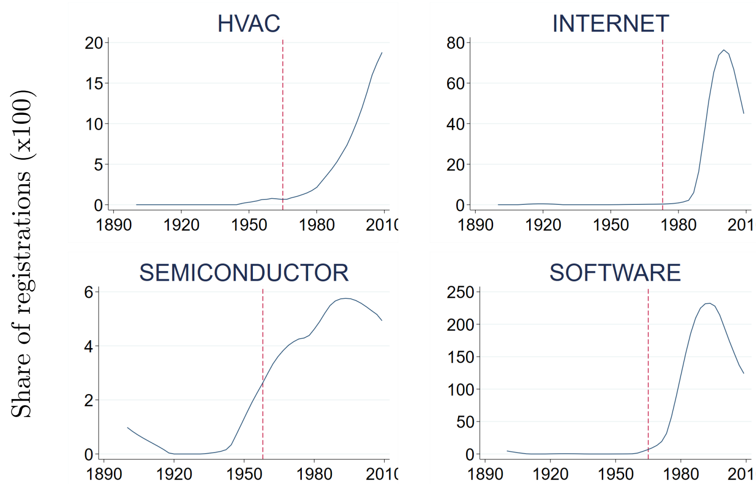
Weitzman, Martin L. 1998. "Recombinant growth," *Quarterly Journal of Economics*, Vol. 113, No. 2, pp. 331–360.

Figure 1: Diffusion of example technologies into new firm names

Panel (A): Early technologies



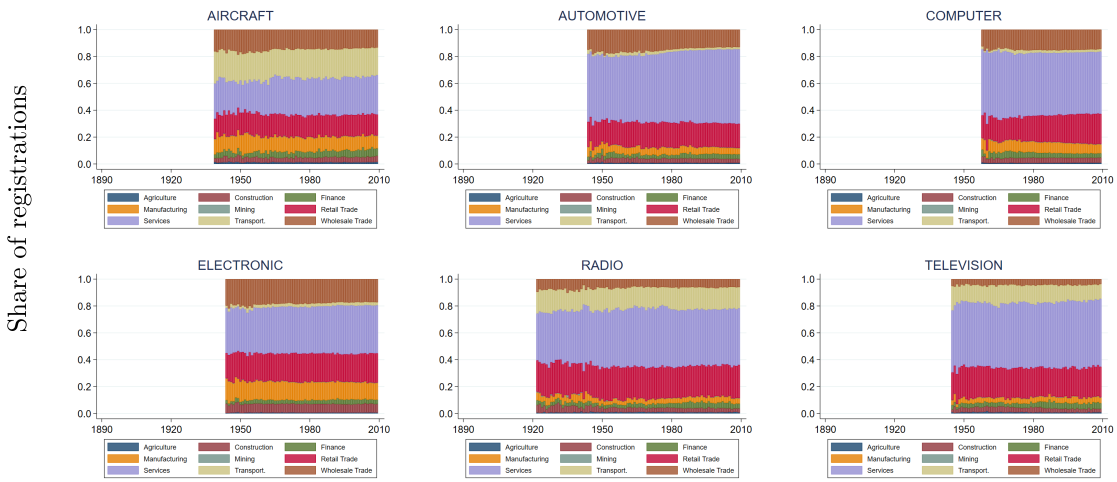
Panel (B): More recent technologies



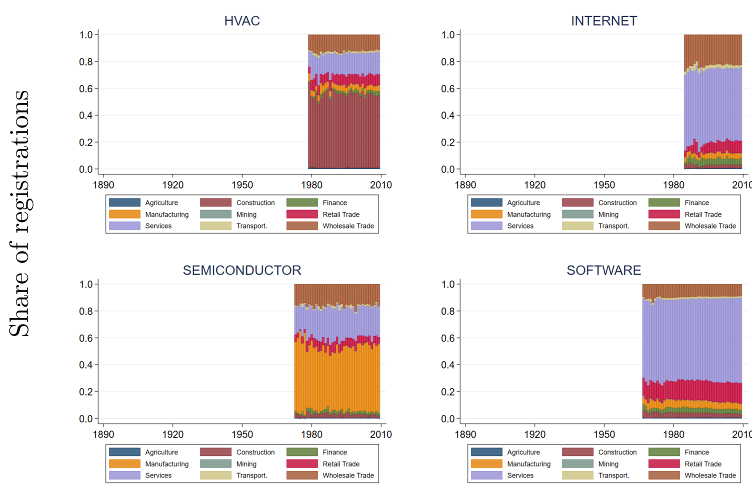
Notes: Figure illustrates entrepreneurial lifecycles for individual technologies, as measured via firm names, for technologies developed before 1950 (Panel A) and after (Panel B). Vertical axis measures the technology's share of business registrations in a given year. The dotted line in each subfigure marks the technology's emergence year, defined as the earliest year after which the technology had at least 10 cumulative firm registrations and ≥10% average growth in registrations over the next five years. Lifecycles plotted using local polynomial smoothing. See text for further details.

Figure 2: Sector composition of new firms created around example technologies

Panel (A): Early technologies

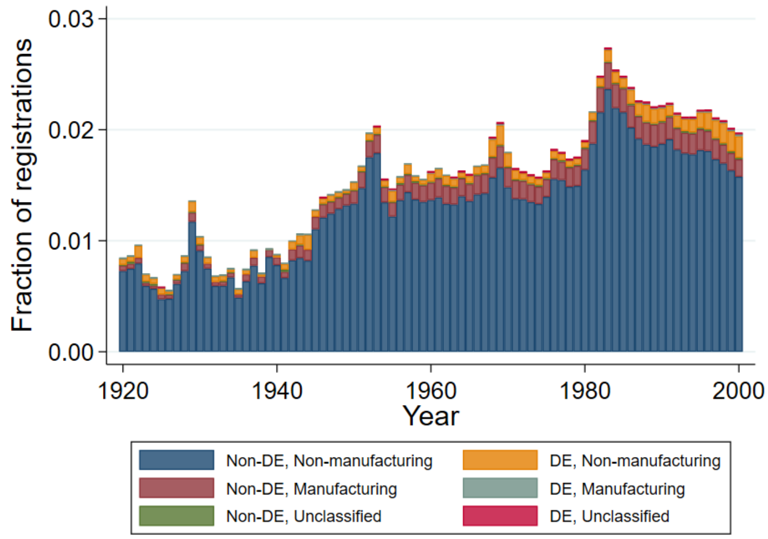


Panel (B): More recent technologies



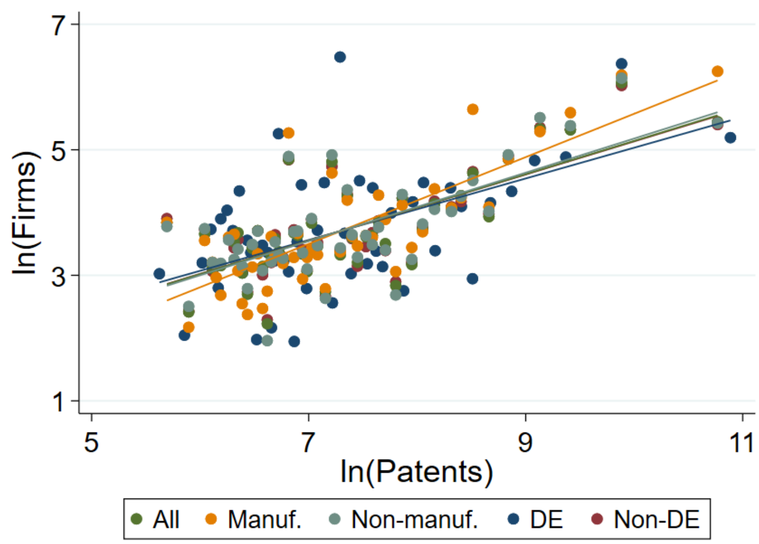
Notes: Figure shows sectoral composition of new business registrations associated with specific technologies, over time, for technologies developed before 1950 (Panel A) and after (Panel B). Each subfigure restricts to technology-years with at least 10 associated registrations, and for presentation purposes restricts to years for which a continuous time series can be constructed without gaps (i.e., the series begins in the year after which the technology always has at least 10 associated firms).

Figure 3: Focal technology-related firm share of U.S. business registrations, 1900-2000



Notes: Figure plots the share of U.S. business registrations over time related to the technologies in our sample, as measured via firm names. We separately report firms registered in Delaware (DE) versus other jurisdictions, and which we measure as manufacturing-related versus in other sectors. See text for further discussion and interpretation of these measures.

Figure 4: Firm creation and patenting in individual technologies



Notes: Figure shows a binned scatterplot of the number of patents (horizontal axis) and number of associated business registrations (vertical axis), as measured via firm names. Scatterplot shown by firm type (all, manufacturing, non-manufacturing, Delaware (DE) registered, non-DE registered) and residualized by invention decade.

