Patent Hunters

Lauren Cohen Umit G. Gurun S. Katie Moon Paula Suh*

Abstract

Analyzing millions of patents granted by the USPTO between 1976 and 2020, we find a pattern where specific patents only rise to prominence after considerable time has passed. Amongst these late-blooming influential patents, we show that there are key players (patent hunters) who consistently identify and develop them. Although initially overlooked, these late-blooming patents have significantly more influence on average than early-recognized patents and are associated with significantly more new product launches. Patent hunters, as early detectors and adopters of these late-blooming patents, are also associated with significant positive rents. Their adoption of these overlooked patents is associated with a 6.4% rise in sales growth ($t = 3.02$), a 2.2% increase in Tobin's Q ($t = 3.91$), and a 2.2% increase in new product offerings ($t = 2.97$). We instrument for patent hunting, and find strong evidence that these benefits are causally due to patent hunting. The rents associated with patent hunting on average exceed those of the original patent creators themselves. Patents hunted tend to be closer to the core technology of the hunters, more peripheral to the writers, and to be in less competitive spaces. Lastly, patent hunting appears to be a persistent firm characteristic and to have an inventor-level component.

Keywords: Innovation chain, innovation rent, patent citations, patent evolution, technological impact, technological trajectories, commercialization.

JEL Classification: O31, O33, L1

---

*Cohen is at Harvard Business School and NBER, Baker Library 279, Soldiers Field, Boston, MA, 02163 [lcohen@hbs.edu]. Gurun is at the University of Texas at Dallas, 800 W Campbell Rd., SM41, Richardson, TX 75080, [umit.gurun@utdallas.edu]. Moon is at the University of Colorado at Boulder, Leeds School of Business, Campus Box 419, Boulder, CO 80309, USA, [katie.moon@colorado.edu]. Suh is at the University of Georgia, Terry College of Business, 620 S. Lumpkin St., Athens, GA, 30602, [paula.suh@uga.edu]. We are grateful for comments from Ufuk Akcigit, Daniel Andrei, Jonathan Cohn, Vincent Glode, Po-Hsuan Hsu, Tong Liu, Holger Mueller, Seula Kim, Josh Krieger, Josh Lerner (discussant), Filippo Mezzanotti (discussant), Christian Opp, Elena Pikulina, Elvira Sojli, Noah Stoffman (discussant), Mark Rompel, Edward Van Wesep, Qiping Xu, Alminas Zaldokas, and seminar participants at 2024 Asian Bureau of Finance and Economic Research Annual Conference, 2024 University of Kentucky Finance Conference, 2024 SFS Cavalcade NA, 2024 Midwest Finance Association meeting, 2024 Frontiers In Finance Conference, 2024 Global Entrepreneurship and Innovation Research Conference, 2023 Northern Finance Association meeting, 2023 European Financial Management Association meeting, 2023 University of Connecticut Finance Conference, 2023 Florida State University Truist Beach Conference, London Business School, Iowa State University, McGill University, the University of Colorado, and the University of Georgia, Korea Advanced Institute of Science & Technology (KAIST).

1 Introduction

Not all ideas that eventually become successful are recognized immediately. Eventual positive realizations take many divergent paths in reaching that point. In this paper, we analyze millions of patents to identify ideas that are not immediately recognized but catch on only later. We show that while these ideas are equally as valuable as ideas that catch on early, the rents along the value chain are shared quite differently. Namely, we provide the first large-sample evidence that there are critical agents in the innovation chain who search out ("hunt") these neglected and overlooked ideas and use them as critical inputs in their innovation and commercialization process. We show that these agents are unique and non-substitutable players in the innovation chain. Moreover, the rents to "patent hunting" are substantial—often the most sizable portion of the entire innovation chain. This patent hunting role, along with the technology, physical capital, and human capital needed to implement it, represents a key component of many innovation chains and a key consideration for agents entering across various innovation stages.