Figure 5: Two entrepreneurial paths to profiting from an innovation

Technology-Enabled Firms	**Ecosystem Innovators**
Often includes: local firms, small businesses	*Often includes: high-growth, innovation-driven startups*
Primary Activity	**Primary Activity**
Arms-length participation in a innovation's value chain, often in a local or specialized niche.	Developing innovation and orchestrating its value chain.
Timing of Opportunity	**Timing of Opportunity**
Technology is sufficiently mature for production and use.	A technology's dominant design has yet to emerge.
Potential Drivers of Long-Term Advantage	**Potential Drivers of Long-Term Advantage**
Price: Location and local market power Niche specialization in complementary activity Exclusive relationship with the innovator	*Price:* Ownership of unique intellectual property Distinctive R&D and manufacturing capabilities Bargaining power over downstream partners
Cost: Focused, efficient activity system Specialized/local market knowledge Local relationships and reputation	*Cost:* Efficiencies in production/manufacturing Economies of scale and scope Bargaining power against suppliers
Potential Challenges	**Potential Challenges**
Can be difficult to capture value, due to limited IP, low entry barriers, labor intensity, or difficulty scaling.	Dependence on tech-enabled ecosystem due to reach, requisite local market knowledge, incongruence.
Example Industries	**Example Businesses**
Print & copy centers (technology rental) Auto dealers, mechanics (distribution, maintenance) Heating & cooling services (installation, repair)	Xerox, Canon, Konica Minolta, etc. (copy machines) Ford, Volkswagen, etc. (automobiles) Carrier, Trane, etc. (HVAC equipment)

Figure 6: Extensions to existing frameworks & new insights

High-tech Industries

Extensions to the industry studies literature:

- Opportunities beyond the core mfg. sector
- Evolution of opportunity over the lifecycle

Key insight:
Entrepreneurial opportunity persists and evolves in the later stages of the technology lifecycle

Technology-Enabled Firms

Extensions to entrepreneurial strategy:

- Technology-first entrepreneurial strategy: choose technology *and then your activity*
- Activity-first entrepreneurial strategy: choose your activity, then your technology

Key insight:
"Choose your technology" entrepreneurial strategies are not only for innovator but also others in the value chain

Ecosystem Innovators

Extensions to ecosystems strategy:

- Arms-length ecosystems: the economy
- Dynamics of ecosystems: specific types of complementors become more or less important over the technology lifecycle

Key insight:
Innovators' ecosystem strategy must be dynamic and prepared to adapt across technology lifecycles

Table 1: Top technology from each decade of the 20th century, and associated characteristics

		Total (to 2009)				Most common	
Decade	Top technology	Firms	Patents	Mfg. share	Other share	paired words	Example firms
1900	WIRELESS	28732	10565	6%	94%	COMMUNICATIONS SERVICES SOLUTIONS	VANGUARD WIRELESS, INC CYVO WIRELESS LLC WIRELESS BUILDERS, LLC
1910	AUTOMOTIVE	70954	2438	5%	95%	SERVICE SERVICES REPAIR	BNA AUTOMOTIVE GROUP, INC. GALYEON AUTOMOTIVE, LLC W.A.M. AUTOMOTIVE, INC.
1920	TELEVISION	12642	7598	4%	96%	CABLE SERVICE RADIO	GABLES TELEVISION, INC. BUSINESS TELEVISION, INC. U.S. PANTS & TELEVISION, INC.
1930	VIDEO	52958	26315	4%	96%	PRODUCTIONS AUDIO SERVICES	LEGAL VIDEO SERVICES, INC. COBB VIDEO, INC. VIDEO EMPORIUM, INC.
1940	SEMICONDUCTOR	1977	88544	48%	52%	EQUIPMENT TECHNOLOGY TECHNOLOGIES	PHILIPS SEMICONDUCTORS INC. SAMSUNG SEMICONDUCTOR, INC. IDEAL SEMICONDUCTOR, INC.
1950	BROADBAND	3643	909	23%	77%	SERVICES COMMUNICATIONS WIRELESS	N2 BROADBAND, INC. ZTARK BROADBAND CORPORATION BROADBAND PROPERTIES, LLC
1960	PAGER	574	565	8%	92%	CELLULAR PLUS REPAIR	PAGERS TO GO, INC. PAGERS PLUS INC. LET'S TALK PAGERS, INC.
1970	SOFTWARE	63181	6198	5%	95%	SOLUTIONS SYSTEMS SERVICES	MINI B SOFTWARE, INC. DENIZEN SOFTWARE, LLC BEST SOFTWARE OF CALIFORNIA
1980	ROBOTIC	1632	756	39%	61%	SYSTEMS AUTOMATION TECHNOLOGY	MAZOR ROBOTICS INC. INTERNATIONAL ROBOTICS INC. FUTURE ROBOTICS, INC.
1990	INTERNET	17268	2387	4%	96%	SERVICES SOLUTIONS MARKETING	DC INTERNET GROUP, LLC SAEROM INTERNET WORLD II INC. WORLDPORT INTERNET, INC.

Notes: Table lists the technology from each decade of the 20th century with the most firms created, alongside (i) the number of associated new firms and patents, (iii) the sectoral composition of those firms, (iv) the three most common paired words, and (v) three example firms.

Table 2: Summary statistics at the technology level

Variable	N	Mean	P10	P25	P50	P75	P90	Max
Patents	860	2919	151	281	909	1653	4915	161347
Firms	658	1040	2	3	19	64	673	98530
Share Delaware jurisdiction	658	0.09	0.00	0.00	0.02	0.08	0.25	1.00
Share non-Delaware jurisdiction	658	0.91	0.75	0.87	0.98	1.00	1.00	1.00
Share manufacturing	625	0.17	0.04	0.06	0.13	0.19	0.33	0.91
Share non-manufacturing	625	0.83	0.67	0.76	0.87	0.92	0.96	1.00

Notes: Table shows summary statistics for technologies in our sample (each measured through 2009).

Table 3: Example technologies' maximum share of business registrations (x1000)

Panel A. Early technologies (pre-1950)

Variable	Max share
AIRCRAFT	3.3
AUTOMOTIVE	2.3
COMPUTER	5.8
ELECTRONIC	4.0
RADIO	4.6
TELEVISION	4.7

Panel B. More recent technologies

Variable	Max share
HVAC	0.2
INTERNET	1.7
SEMICONDUCTOR	0.1
SOFTWARE	2.8

Notes: Table lists ten example technologies and their peak share of new business registrations in our data.

Table 4: Dispersion of entrepreneurial activity across the technology lifecycle

		Activity			Sector	
	(1)	(2)	(3)	(4)	(5)	(6)
	Top 1	HHI	C.V.	Top 1	HHI	C.V.
11-20 years into lifecycle	-0.058	-0.059	-0.112	-0.011	-0.011	-0.025
	(0.014)	(0.016)	(0.030)	(0.005)	(0.005)	(0.012)
21-30 years into lifecycle	-0.075	-0.072	-0.143	-0.026	-0.023	-0.059
	(0.019)	(0.023)	(0.043)	(0.006)	(0.005)	(0.013)
31-40 years into lifecycle	-0.108	-0.110	-0.213	-0.031	-0.030	-0.068
	(0.025)	(0.031)	(0.058)	(0.007)	(0.006)	(0.015)
41-50 years into lifecycle	-0.135	-0.140	-0.273	-0.035	-0.032	-0.078
	(0.032)	(0.039)	(0.073)	(0.008)	(0.007)	(0.018)
51-60 years into lifecycle	-0.174	-0.183	-0.355	-0.038	-0.033	-0.086
	(0.038)	(0.046)	(0.087)	(0.009)	(0.008)	(0.021)
61+ years into lifecycle	-0.215	-0.242	-0.459	-0.044	-0.036	-0.096
	(0.048)	(0.058)	(0.111)	(0.011)	(0.009)	(0.025)
N	4348	4348	4348	12308	12308	12308
R^2	0.32	0.39	0.39	0.54	0.51	0.56
Token FEs	Y	Y	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y	Y	Y
Y mean	0.88	0.84	1.96	0.58	0.48	0.73

Notes: Table estimates changes in the dispersion of firm creation across economic sectors or the value chain over time. Dispersion measures are (i) the top activity/sector share ("Top 1"; Columns 1 and 4), the activity/sector HHI ("HHI"; Columns 2 and 5), and the coefficient of variation ("C.V."—in essence, the standard deviation of activity/sector shares; Columns 3 and 6). SEs clustered by technology in parentheses.

Table 5: Changing nature of entrepreneurial entry across the technology lifecycle

Panel A. Economic sectors

	Number of firms into lifecycle					
	(1)	(2)	(3)	(4)	(5)	(6)
Sector	1-100	101-200	201-500	501-1000	1001-2000	2001+
Manufacturing	1.00	0.94	0.83	0.75	0.68	0.39
Services	1.00	1.00	1.01	1.02	1.05	1.25
All other	1.00	1.02	1.07	1.09	1.09	1.06

Panel B. Value chain activities

	Number of firms into lifecycle					
	(1)	(2)	(3)	(4)	(5)	(6)
Activity	1-100	101-200	201-500	501-1000	1001-2000	2001+
Research	1.00	0.64	0.37	0.42	0.36	0.19
Manufacturing	1.00	0.92	0.73	0.73	0.51	0.27
Sales	1.00	1.14	1.39	1.88	1.68	1.27
Distribution	1.00	1.22	1.09	1.65	1.35	1.72
Service	1.00	1.36	1.69	1.80	1.68	1.99

Panel C. Specific paired words

	Number of firms into lifecycle					
	(1)	(2)	(3)	(4)	(5)	(6)
Paired word	1-100	101-200	201-500	501-1000	1001-2000	2001+
Industry	1.00	0.80	0.50	0.53	0.46	0.27
Manufacturing	1.00	0.96	0.73	0.72	0.53	0.27
Engineering	1.00	0.96	0.68	0.53	0.58	0.47
Research	1.00	0.63	0.37	0.42	0.36	0.19
Development	1.00	0.58	0.55	0.63	0.57	0.68
Product	1.00	0.65	0.64	0.50	0.52	0.36
Maintenance	1.00	0.90	1.21	0.94	1.11	1.67
Part	1.00	2.55	3.58	5.03	8.07	6.13
Service	1.00	1.30	1.57	1.59	1.55	1.82
Equipment	1.00	1.11	1.69	1.62	1.24	0.80
Supply	1.00	1.07	1.29	1.73	1.64	1.01

Notes: Table evaluates changes in the relative frequency of firm creation by economic sector (Panel A), activity (Panel B), and specific paired words in firm names (Panel C). To produce this table, we identify firms in the first 1-100, 101-200, etc. firms of a given technology's lifecycle. We then aggregate these to measure the share of all "first 100" firms which are associated with each sector or activity, as measured via firm names. To ease interpretation we then index these shares to that of the first 100 firms, which is therefore always equal to one and provides a reference point to compare against as firm creation changes over time.

Table 6: Upstream to downstream industry evolution

	Activity		Sector	
	(1)	(2)	(3)	(4)
	Manuf.	Services	Manuf.	Services
11-20 years into lifecycle	-0.026	0.043	-0.010	0.009
	(0.032)	(0.044)	(0.005)	(0.005)
21-30 years into lifecycle	-0.083	0.122	-0.013	0.024
	(0.035)	(0.047)	(0.005)	(0.006)
31-40 years into lifecycle	-0.094	0.112	-0.022	0.027
	(0.035)	(0.052)	(0.006)	(0.007)
41-50 years into lifecycle	-0.122	0.166	-0.028	0.035
	(0.035)	(0.057)	(0.007)	(0.008)
51-60 years into lifecycle	-0.137	0.149	-0.028	0.038
	(0.039)	(0.064)	(0.009)	(0.010)
61+ years into lifecycle	-0.154	0.166	-0.034	0.048
	(0.041)	(0.069)	(0.010)	(0.012)
N	4348	4348	12308	12308
R^2	0.33	0.44	0.76	0.69
Token FEs	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y
Y mean	0.07	0.69	0.19	0.35

Notes: Table estimates changes in the share of firm creation across value chain activities and economic sectors, comparing manufacturing and services. SEs clustered by technology in parentheses.

Table 7: Firm creation and the dispersion of activity in relation to innovation characteristics

Panel A: Level of firm creation

		By form		By sector	
	(1)	(2)	(3)	(4)	(5)
	All	DE	Non-DE	Manuf.	Other
Std(Max value of first 100 patents)	0.004	0.037	0.010	0.030	0.045
	(0.086)	(0.085)	(0.092)	(0.094)	(0.089)
Std(Max breakthru in first 100 patents)	0.374	0.290	0.374	0.325	0.344
	(0.129)	(0.161)	(0.129)	(0.133)	(0.138)
Std(Combinations in first 100 patents)	0.250	0.233	0.252	0.356	0.273
	(0.097)	(0.104)	(0.097)	(0.111)	(0.102)
N	615	337	611	583	584
R^2	0.09	0.10	0.09	0.12	0.09
Decade FEs	Y	Y	Y	Y	Y

Panel B: Dispersion of firm creation

		Activity			Sector	
	(1)	(2)	(3)	(4)	(5)	(6)
	Top 1	HHI	C.V.	Top 1	HHI	C.V.
Std(Max value of first 100 patents)	-0.000	0.005	0.011	0.007	0.006	0.018
	(0.014)	(0.016)	(0.029)	(0.007)	(0.007)	(0.016)
Std(Max breakthru in first 100 patents)	-0.012	-0.019	-0.033	-0.007	-0.005	-0.008
	(0.012)	(0.014)	(0.026)	(0.006)	(0.005)	(0.013)
Std(Combinations in first 100 patents)	-0.028	-0.036	-0.068	-0.007	-0.007	-0.019
	(0.012)	(0.014)	(0.027)	(0.005)	(0.004)	(0.012)
N	337	337	337	584	584	584
R^2	0.10	0.11	0.11	0.02	0.03	0.03
Decade FEs	Y	Y	Y	Y	Y	Y

Notes: Panel (A) estimates the relationship between total firm creation in a given technology and characteristics of its first 100 patents. Column (1) measures all firms; Column (2) and (3), manufacturing and non-manufacturing firms; and Columns (4) and (5), Delaware jurisdiction vs. other firms. Panel (B) estimates differences in the dispersion of firm creation in a given technology as a function of the characteristics of its first 100 patents. Dispersion measures are (i) the top activity/sector share ("Top 1"; Columns 1 and 4), the activity/sector HHI ("HHI"; Columns 2 and 5), and the coefficient of variation ("C.V."—in essence, the standard deviation of activity/sector shares; Columns 3 and 6). Robust SEs in parentheses.

Table 8: Annual firm creation as a function of current and recent patenting

Panel A: Current patenting

		By form		By sector	
	(1)	(2)	(3)	(4)	(5)
	All	DE	Non-DE	Manuf.	Other
Ln(Patents)_t	0.290	0.229	0.296	0.320	0.301
	(0.039)	(0.051)	(0.040)	(0.041)	(0.041)
N	12153	3297	11793	11153	11231
R^2	0.82	0.71	0.83	0.76	0.83
Token FEs	Y	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y	Y
Y mean	1.61	1.21	1.59	-0.39	1.44

Panel B: Recent patenting

		By form		By sector	
	(1)	(2)	(3)	(4)	(5)
	All	DE	Non-DE	Manuf.	Other
Ln(Patents)_{t-1}	0.214	0.182	0.217	0.267	0.233
	(0.033)	(0.066)	(0.033)	(0.038)	(0.035)
Ln(Patents)_{t-2}	0.092	0.045	0.088	0.068	0.084
	(0.021)	(0.054)	(0.021)	(0.029)	(0.022)
Ln(Patents)_{t-3}	-0.001	0.049	-0.007	0.024	-0.007
	(0.019)	(0.047)	(0.019)	(0.031)	(0.021)
Ln(Patents)_{t-4}	0.004	-0.029	0.016	-0.028	0.011
	(0.026)	(0.056)	(0.025)	(0.035)	(0.027)
N	11251	3125	10927	10366	10419
R^2	0.84	0.73	0.84	0.77	0.85
Token FEs	Y	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y	Y
Y mean	1.65	1.23	1.63	-0.33	1.47

Notes: Panel (A) estimates relationship between annual firm creation and patenting in a given technology, and Panel (B) the relationship between annual firm creation and recent patenting. Observations in both panels are at the technology-year level. Column (1) measures all firms; Column (2) and (3), manufacturing and non-manufacturing firms; and Columns (4) and (5), Delaware jurisdiction vs. other firms. SEs clustered by technology in parentheses.

Online Appendix

Beyond Innovation: The Composition and Dynamics of Technology-Enabled Entrepreneurship

Innessa Colaiacovo, Daniel P. Gross, and Jorge A. Guzman

A Sector Classification Method

In this appendix, we describe our sector classification procedure in detail, and provide supplementary validation results to those in the body of the paper. Portions of this section may reproduce passages in the paper, adding detail and context. We begin as follows.

Let i index firms, j tokens, s sectors. Each firm i has J_i tokens in its name, and there are J unique tokens in the corpus. Let τ_{js} represent a token-sector score, which will be defined below as the fraction of token j's total occurrences that are in each sector s. φ_{is} will represent a firm-sector score, which will be an average of τ_{js} across all J_i tokens in firm i's name. τ_{js} and φ_{is} will thus measure each token or firm's association with each of $S = 10$ sectors. To aid the exposition below, let n_{js} represent the number of times token j is in the name of firms in sector s, and $N_J = \sum_{s=1}^{10} n_{js}$ represent the number of firms in which token j appears overall.

A.1 Measuring token-sector scores

We first undertake several steps to prepare the D&B data for training and validating our classification procedure. We begin by cleaning firm names of stop words, special characters, tokens of length $\leq$2, and all-numeric tokens. We then split out the tokens in firm names, and reshape the firm-level data to a token-level dataset, with one observation per firm-token. Because the D&B company name field is fixed in width, words in firm names are sometimes abbreviated. We thus extend our data cleaning effort by creating an extensive, part-manual and part-automated crosswalk from tokens which appear to be abbreviations or minor misspellings to full words. A total of roughly 65,000 abbreviated tokens get crosswalked to full words through this approach. For example, "IMPRVMNT" is updated to "IMPROVEMENT", and "PIZZARIA" to "PIZZERIA". We believe these changes ultimately increase the fidelity of our data to the underlying economic reality and will improve the quality of our sector classification. A total of 65,000 unique tokens in the D&B 1990 sample are revised through these methods, out of roughly 10 times as many tokens in the data. The three most common words in this sample are "SERVICE", "CONSTRUCTION", and "SHOP"—words which are (seemingly) immediately revealing of the sector that a firm with those words is in, and illustrate the face validity of our proposed approach.

We take two approaches to further reducing (or not) the set of tokens which we will use to predict firms' economic sector. In the first variant, we deploy this set of tokens as is. In our second variant, we remove additional stop words, excluding those which are *not* an English-language word but are any of (i) a word in a U.S. state name, (ii) a word in a U.S. city name, (iii) a common given name, or (iv) a common surname.[1] We exclude these tokens because we think they are both frequently occurring in firm names, due to geographic specialization or eponymy, and because they are sector-agnostic and thus may obscure our sectoral predictions. The resulting dataset has firm-tokens and their associated sector, obtained from each firm's D&B-reported SIC code.

We then measure τ_{js} as sector s's share of all uses of token j: $\tau_{js} = \frac{n_{js}}{N_J}$. Although we experimented with other approaches to scoring tokens' associations with each sector, we found that this

[1]Our sample of given names consists of the union of the 1,000 most common baby names for birth years 1880 to 2009, according to the Social Security Administration, obtained from https://github.com/hadley/data-baby-names. Our sample of surnames consists of all surnames which appeared $\geq$100 times in the 2010 census, obtained from https://www.census.gov/topics/population/genealogy/data/2010_surnames.html. In addition to excluding these names, we also exclude variants of these names plus the letter 'S', to account for the sizable number of businesses named in the possessive (e.g., "MIKE'S FOOD & SPIRITS" or "DAVE'S FRESH PASTA").

simple approach outperformed others across many validation tests.[2] By construction, each token's sector-level scores τ_{js} will add to 1 when summed across sectors.