To explore these rich components of the innovation value chain, we examine the past nearly 50 years of patenting in the United States and identify those patents that eventually do gain prominence, becoming influential patents. While many of those are identified early on as influential patents that other innovators build upon, a sizable portion only do so later in their lives, being described as "late-bloomer" patents. Namely, these patents eventually become influential, however are not recognized initially and such, and not so until much later in their life. As all patents are, upon approval, publicly available for other innovators to read, build upon, and cite, one could imagine that conditioning on patents that end up as influential, those that are passed over initially are of lower ending value on average. However, we find that this is empirically not the case. These late-bloomer patents, when compared with patents that are recognized earlier, appear just as valuable on average and, by some quality markers, have even more impact.

To better understand our approach, consider the example of GPU (Graphical Processing Unit) computing technology. In 1991, US Patent #5,025,407 was granted to Texas Instruments, Inc. Texas Instruments is a publicly traded technology firm based in Dallas, TX specializing in semiconductors and other circuitry technology. This patent was in the technology classes of both G06F (Electric digital data processing) and G06T (Image data processing), as shown in Panel (a) of Internet Appendix Figure A1. Texas Instruments's core technology class was H01L (Semiconductor devices).

In its early years, the patent garnered comparatively little attention. In fact, it was in 2006, the fifteenth year following its grant, that it saw a large and record spike in citations and innovation build-out as shown in Internet Appendix Figure A1 Panel (b). This surge was largely driven by a single firm: Nvidia, Corp. Nvidia (also a publicly traded technology company) was founded in the early 1990s and is based in Santa Clara, CA. It was founded specifically focusing on the promise of graphics technology, with its technology proximity much closer to the Texas Instruments patent than the inventing firm itself. Nvidia used this patent to continue to develop and build out its product portfolio, which both contributed to—and was positively buoyed by—the positive demand trends of gaming (especially mobile), and GPU-reliant Artificial Intelligence, machine learning, and crypto-asset demands. Subsequent to this period led by Nvidia, prominent innovators like Apple, Inc. entered the fray, focusing particularly on the computationally intensive requirements that emerged in the years leading up to and throughout the early 2020s. Interestingly, the patent's inventor—Texas Instruments—did not end up taking further part in building out GPU computing technologies in earnest, nor did it continue to take and build on this critical patent. Instead, other key players, such as Nvidia, years later took the patent, built upon it, and developed and commercialized an industry of products in the space.

We demonstrate that instances like this are far from rare, occurring regularly and extensively across the spectrum of patenting and innovation. In Internet Appendix Figure A2, we catalog a selection of other notable technologies—including battery packs, carbon refrigerators, liquid dish-washing detergents, and control logic interfaces for embedding microprocessors in gate arrays—that display analogous trends to Texas Instruments and Nvidia. In particular, patent hunters, which we model and define as directed-search innovators who search out and integrate overlooked innovation (*e.g.,* Nvidia) routinely, are critical players in the development and commercialization of late-blooming, high-impact patents across their innovation chain. Moreover, we find that patent-hunting is associated with sizable rents across various dimensions of value. For instance, patent hunters' sales growth increases significantly following their discovery and incorporation of the hunted patent by 6.4% ($t = 3.02$) on average. Tobin's Q also increases by about 2.2% ($t = 3.91$), coupled with a significant expansion of 2.2% ($t = 2.97$) in the real quantity of new products developed by the patent hunter.

We explore the characteristics of firms at various stages of the innovation value chain. The writers of the initially overlooked, late-bloomer patents tend to be older, larger, value firms such as General Electric, Eastman Kodak, and Xerox. In contrast, patent hunters tend to be smaller, consumer-focused, growth firms such as Sandisk, Broadcom, and Tivo Corp. Moreover, patent hunting appears to be a persistent firm characteristic. The same firms continue hunting over time, and the rents to their hunting appear to have a learning component, with additional hunted patents yielding increasing benefits to the hunter over time.