A.2 Measuring firm-sector scores

To get from token-sector scores to firm-sector scores (our target output), we need to aggregate across tokens in firm names—reducing the dimensionality of each firm i from $J_i \times 10$ to 10. We experimented with two approaches: straight averages and weighted averages, weighting by (the square root of) a token's total usage. This latter approach will overweight more-common words, whose sector associations we might measure more precisely. Formally, these two approaches to measuring the firm-sector score φ_{is} can be characterized as follows:

Approach 1: $$\varphi_{is}^{uwtd} = \frac{1}{J_i} \sum_{j=1}^{J_i} \tau_{js}$$

Approach 2: $$\varphi_{is}^{wtd} = \frac{1}{\sum_{j=1}^{J_i} \sqrt{N_j}} \sum_{j=1}^{J_i} \sqrt{N_j} \tau_{js}$$

To illustrate, suppose there are two tokens (A and B) and two sectors (1 and 2). To make this example concrete, let us specifically imagine a firm with the name "Master Plumbing", and that there are two sectors, *Construction* and *Retail Trade*. We don't ex-ante know what sector the firm is in. It might be a construction contractor. It might be a plumbing fixture store. But the words in its name give us clues—especially when we can gauge how often these words associate with each of these sectors in our training data. For this example, let us denote "Master" and "Plumbing" as tokens A and B, and *Construction* and *Retail Trade* as sectors 1 and 2.

Suppose that in our training sample, token A ("Master") is used in 225 firm names, and appears 50% of the time in sector 1 (*Construction*) and 50% in sector 2 (*Retail Trade*). Token B ("Plumbing"), on the other hand, is in 25 firm names, and appears 90% of the time in sector 1 and 10% in sector 2. Table A.1 provides the computed sector scores under each of these two approaches. In practice, we find that the simpler aggregation method (unweighted averages) better predicts the true distribution, which makes this our preferred approach.

Table A.1: Firm-sector scores for "Master Plumbing" (contrived example)

	Sector 1	Sector 2
Approach 1: Unweighted	$\varphi_{i1} = \frac{0.5+0.9}{2} = 0.7$	$\varphi_{i2} = \frac{0.5+0.1}{2} = 0.3$
Approach 2: Weighted	$\varphi_{i1} = \frac{\sqrt{225}\cdot 0.5+\sqrt{25}\cdot 0.9}{\sqrt{225}+\sqrt{25}} = 0.6$	$\varphi_{i2} = \frac{\sqrt{225}\cdot 0.5+\sqrt{25}\cdot 0.1}{\sqrt{225}+\sqrt{25}} = 0.4$

For a real example from our data, consider the firm "Anderson Home Appliances", whose words can be fractionally classified across each of the original 10 economic sectors. In our data, these words associate with sectors as shown in Table A.2 (where only the top five sectors are shown), leading to the firm-level scores shown in the last row of the table. Our procedure predicts *Retail Trade* and *Services* as equally likely sectors. The actual primary sector reported for this firm in

[2] An alternative approach we explored was to square these shares, to overweight high token-sector associations. Doing so, however, reduced the rate at which we correctly predicted firms' actual sector.

the D&B data is *Services*, and the secondary sector reported in D&B is *Retail Trade*—though we might infer that Anderson Home Appliances is related to *Construction* too.

Table A.2: Firm-sector scores for "Anderson Home Appliances" (true example)

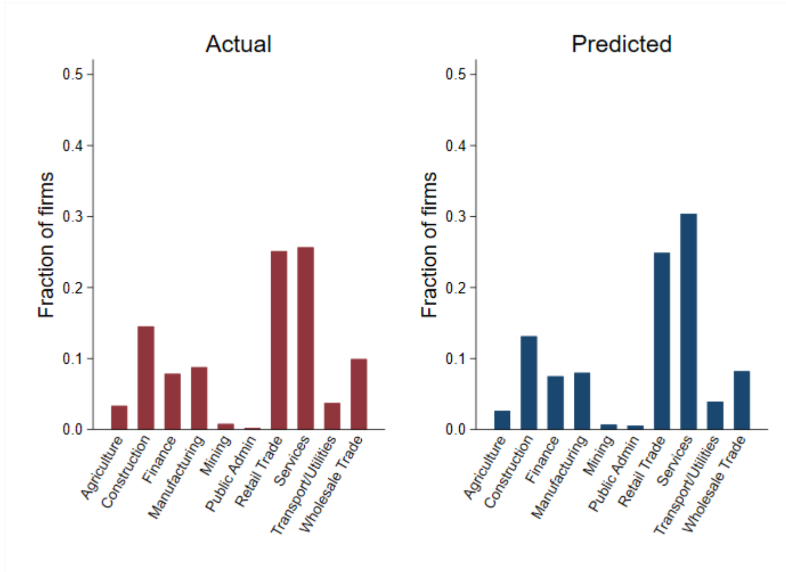
	Token-sector scores and cross-token averages				
Word	Agriculture	Construction	Fin./Ins./RE	Retail Trade	Services
Anderson	8%	17%	10%	14%	34%
Home	0%	20%	10%	17%	48%
Appliances	0%	2%	0%	71%	21%
Average	3%	13%	7%	34%	34%

A.3 Validating our procedure

We validate this sector classification procedure in several ways. We begin by validating within our D&B test sample, where we can classify firms into sectors and compare our predictions to what is reported in the data. We focus our validation evidence on our more restrictive variant, where we remove place and person names from firm names before classifying sectors.

Figure A.1 begins with a high level view, aggregating sector shares across the sample and showing the overall frequency of firms in each of our 10 sectors as observed (left panel) and as predicted (right panel). The distributions line up closely, though we slightly overpredict *Services* and slightly underpredict *Wholesale Trade*. Table A.3 provides a more nuanced view of our predictive validity, showing, for each sector (first column), how often firms predicted to be in that sector are in fact reported by D&B to be in that sector (second column), as well as how often in each such case our second-most likely prediction is the D&B-reported sector (third column). The final column shows the sum: the probability that one of the top two predicted sectors is the reported sector. Across all predicted sectors, we match D&B 70 to 90% of the time with our top one or two predictions. We consider these rates to be high, particularly as we do not expect to match D&B 100% of the time, given that many firms span boundaries between sectors.

Figure A.1: Predicted vs. actual distribution of firms across sectors in D&B testing sample



Notes: Figure shows the predicted and actual distribution of sectors in our D&B test sample, using the sector classification algorithm described in the paper.

Table A.3: Sector classification algorithm: Predictive performance, by sector

	Fraction of firms where:		
Predicted sector	Prediction correct	Runner-up prediction correct	Total
Agriculture, Forestry, & Fishing	78%	9%	87%
Construction	84%	7%	91%
Finance, Insurance, & Real Estate	82%	8%	90%
Manufacturing	67%	15%	82%
Mining	58%	13%	71%
Public Administration	81%	9%	91%
Retail Trade	73%	13%	86%
Services	66%	16%	83%
Transportation & Public Utilities	84%	6%	91%
Wholesale Trade	57%	21%	78%
Overall	71%	14%	85%

Notes: Table shows, for each sector (first column), how often firms predicted to be in that sector (i.e., whose top predicted sector is that shown in the left column) are in fact reported by D&B to be in that sector (second column), as well as how often in each such case our second-most likely prediction is the D&B-reported sector (third column).

In Table A.4, we break these patterns down even further. Here we document, for each top-predicted sector (left column), the fraction of firms with that prediction that are reported by D&B in each of our 10 sectors. The first features of this table to observe is that the mass is overwhelmingly concentrated along the diagonal (which reproduces Column 2 of Table A.3). In addition, the table also illustrates the sectors that are likely to be jointly present—or incorrectly classified, depending on interpretation. Firms which we predict to be in *Wholesale Trade* (row 10), for example, are 18% of the time actually reported as being in *Retail Trade*—illustrating that firms may sometimes be a bit of each sector, and the occasional challenging of distinguishing the two.

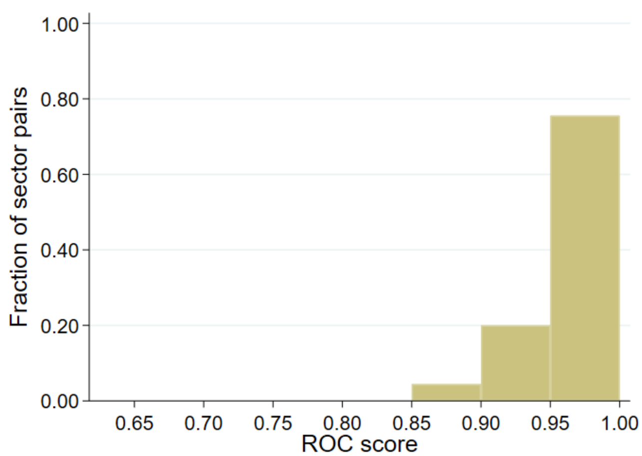
Table A.4: Sector classification algorithm performance: Predicted vs. actual sector

When top predicted sector is...		Pr(Actual sector is...)									
Sector	Description	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1	Agriculture, Forestry, & Fishing	**78%**	2%	2%	1%	0%	0%	7%	5%	1%	4%
2	Construction	1%	**84%**	1%	3%	1%	0%	3%	4%	1%	3%
3	Finance, Insurance, & Real Estate	1%	4%	**82%**	1%	0%	0%	3%	7%	1%	2%
4	Manufacturing	1%	4%	1%	**67%**	0%	0%	7%	9%	1%	11%
5	Mining	1%	4%	4%	4%	**58%**	0%	9%	6%	3%	12%
6	Public Administration	1%	1%	4%	1%	0%	**81%**	2%	6%	3%	2%
7	Retail Trade	2%	2%	2%	5%	0%	0%	**73%**	9%	1%	7%
8	Services	2%	4%	4%	4%	0%	1%	11%	**66%**	2%	5%
9	Transportation & Public Utilities	1%	2%	1%	1%	0%	0%	3%	5%	**84%**	2%
10	Wholesale Trade	2%	3%	1%	8%	2%	0%	18%	7%	2%	**57%**

Notes: Table shows, for each sector, how often firms predicted to be in that sector (i.e., whose top predicted sector is that shown in the left column) are fact reported by D&B to be in each of the ten sectors. The diagonal reproduces the rates from the second column of Table A.3.