Next, we examine the nature of the patents that become the target of hunting. One might expect that, even if a patent is overlooked by other agents, the patenting firm itself should be aware of the patent and able to build upon it. Therefore, one might expect there to be a reason why the original patent writer does not further develop the patent and its technology area, allowing the patent to be hunted and developed by an outside firm. We identify several systematic markers of these patents consistent with this notion. First, and in line with the Texas Instruments-Nvidia example, we find that patents that are hunted by outside firms are on average significantly more distant from the patent writers' core technology areas than the writers' average patent, and conversely significantly closer to the core technology area of the respective patent hunters. Second, we find that these hunted patents are in technology spaces that are less competitive at their time of patenting. This may exert less time pressure on writers to develop the patent immediately, reinforcing their focus on closer and more competitive spaces first.

Late-blooming hunted patents also take a diverging path from other influential patents. We precisely define influential patents, using a non-parametric measure of the impact, as those that receive citations in the 95th percentile or higher adjusted for vintage and technology class. We take this approach as past literature has shown this right-tail parameterization to more accurately capture the commercializability of a patent and correlate more strongly with patent value, reflecting the highly skewed distribution of patent citations (Trajtenberg, 1990; Sampat and Ziedonis, 2004).

We then collect micro-level data on the individual inventor level to explore inventor-level components in the patent-hunting process. We find two dynamics reinforcing an important inventor-level determinant of the process. First, we identify firms that end up hiring the inventor of the original patent that they "hunted." In the Texas Instruments-Nvidia example above, this would be Nvidia hiring one of David Gulley and Jerry Van Aken from Texas Instruments (the two listed patent inventors on the patent) following their publishing of the patent. We find that in these cases, the benefits accruing to the patent hunting firm (sales growth and Tobin's Q) more than double. It surely could be that hiring the original inventors is part of a broader scope of investments in the new technology space (and so the real effects that we measure are not solely due to bringing in the original inventors of the focal patent itself). Nonetheless, these inventor acquisitions along with the hunted patent are strongly related to greater economic and statistically reliable value derived from patent hunting. Second, we find that inventors themselves who engage in patent hunting (so the inventors at Nvidia who cited the Texas Instruments patent) are significantly more likely to continue patent hunting across different workplaces (*e.g.,* when moving to Tesla). Specifically, those inventors are 7% ($t = 11.23$) more likely to continue to be hunting inventors at subsequent firms. Much like the previous finding, it is difficult to disentangle whether this effect is attributable to the inventor herself or simply reflects the inventor's selection of firms that also engage in patent hunting. Yet, irrespective of the specific mechanism, this finding suggests that there is an inventor-level component of the hunting process.

Stepping back, given the large, positive rents that are associated with patent hunting – larger, in fact, than those that accrue even to the original patent writer itself, on average— one might wonder why anyone would choose to be a patent writer of the original patent. First, as discussed earlier, there are a number of moderating effects of patent hunting. The rents to patent hunting attenuate when the patent in search is complex, hence, difficult to assess the technological fit, or when there exists intense competition from other patent hunters in the same technology class. Both scenarios are consistent with increased search costs (lower equilibrium rents) impacting the benefits from patent hunting and moreover suggest that the patent hunting mechanism might follow a search cost dynamic. Motivated by this, we develop a simple search model in Section 3 to frame the understanding around this. Second, there are firm-specific characteristics that cause cross-sectional and time-series heterogeneity in the relative costs and benefits of patent hunting. While labor may appear somewhat mobile (so that patent writing firms like Texas Instruments could simply attract researchers from Nvidia with sufficiently high wages, benefits, etc.), there are many non-transferrable characteristics such as location, agglomeration, intangible capital (*e.g.,* brand), complementarities and other firm-specific components that make patent hunting uniquely and privately valuable for certain firms while unprofitable for others. In that sense, no firm is necessarily solving sub-optimally along the innovation chain deciding to be a writer or hunter.

Our paper provides the first large-scale empirical evidence on the economic rents and value chain implications of patent hunting. The substantial and persistent payoffs we uncover suggest promising avenues for further research. The paper is structured as follows: Section 2 reviews relevant literature. Section 3 develops a search-cost model of the innovator's write-or-hunt decision process. Section 4 details the data and sample selection methodology. Section 5 analyzes the dynamics of early- and late-bloomer patents. Section 6 presents our main findings on patent hunters and their accrued benefits. Section 7 examines the incentives and characteristics of firms and inventors engaged in patent hunting. Section 8 examines the causal impact of patent hunting using inventor movements, and patent characteristics to determine if patent hunting is a valuable skill or a sideshow. Section 9 concludes.

2 Literature Review

Our paper adds to the literature studying the path of knowledge production, technological innovation, and their impacts on economic growth. Along these lines, Weitzman (1998) presents a model to analyze the determinants of long-term growth by considering a production function of new knowledge that uses new configurations of old knowledge as an input (recombinant growth). The paper emphasizes the role of building upon existing ideas to create new knowledge. In the academic research setting, analyzing citation patterns of over 18 million scientific papers, Uzzi, Mukherjee, Stringer, and Jones (2013) show that the most influential research tends to be rooted in conventional combinations of existing knowledge but also includes elements of unusual combinations. Similarly, Escolar, Hiraoka, Igami, and Ozcan (2023) focus on technological trajectories and the reuse of knowledge in subsequent inventions, emphasizing the significance of combining dissimilar technological components with strong scientific content to shape trajectories with high technological impact. In a similar dynamic setting, past work such as Gal-Or (1987) and Chamley and Gale (1994) have modeled Stackelberg equilibria, where distinct advantages exist for certain agents who benefit from a second-mover advantage and entry strategy in the traditional leader-follower set-ups. Glode and Ordonez (2023) model surplus-creating and surplus-appropriating activities of firms, relating these decisions to industry-wide technological shocks that change firms' incentives to invest in one activity over the other.

Pezzoni, Veugelers, and Visentin (2022) investigate the reuse of technologies in subsequent inventions and explore the various factors influencing this process. Using European patent data spanning from 1985 to 2015, they identify new combinations of existing technological components, marking the inception of technological trajectories. Instead of relying on citations to trace these trajectories, the authors identify a new technology as the first occurrence of a particular combination of IPC classes in a patent, akin to the approach taken in Strumsky and Lobo (2015) and Verhoeven, Bakker, and Veugelers (2016). Pezzoni et al. (2022) also observe that technological trajectories tend to follow an S-shaped curve, with variations in their adoption time and maximum technological impact. We add richness to this literature by not only exploring firms, even down to individual inventors, who specialize in critical components of this process, but also by examining the expansive set of value implications of identifying and building on a patent within its technological trajectory.

Our paper also contributes to research exploring the ramifications of technological spillovers on both economic growth and technological advancement. For example, Bloom, Schankerman, and Van Reenen (2013) investigate the pivotal role played by technology spillovers in stimulating economic growth, underscoring the significance of incentivizing research and development (R&D) and considering the scale of firms involved. Furthermore, Kelly, Papanikolaou, Seru, and Taddy (2021) employ textual analysis of patent documents to gauge technological innovation. They identify patents that stand apart from previous work but have high relevance to subsequent innovations. They classify these distinctive patents as breakthrough innovations, investigate the domains in which these breakthroughs occur, and establish connections between these breakthrough innovations and overall total factor productivity. Collectively, these studies offer valuable insights into the intricate interplay between knowledge creation, technological progress, and economic growth. Our contribution to this body of research lies in the identification of specific technology groups where spillovers may require time to materialize, along with mechanistic, micro-evidence on how technological growth and factor productivity materialize through these novel important agents of patent hunters.