As another check, we make all unique pairs of sectors, and for each pair, we filter to firms whose reported sector is one of the two and use receiver-operating characteristic (ROC) analysis to evaluate for how many firms the more probable sector is the reported one. Figure A.2 plots the distribution of ROC scores among these pairs. The scores are near one, indicating that in each of these two-sector horseraces, we nearly always predict the D&B-reported sector.

Figure A.2: Distribution of ROC scores for prediction in sector pairs in D&B testing sample



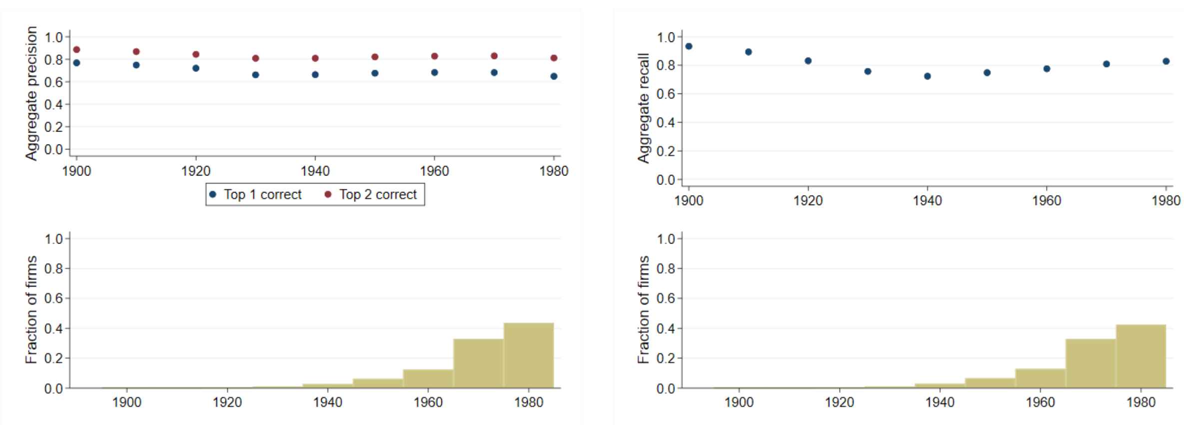
Notes: Figure shows distribution of ROC scores over sector pairs, evaluating across our test sample how often the more-predicted sector in each pair is the firm's reported sector, conditional on one of the two sectors in the pair being the true sector. With 10 sectors, the distribution in this figure is over $\sum_{s=1}^{9} s$ pairs. In the vast majority of these pairings, our algorithm predicts a firm's true sector over the alternative >95% of the time.

In additional validation tests, we examine the stability of this performance across time, for firms with different D&B-reported founding years, and we find that our procedure's performance is stable in both the rate at which its predictions match D&B-reported sectors, and in the fraction of firms that get classified to a sector (Figure A.3). We also perform validation tests against an independent third sample of firms in the Infogroup USA 2000 data extract. The performance of our procedure in this sample is similar to that in our D&B test sample, suggesting that the results above are not driven by distinctive shared features of the D&B training and test samples.

Figure A.3: Precision of sector predictions vs. rate of prediction in D&B testing sample

Panel (A): Pr(True sector is top-predicted)

Panel (B): Fraction of firms classified w/ sector



Notes: The top chart in each panel shows how precision and recall rates vary over time (across firm registration decade) in the D&B testing sample. Bottom panel shows the distribution of firms in this sample across decades. Aggregate precision and recall is a weighted average of those in the upper scatterplots, weighted by the density in the lower histograms.

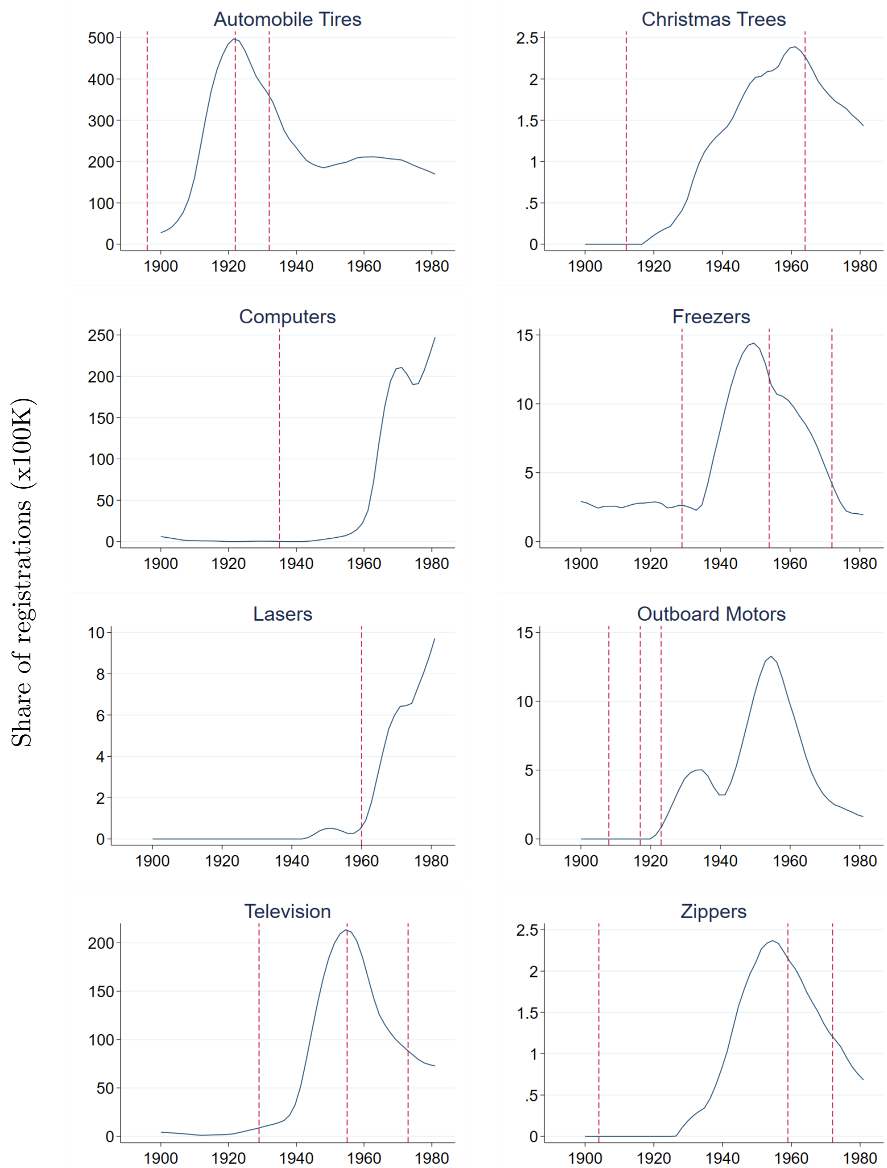
A.4 Classifying the business registration sample

Our final step is to classify firms in our business registration data to economic sectors. We clean firm names by the same approach applied to D&B firms in Section A.1, up to expanding abbreviations and correcting minor misspellings, and again remove place and person names that are not English-language words. We then apply the token-sector scores calculated in Section A.1 to these firms' names, and aggregate to the firm level following Section A.2. The resulting dataset has sector scores for each registered firm, across each of ten economic sectors (*Agriculture* to *Wholesale Trade*). Recall that these scores will add to one, and can alternatively be interpreted as the probability that a firm is in a given sector, or the firm's fractional association with that sector—which allows firms to span sectoral boundaries, as many firms do in practice.

This classification nevertheless bears limitations. One is that some words in the set of firm names in the registration sample may not be present in the Dun & Bradstreet sample, and thus will be unable to be scored. Due to this, as well as the removal of stop words, some firms in the registration data may not be able to have their sector predicted. We are reassured, however, by the fact that even under our more restrictive sector classification procedure, we classify 90% of firms—and this rate is stable over time. We believe this is high enough to generate useful measures of the sectoral distribution of U.S. entrepreneurship over long periods.

B Validating name-based firm measures

Figure B.1: Comparison of SCP-derived entrepreneurial lifecycles to Klepper and Graddy (1990) product lifecycles for select products



Notes: Figure shows the smoothed rate of firm creation (across all sectors) in select products studied by Klepper and Graddy (1990)—who previously identified periods of growth, contraction (shakeout), and stability in the number of manufacturing firms active in each of 46 product categories. We link our business registration data to these products via a manually-curated set of words which we search for in firm names, and focus our analysis here on eight technologies which appear in enough firm names to be able to measure their lifecycles in our data. Each subfigure shows up to three vertical, dashed lines: the first line marks the beginning of Klepper and Graddy Stage 1 (takeoff); if present, the second line marks the start of Stage 2 (shakeout); and if present, the third line marks the start of Stage 3 (stability). Lifecycles plotted using local polynomial smoothing.