In addition, our paper also contributes to the ongoing discourse surrounding the merits and demerits of a patent system where inventors publicly unveil their innovations in exchange for patent protection, and the potential ramifications of such a system on future innovations. One perspective posits that the imperative for patent disclosure might dissuade individual inventors, potentially eroding the incentives inherent in the patent system. Conversely, patent disclosure and protection could serve as a mechanism to stimulate new ideas and foster innovation (as discussed by Williams (2017)). On the empirical side, Furman, Nagler, and Watzinger (2021) examine the role of information disclosure via patents, showing that increased accessibility to technical knowledge significantly bolsters local patenting and business establishment. Their findings underscore the role of patent disclosure in advancing cumulative innovation. In contrast, Kim and Valentine (2021) report that firms compelled to disclose their innovations more promptly by the American Inventor's Protection Act (AIPA) of 1999 reduce their R&D investments and generate fewer innovations (see also Graham and Hegde (2015) and Hegde, Herkenhoff, and Zhu (2015)). Our paper adds to this literature by showing a new merit of a patent disclosure system—whereby agents such as patent hunters can detect, and build upon, initially overlooked ideas to create new knowledge.

In the case of scientific academic publications, Garfield (1970) proposes that the use of scientific citation indices not only prevents inadvertent neglect of useful work but also reduces duplicative effort in research and publication. Subsequently, Garfield (1980, 1989a,b, 1990) provide concrete examples of articles that experienced delayed recognition within the scholarly community. Expanding on this notion, Glänzel, Schlemmer, and Thijs (2003), through an extensive literature survey, offer an estimate of the prevalence of delayed recognition and explore common characteristics shared by papers that receive belated recognition. Van Raan (2004) proposes a framework for measuring delayed recognition, often referred to as *sleeping beauties*, along three dimensions: (i) the length of sleep, signifying the duration of the "sleeping period"; (ii) the depth of sleep, denoting the average number of citations during the sleeping period; and (iii) awake intensity, indicating the number of citations accumulated after the sleeping period.

In a similar vein, Ke, Ferrara, Radicchi, and Flammini (2015) reexamine the concept of delayed recognition of academic publications, introducing parameter-free methods to identify papers that might escape detection by the methods proposed by Van Raan (2004). Van Raan and Winnink (2019) document instances in medical research where publications went unnoticed for several years after their initial release, only to suddenly garner citation attention subsequently. Moreover, Van Raan (2017) and Hou and Yang (2019) extend the analysis of delayed recognition from scientific papers to patents, exploring various evolutionary trajectories of patents in this context. Our paper extends the concept of delayed recognition in the realm of patents, specifically mapping the benefits and costs experienced by innovators whose patents gain recognition after a significant delay. Further, it again moves a step beyond this to examine critical agents in the recognition process, and the value that accrues to these agents, who mechanistically bring about the recognition of these initially overlooked ideas.

3 Theoretical Framework

In this section, we provide a model of firm innovation: namely, a search-cost model of whether to write or hunt for a patent. Time is discrete, and the horizon is infinite. At each time $t > 0$, a firm can either write a new patent itself or hunt for a patent that has been produced precedingly by others. If the firm writes the patent itself, the patent will help produce a good at zero marginal cost in each period t. Consumers value this good as $v > 0$, and the price is $p = \alpha v$, where $\alpha \in (0,\ 1)$ denotes the surplus that the firm can capture in the market by commercializing the patent.