Table B.1: SIC industries most heavily associated with specific technologies in D&B sample

	Largest SIC		2nd-largest SIC		3rd-largest SIC	
Technology	Share	Description	Share	Description	Share	Description
AIRCRAFT	16.8%	Airports, flying fields, and services	13.5%	Transportation equipment and supplies	13.1%	Automotive dealers, nec
ALTERNATOR	56.2%	Automotive repair shops, nec	12.1%	Motor vehicle supplies and new parts	10.2%	General automotive repair shops
ATOMIC	7.4%	Labor organizations	3.7%	Eating places	3.1%	Electrical work
AUDIO	21.1%	Radio, television, and electronic stores	8.9%	Radio and television repair	7.0%	Radio And Television Stores
AUTOMOBILE	22.9%	Membership organizations, nec	13.0%	New and used car dealers	10.6%	Used car dealers
AUTOMOTIVE	45.4%	General automotive repair shops	15.4%	Motor vehicle supplies and new parts	7.8%	Auto and home supply stores
BUTANE	71.6%	Liquefied petroleum gas dealers	9.2%	Petroleum products, nec	3.9%	Toys and hobby goods and supplies
CARDIAC	55.3%	Offices and clinics of medical doctors	9.1%	Medical laboratories	6.1%	Specialty outpatient clinics, nec
CATALYST	14.9%	Management consulting services	5.6%	Industrial inorganic chemicals, nec	5.1%	Business consulting, nec
COMPUTER	19.1%	Computer and software stores	13.5%	Computer related services, nec	9.8%	Computer maintenance and repair
COMPUTERIZED	12.9%	Accounting, auditing, and bookkeeping	5.9%	Data processing and preparation	4.9%	Custom computer programming services
DIESEL	60.6%	General automotive repair shops	10.1%	Industrial machinery and equipment	6.1%	Repair services, nec
ELECTRONIC	15.7%	Radio and television repair	15.4%	Electronic parts and equipment, nec	8.8%	Radio, television, and electronic stores
EXTRUSION	17.2%	Aluminum extruded products	12.7%	Plastics products, nec	11.9%	Special dies, tools, jigs, and fixtures
FIBERGLASS	12.5%	Boatbuilding and repairing	10.9%	Repair services, nec	9.0%	Special trade contractors, nec
FUSION	5.0%	Eating places	4.4%	Custom computer programming services	3.7%	Business services, nec
GENETIC	12.2%	Livestock services, except veterinary	11.2%	Medical laboratories	7.7%	Commercial physical research
HELICOPTER	19.9%	Air transportation, nonscheduled	11.3%	Crop planting and protection	9.2%	Air transportation, scheduled
HVAC	79.8%	Plumbing, heating, air-conditioning	6.2%	Warm air heating and air conditioning	2.4%	Refrigeration service and repair
INTERNET	32.6%	Telephone communication, except radio	12.8%	Data processing and preparation	7.3%	Computer related services, nec
KARAOKE	69.0%	Entertainers and entertainment groups	5.7%	Radio, television, and electronic stores	3.1%	Eating places
LAMINATE	13.2%	Wood partitions and fixtures	6.9%	Special trade contractors, nec	6.4%	Lumber, plywood, and millwork
MACHINING	75.0%	Industrial machinery, nec	2.8%	Special dies, tools, jigs, and fixtures	1.7%	Repair services, nec
MAILBOX	58.6%	Business services, nec	8.3%	Direct mail advertising services	3.6%	Hardware stores
MICRO	10.6%	Computer and software stores	8.9%	Computers, peripherals, and software	6.3%	Computer related services, nec
MICROFILM	40.4%	Business services, nec	29.6%	Business Services, nec	5.5%	Office equipment
MICROWAVE	15.4%	Electronic components, nec	14.3%	Household appliance stores	13.1%	Electrical repair shops
MOTORCYCLE	46.8%	Motorcycle dealers	18.9%	Repair services, nec	9.5%	Automotive dealers, nec
MULTIMEDIA	8.7%	Data processing and preparation	8.4%	Custom computer programming services	7.7%	Motion picture and video production
NUCLEAR	10.6%	Offices and clinics of medical doctors	6.4%	Medical laboratories	3.7%	Services, nec
ONLINE	19.1%	Telephone communication, except radio	8.8%	Data processing and preparation	6.2%	Computer related services, nec
ORTHODONTIC	54.0%	Offices and clinics of dentists	32.1%	Dental laboratories	7.3%	Dental equipment and supplies
PAGER	40.7%	Radiotelephone communication	25.9%	Miscellaneous retail stores, nec	15.9%	Electronic parts and equipment, nec
PAINTBALL	38.6%	Sporting goods and bicycle shops	25.3%	Amusement and recreation, nec	6.2%	Hobby, toy, and game shops
PHONE	25.6%	Miscellaneous retail stores, nec	9.2%	Business services, nec	8.9%	Telephone communication, except radio
PIPELINE	26.5%	Water, sewer, and utility lines	13.2%	Crude petroleum pipelines	12.7%	Natural gas transmission
PLYWOOD	27.5%	Lumber, plywood, and millwork	27.1%	Lumber and other building materials	7.0%	Hardwood veneer and plywood
POLYMER	13.4%	Plastics materials and resins	11.2%	Plastics materials and basic shapes	8.1%	Plastics products, nec
PROCESSOR	11.0%	Meat packing plants	6.3%	Scrap and waste materials	3.6%	Livestock services, except veterinary
PROPANE	75.3%	Liquefied petroleum gas dealers	6.9%	Petroleum products, nec	2.8%	Fuel oil dealers
PROSTHETIC	37.0%	Surgical appliances and supplies	30.8%	Miscellaneous retail stores, nec	18.6%	Dental laboratories
QUANTUM	5.1%	Management consulting services	3.6%	Business services, nec	3.5%	Business consulting, nec
RADIO	35.7%	Radio and television repair	17.1%	Radio And Television Stores	17.1%	Radio broadcasting stations
SEMICONDUCTOR	37.6%	Semiconductors and related devices	31.3%	Electronic parts and equipment, nec	2.0%	Engineering services
SNOWMOBILE	62.9%	Automotive dealers, nec	8.1%	Repair services, nec	4.3%	Membership sports and recreation clubs
SOFTWARE	45.8%	Prepackaged software	23.2%	Custom computer programming services	8.9%	Computer and software stores
TELECOMMUNICATION	27.0%	Telephone communication, except radio	11.6%	Electrical work	9.8%	Business consulting, nec
TELEVISION	39.5%	Radio and television repair	18.6%	Radio And Television Stores	8.6%	Television broadcasting stations
TRACTOR	28.4%	Farm and garden machinery	10.0%	Repair services, nec	8.4%	Miscellaneous retail stores, nec
TYPEWRITER	47.9%	Miscellaneous retail stores, nec	27.8%	Repair services, nec	7.5%	Commercial Machines And Equipment (Selling)
URETHANE	24.2%	Plastering, drywall, and insulation	20.3%	Roofing, siding, and sheetmetal work	6.0%	Special trade contractors, nec
VIDEO	43.7%	Video tape rental	9.7%	Motion picture and video production	8.1%	Radio, television, and electronic stores
VINYL	20.3%	Roofing, siding, and sheetmetal work	8.9%	Floor covering stores	7.8%	Top and body repair and paint shops
WINDSHIELD	56.2%	Automotive glass replacement shops	15.4%	Paint, glass, and wallpaper stores	11.0%	Automotive repair shops, nec
WIRELESS	37.3%	Radiotelephone communication	32.4%	Miscellaneous retail stores, nec	6.5%	Telephone communication, except radio
WIRELINE	86.3%	Oil and gas field services, nec	1.4%	Drilling oil and gas wells	1.2%	Water, sewer, and utility lines

Notes: Table lists technologies in our sample with $\geq$500 associated firms in our D&B data, and the 4-digit SIC industries most heavily associated with each. The close relationship between the technology and their firms' principal industries suggests that the presence of a technology in a firm's name generally suggests that the firm's principal line of business is related to that technology.

Table B.2: List of technology terms and related statistics from firm names

Technology	Patents	Firms	Year of 10th firm	Most paired word	2nd most paired word	3rd most paired word
SUBSTRATE	161347	80	1992	TECHNOLOGY	SOLUTION	LASER
SEMICONDUCTOR	88544	1977	1960	TECHNOLOGY	EQUIPMENT	SERVICE
POLYMER	85253	4184	1952	TECHNOLOGY	SYSTEM	PRODUCT
CATALYST	77798	3174	1964	SERVICE	CONSULTING	TECHNOLOGY
TRANSISTOR	76986	75	1971	DEVICE	SERVICE	ALL
SENSOR	73909	1231	1967	TECHNOLOGY	SYSTEM	CONTROL
ROTOR	62808	447	1956	WING	SERVICE	AVIATION
COMPUTER	57295	98530	1948	SERVICE	SYSTEM	SOLUTION
AMPLIFIER	48044	118	1989	CALIFORNIA	TELEVISION	CUSTOM
CAPACITOR	46707	119	1967	TECHNOLOGY	SALE	CALIFORNIA
RESISTOR	38977	50	1982	PRECISION	ELECTRONIC	GLOVER
ELECTRONIC	35422	59984	1923	SERVICE	SYSTEM	COMMUNICATION
COPOLYMER	32494	30	1984	LION	CHEMICAL	DSM
PROCESSOR	32428	2665	1937	FOOD	MEAT	GAS
TRANSDUCER	27111	114	1968	TECHNOLOGY	SYSTEM	RESEARCH
VIDEO	26315	52958	1941	PRODUCTION	AUDIO	SERVICE
DIODE	24390	65	1981	LASER	TECHNOLOGY	HOLDING
AUTOMOBILE	24072	6624	1900	CLUB	SERVICE	DEALER
VINYL	21286	4689	1954	SIDING	PRODUCT	CARPET
RADIO	20932	21350	1910	CLUB	AMATEUR	SERVICE
WAVELENGTH	19665	178	1979	COMMUNICATION	PRODUCTION	TECHNOLOGY
AIRCRAFT	18820	18145	1922	SERVICE	SALE	LEASING
SIDEWALL	18030	19	0	ROOFING	PRODUCT	SPECIALTY
PLANAR	18004	74	1983	SYSTEM	TECHNOLOGY	IPEC
THERMOPLASTIC	17216	166	1966	CAL	SYSTEM	ADVANCED
TRACTOR	16566	11879	1913	SERVICE	EQUIPMENT	TRUCK
ANALOG	15469	384	1973	DIGITAL	DESIGN	SYSTEM
WAVEGUIDE	15196	37	2001	CONSULTING	COMMUNICATION	ENGINEERING
AUDIO	14143	17062	1948	VIDEO	VISUAL	SERVICE
PHOTOSENSITIVE	13246	3	0			
POLYESTER	13150	55	1995	FILM	MITSUBISHI	NESTE
DATABASE	12350	1791	1978	SOLUTION	SYSTEM	SERVICE
PARTICULATE	11167	19	0	CONTROL	SYSTEM	TECHNOLOGY
EPOXY	11094	519	1968	SYSTEM	COATING	FLOORING
CIRCUITRY	10819	54	1987	ADVANCED	BOARD	DESIGN
WIRELESS	10565	28732	1924	COMMUNICATION	SERVICE	SOLUTION
COMPARATOR	10560	10	0	SERVICE	PRODUCT	SYSTEM
MOLECULE	10293	75	1990	DESIGNED	IMMUNO	RESEARCH
PHARMACEUTICALLY	10257	2	0	ELEGANT	FORMERLY	PHARMACEUTIC
PIEZOELECTRIC	10242	2	0	DEVICE	INVESTOR	TECHNOLOGY
FORMALDEHYDE	10189	12	0	FOAM	INSULATION	UREA
COOLANT	9818	74	1989	SERVICE	SYSTEM	MANAGEMENT
ENZYME	9717	225	1965	TECHNOLOGY	RESEARCH	PRODUCT
MICROWAVE	8833	1633	1955	SERVICE	SYSTEM	TECHNOLOGY
ULTRASONIC	8772	401	1964	CLEANING	BLIND	SERVICE
INVERTER	8620	27	1999	POWER	SERVICE	SOLAR
SERVO	8472	231	1967	SYSTEM	TECHNOLOGY	DYNAMIC
REFRACTIVE	8212	264	1982	SURGERY	CENTER	LASER
BINARY	8211	594	1975	SYSTEM	SOLUTION	TECHNOLOGY

Notes: Table lists the 50 top technologies in our sample (by number of patents) and (i) their number of associated firms in our data, (ii) the year of the tenth firm, and (iii) the top three words which co-occur in the names of firms with that technology (excluding stop words).

Table B.3: Example firms for each technology (randomly chosen)

Technology	Patents	Firms	Example 1	Example 2	Example 3
COMPUTER	57295	98530	KEPLER COMPUTER INC.	ST. CROIX COMPUTERS, LLC	COMPUTER DIRECTIONS, INC.
AUTOMOTIVE	2438	70954	GALYEON AUTOMOTIVE, LLC	W.A.M. AUTOMOTIVE, INC.	BNA AUTOMOTIVE GROUP, INC.
SOFTWARE	6198	63181	MINI B SOFTWARE, INC.	DENIZEN SOFTWARE, LLC	BEST SOFTWARE OF CALIFORNIA
ELECTRONIC	35422	59984	DM ELECTRONICS INC.	COLT ELECTRONICS CO.	GEMCO ELECTRONICS, INC.
VIDEO	26315	52958	COBB VIDEO, INC.	VIDEO EMPORIUM, INC.	LEGAL VIDEO SERVICES, INC.
WIRELESS	10565	28732	CYVO WIRELESS LLC	VANGUARD WIRELESS, INC	WIRELESS BUILDERS, LLC
RADIO	20932	21350	RADIO FREE FRISCO LLC	MUSIC SOUND RADIO, INC.	CONTINENTAL RADIO ASSOCIATION
AIRCRAFT	18820	18145	DAMAR AIRCRAFT CORPORATION	LAIRD AIRCRAFT CORPORATION	AIRCRAFT-MARINE PRODUCTS INC
INTERNET	2387	17268	DC INTERNET GROUP, LLC	WORLDPORT INTERNET, INC.	SAEROM INTERNET WORLD II INC.
AUDIO	14143	17062	MARINE AUDIO INCORPORATED	AUDIO INTERNATIONAL, INC.	SAPPHIRE AUDIO & VIDEO, LLC
TELECOMMUNICATION	2701	15645	MJM TELECOMMUNICATIONS, INC.	DB SAT TELECOMMUNICATIONS LLC	K & M TELECOMMUNICATIONS INC.
TELEVISION	7598	12642	GABLES TELEVISION, INC.	BUSINESS TELEVISION, INC.	U.S. PANTS & TELEVISION, INC.
ONLINE	319	12423	REO ONLINE INVESTORS, LLC	SALVAGE AUTOS ONLINE, INC.	SIGHTLINES ONLINE SALES, LLC
MICRO	6560	12377	MICRO FLUIDS, L.L.C.	CENTURY MICRO TEK, INC.	MICRO CONSTRUCTION, INC
TRACTOR	16566	11879	TRACTOR 6943, INC.	KELLEY TRACTOR COMPANY INC	C & W TRACTOR COMPANY, INC.
PIPELINE	5007	11485	SEVEN-A PIPELINE, LLC	JUNIOR'S PIPELINE CO., INC.	INTERCHANGE PIPELINE COMPANY
MOTORCYCLE	2465	8315	MOAK'S MOTORCYCLES, INC.	MOTORCYCLE RENTALS, INC.	NEVADA MOTORCYCLE ADVENTURES
DIESEL	1842	8132	DIESEL DIRECT LLC	DIESEL WORKS, INC.	DIESEL DISPATH INC
PHONE	2883	7345	THE ENTER-PHONE, INC.	CALIFORNIA PHONES LP-16	FRANKY THE PHONE GUY INC
MACHINING	5514	7154	J & W MACHINING, INC.	MOHAWK VALLEY MACHINING, INC.	MACON PRECISION MACHINING LLC
AUTOMOBILE	24072	6624	WESTERN AUTOMOBILE COMPANY	BIG FOUR AUTOMOBILE COMPANY	ILLINOIS AUTOMOBILE CLUB INC
MULTIMEDIA	1351	5562	CREAM MULTIMEDIA INC	MARION MULTIMEDIA INC.	STUDIO H MULTIMEDIA, INC.
VINYL	21286	4689	T & C VINYL, LLC	HANSEN?S VINYL SIDING LLC	VINYL-MASTER INTERNATIONAL
HELICOPTER	1687	4478	T A HELICOPTERS, LLC	HELICOPTER PATROL, INC.	VACA VALLEY HELICOPTERS, INC.
PROPANE	2483	4444	PROPANE CORPORATION	SUBURBAN PROPANE, L.P.	PLANTATION PROPANE, INC.
HVAC	282	4278	AIR-FLO HVAC, INC.	HVAC SOLUTIONS, LLC	AMETECH HVAC CORPORATION
POLYMER	85253	4184	HTI POLYMER, INC.	TICONA POLYMERS, INC.	PORTLAND POLYMERS, LLC
BROADBAND	909	3643	N2 BROADBAND, INC.	BROADBAND PROPERTIES, LLC	ZTARK BROADBAND CORPORATION
ORTHODONTIC	1029	3584	A & A ORTHODONTICS, INC.	JENNINGS ORTHODONTICS, PLLC	R. NIKODEM ORTHODONTIC, INC.
FIBERGLASS	951	3217	H AND H FIBERGLASS, INC.	BLADOW'S FIBERGLASS, INC.	BOSS FIBERGLASS REPAIR LLC
CATALYST	77798	3174	CATALYST PROPERTIES, LLC	CATALYST IT SERVICES, INC.	SOLUTIONS CATALYST GROUP, LLC
PROCESSOR	32428	2665	PET PROCESSORS, LLC	STEEL PROCESSORS, LLC	APPLIED PROCESSORS CORP.
PROSTHETIC	1425	2560	VIVO PROSTHETICS, INC.	CREATIVE PROSTHETICS, INC.	DALE CLARK PROSTHETICS, INC.
CARDIAC	2302	2434	CARDIAC PACEMAKERS, INC.	CARDIAC REHABILITATION, INC.	CARDIAC PATHWAYS CORPORATION
PAINTBALL	197	2428	JP PAINTBALL LLC	NYS PAINTBALL SUPPLY LLC	GLADIATOR PAINTBALL, INC.
GENETIC	482	2225	GENETIC ID, INC.	BRAND GENETICS, LLC	ULTIMATE GENETICS, LLC
COMPUTERIZED	486	2000	COMPUTERIZED SPORTS, INC.	COMPUTERIZED SERVICES, INC.	COMPUTERIZED MARKETING, INC.
SNOWMOBILE	678	1994	ALBANY SNOWMOBILE CLUB, INC.	SUNRISE LAKE SNOWMOBILE CLUB	SKYLINE SNOWMOBILE CLUB, INC.
SEMICONDUCTOR	88544	1977	IDEAL SEMICONDUCTOR, INC.	PHILIPS SEMICONDUCTORS INC.	SAMSUNG SEMICONDUCTOR, INC.
DATABASE	12350	1791	MATERIAL DATABASE CORPORATION	DATABASE DEVELOPERS GROUP LLC	DATABASE RESEARCH CORPORATION
MICROWAVE	8833	1633	CVN MICROWAVE, INC.	ASTERIA MICROWAVE LLC	MID-TEXAS MICROWAVE, INC.
ROBOTIC	756	1632	MAZOR ROBOTICS INC.	FUTURE ROBOTICS, INC.	INTERNATIONAL ROBOTICS INC.
INTEGRATOR	3184	1561	ELITE INTEGRATORS LLC	BAUD SYSTEM INTEGRATORS LLC	TECHNOLOGY INTEGRATORS INC.
ULTRASOUND	2522	1541	ABC ULTRASOUND, LLC	BUFFALO ULTRASOUND, INC.	PYRAMID ULTRASOUND, INC.
PLYWOOD	721	1422	PLYWOOD IMPORTERS, LTD.	CITY PLYWOOD CENTER, INC.	VANCOUVER PLYWOOD CO., INC.
WATERMARK	1043	1321	WATERMARK REALTY, LLC	WATERMARK SIGNAGE, INC.	THE WATERMARK GROUP, INC.
SENSOR	73909	1231	CLOUDTECH SENSORS, INC.	SENSOR TECHNOLOGIES, LLC	INDUSTRIAL SENSORS, INC.
WINDSHIELD	6383	1079	MR. WINDSHIELD, INC.	WINDSHIELD MAN, INC.	B C C WINDSHIELD REPAIR LLC
WEBSITE	265	1071	WEBSITE EXPRESS	CHEETAH WEBSITES INC.	ICON WEBSITE DESIGN LLC

Notes: Table lists the 50 top technologies in our sample (by number of firms) and provides up to three examples of firms with that word in their name. Example firms restricted to those with less than 30 characters in firm name for readability.

C Supplementary Results

The following tables reproduce Tables 4 to 8 of the paper, limiting our technology sample to only nouns. The results are quantitatively and statistically similar.

Table C.1: Dispersion of entrepreneurial activity across the technology lifecycle
robustness check: excluding non-nouns from the technology sample

	Activity			Sector		
	(1)	(2)	(3)	(4)	(5)	(6)
	Top 1	HHI	C.V.	Top 1	HHI	C.V.
11-20 years into lifecycle	-0.059	-0.063	-0.119	-0.012	-0.012	-0.029
	(0.015)	(0.017)	(0.032)	(0.005)	(0.005)	(0.012)
21-30 years into lifecycle	-0.074	-0.073	-0.145	-0.026	-0.022	-0.059
	(0.021)	(0.025)	(0.045)	(0.006)	(0.005)	(0.013)
31-40 years into lifecycle	-0.110	-0.116	-0.222	-0.031	-0.029	-0.067
	(0.026)	(0.033)	(0.062)	(0.007)	(0.006)	(0.015)
41-50 years into lifecycle	-0.135	-0.145	-0.284	-0.034	-0.029	-0.074
	(0.034)	(0.042)	(0.078)	(0.008)	(0.007)	(0.018)
51-60 years into lifecycle	-0.182	-0.194	-0.375	-0.033	-0.028	-0.075
	(0.040)	(0.049)	(0.093)	(0.010)	(0.008)	(0.021)
61+ years into lifecycle	-0.214	-0.236	-0.451	-0.041	-0.033	-0.092
	(0.051)	(0.063)	(0.119)	(0.012)	(0.009)	(0.026)
N	3771	3771	3771	10421	10421	10421
R^2	0.32	0.38	0.38	0.56	0.54	0.58
Token FEs	Y	Y	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y	Y	Y
Y mean	0.88	0.84	1.96	0.58	0.48	0.74

Notes: Table estimates changes in the dispersion of firm creation across economic sectors or the value chain over time. Dispersion measures are (i) the top activity/sector share ("Top 1"; Columns 1 and 4), the activity/sector HHI ("HHI"; Columns 2 and 5), and the coefficient of variation ("C.V."—in essence, the standard deviation of activity/sector shares; Columns 3 and 6). SEs clustered by technology in parentheses.

Table C.2: Upstream to downstream industry evolution
robustness check: excluding non-nouns from the technology sample

	Activity		Sector	
	(1)	(2)	(3)	(4)
	Manuf.	Services	Manuf.	Services
11-20 years into lifecycle	-0.031	0.047	-0.007	0.008
	(0.038)	(0.046)	(0.005)	(0.006)
21-30 years into lifecycle	-0.101	0.109	-0.011	0.026
	(0.040)	(0.051)	(0.006)	(0.007)
31-40 years into lifecycle	-0.119	0.099	-0.023	0.029
	(0.040)	(0.056)	(0.007)	(0.008)
41-50 years into lifecycle	-0.149	0.148	-0.026	0.036
	(0.040)	(0.060)	(0.007)	(0.009)
51-60 years into lifecycle	-0.168	0.136	-0.024	0.037
	(0.044)	(0.067)	(0.009)	(0.010)
61+ years into lifecycle	-0.178	0.165	-0.030	0.046
	(0.046)	(0.073)	(0.010)	(0.012)
N	3771	3771	10421	10421
R^2	0.33	0.45	0.76	0.70
Token FEs	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y
Y mean	0.07	0.70	0.18	0.35

Notes: Table estimates changes in the share of firm creation across value chain activities and economic sectors, comparing manufacturing and services. SEs clustered by technology in parentheses.

Table C.3: Firm creation and the dispersion of activity in relation to innovation characteristics
robustness check: excluding non-nouns from the technology sample

Panel A: Firm creation

		By form		By sector	
	(1)	(2)	(3)	(4)	(5)
	All	DE	Non-DE	Manuf.	Other
Std(Max value of first 100 patents)	-0.036	-0.028	-0.035	-0.034	0.004
	(0.080)	(0.083)	(0.086)	(0.085)	(0.084)
Std(Max breakthru in first 100 patents)	0.370	0.309	0.368	0.351	0.336
	(0.143)	(0.181)	(0.142)	(0.149)	(0.152)
Std(Combinations in first 100 patents)	0.285	0.196	0.300	0.400	0.297
	(0.107)	(0.119)	(0.106)	(0.127)	(0.115)
N	494	280	490	468	469
R^2	0.12	0.12	0.13	0.15	0.12
Decade FEs	Y	Y	Y	Y	Y

Panel B: Dispersion of activity

		Activity			Sector	
	(1)	(2)	(3)	(4)	(5)	(6)
	Top 1	HHI	C.V.	Top 1	HHI	C.V.
Std(Max value of first 100 patents)	0.012	0.022	0.040	0.010	0.009	0.027
	(0.016)	(0.017)	(0.030)	(0.008)	(0.008)	(0.018)
Std(Max breakthru in first 100 patents)	-0.019	-0.028	-0.050	-0.008	-0.007	-0.015
	(0.013)	(0.015)	(0.028)	(0.006)	(0.006)	(0.015)
Std(Combinations in first 100 patents)	-0.020	-0.028	-0.053	-0.010	-0.009	-0.027
	(0.013)	(0.015)	(0.028)	(0.006)	(0.005)	(0.014)
N	278	278	278	469	469	469
R^2	0.11	0.13	0.12	0.02	0.04	0.04
Decade FEs	Y	Y	Y	Y	Y	Y

Notes: Panel (A) estimates the relationship between total firm creation in a given technology and characteristics of its first 100 patents. Column (1) measures all firms; Column (2) and (3), manufacturing and non-manufacturing firms; and Columns (4) and (5), Delaware jurisdiction vs. other firms. Panel (B) estimates differences in the dispersion of firm creation in a given technology as a function of the characteristics of its first 100 patents. Dispersion measures are (i) the top activity/sector share ("Top 1"; Columns 1 and 4), the activity/sector HHI ("HHI"; Columns 2 and 5), and the coefficient of variation ("C.V."—in essence, the standard deviation of activity/sector shares; Columns 3 and 6). Robust SEs in parentheses.

Table C.4: Annual firm creation as a function of current and recent patenting
robustness check: excluding non-nouns from the technology sample

Panel A: Current patenting

		By form		By sector	
	(1)	(2)	(3)	(4)	(5)
	All	DE	Non-DE	Manuf.	Other
Ln(Patents)$_t$	0.293	0.241	0.297	0.319	0.305
	(0.042)	(0.054)	(0.044)	(0.045)	(0.044)
N	10326	2867	10033	9458	9547
R^2	0.83	0.72	0.83	0.76	0.84
Token FEs	Y	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y	Y
Y mean	1.67	1.25	1.65	-0.38	1.51

Panel B: Recent patenting

		By form		By sector	
	(1)	(2)	(3)	(4)	(5)
	All	DE	Non-DE	Manuf.	Other
Ln(Patents)$_{t-1}$	0.214	0.196	0.214	0.259	0.235
	(0.037)	(0.073)	(0.037)	(0.042)	(0.039)
Ln(Patents)$_{t-2}$	0.107	0.078	0.100	0.097	0.095
	(0.021)	(0.057)	(0.021)	(0.030)	(0.023)
Ln(Patents)$_{t-3}$	-0.002	0.050	-0.010	0.018	-0.006
	(0.020)	(0.052)	(0.020)	(0.033)	(0.023)
Ln(Patents)$_{t-4}$	-0.009	-0.064	0.006	-0.047	-0.004
	(0.029)	(0.061)	(0.028)	(0.038)	(0.030)
N	9565	2731	9302	8798	8860
R^2	0.84	0.73	0.84	0.78	0.85
Token FEs	Y	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y	Y
Y mean	1.71	1.27	1.69	-0.32	1.55

Notes: Panel (A) estimates relationship between annual firm creation and patenting in a given technology, and Panel (B) the relationship between annual firm creation and recent patenting. Observations in both panels are at the technology-year level. Column (1) measures all firms; Column (2) and (3), manufacturing and non-manufacturing firms; and Columns (4) and (5), Delaware jurisdiction vs. other firms. SEs clustered by technology in parentheses.