Alternatively, the company can engage in directed search, incurring a cost c, to identify patents within the same area with potentially greater commercialization value than those developed internally (*i.e.*, $V - L - A \geq v$). The directed search is comprised of both (i) assessing the expected value of complementarities of internal vs. external technology and innovation given its existing assets and innovation capital in place, and (ii) conducting the search across the innovation space, both new and old, potentially reconfigurable, ideas. V represents the value of the goods produced using the identified patent with directed search. L is the cost of acquiring a technology license from the original innovator. A is the expenses associated with adapting the product for successful commercialization. It is important to note that an innovator is more likely to demand a higher licensing fee when the original technology closely aligns with its core technologies. The likelihood of discovering this second category of patent is $\lambda \in (0,\ 1)$. Upon identifying such a patent, the company transitions from producing the previous generation of goods to creating novel products. In each time period, there exists a probability $\delta \in (0,\ 1)$ that the scenario concludes, bringing the game to an end.

In light of these fundamental parameters, we provide a comparative analysis of a firm's strategic choices, delineated by two distinct avenues:

a) Firm valuation in the absence of patent exploration:

The cumulative summation of $(1-\delta)^t(1-\lambda)^t\alpha v$ over time gives us $\frac{\alpha v(1-\delta-\lambda+\lambda\delta)}{\delta+\lambda-\lambda\delta}$.

b) Firm valuation upon discovery of a hunted patent at time t:

$$\sum_{t=1}^{\infty}(1-\delta)^t((1-\lambda)^t\alpha(v-c)+(1-\lambda)^{t-1}\lambda H)),$$

where $H = \alpha(V-L-A)(1-\delta)\delta$ is the perpetuity value of the newly hunted technology.

The net benefit of following a directed search strategy to find a hunt is then given by the difference between these values:

$$\alpha(V-L-A)(1-\delta)\lambda - \ c\delta(1-\lambda).$$

This expression is increasing in

(i) λ – As the probability of finding a suitable patent increases, the avidity of the pursuit escalates;

(ii) α – The magnitude of commercialization potential amplifies the incentive for hunting.

In contrast, the expression decreases with

(i) c – A higher cost attached to patent hunting, reduces its net benefit;

(ii) L and A – The appeal of hunting decreases when licensing fees L and adaptation costs A become prohibitive.

4 Data and Sample Selection

Our data come from a number of sources. U.S. Patent Data are obtained from Thomson Innovation, which covers all patents granted in the United States between 1835 and 2020 and contains information on backward and forward citations. We merge this patent data with PatentsView data, which provides information on assignees, claims, inventors, and examiners of the patents granted since 1976. Our main analyses focus on U.S. public firms, for which we collect financial data and firm characteristic data from the S&P Compustat Database. To identify patents belonging to U.S. public firms, we match patent numbers to the University of Chicago's Center for Research in Securities Prices (CRSP) identifier PERMNO, using Kogan, Papanikolaou, Seru, and Stoffman (2017) data. The data on new products of U.S. public firms are obtained from Mukherjee, Thornquist, and Žaldokas (2022). Lastly, we also use the FactSet Revere Relationship data and Compustat customer segment data to identify the relations between citing and cited firms, including customer-, supplier-, and competitor-relationships, along with patent licensing-fee relationships. Our final data for the analyses consists of all public U.S. firm patents between 1976 and 2020. See Internet Appendix Table A1 for more details on our sample selection procedure.

5 Influential and late-bloomer patents

We define a patent as an influential patent if the cumulative citation count of the patent within its CPC class and grant year cohort is in the 95th percentile or higher at any point in time during the patent term of 20 years since its grant.[1] To make a fair comparison of citations across patents granted in different years and also to consider the issues of truncation and citation-clustering raised in Lerner and Seru (2022), we (i) only consider citations during the term of patenting (*i.e.,* the first 20 years) and (ii) compute the cumulative citation percentiles within the same grant year and CPC class. By doing so, we restrict our sample to patents granted between 1976 and 1999, so that the last cohort of patents granted in 1999 has complete 20-year citation data ending in 2020. We remove self-citations from the citation count as we aim to capture the use of patents by external users.

By definition, influential patents are significantly more successful patents than the bottom 95% patents based on the number of forward citations. We illustrate the stark differences in citations between influential and bottom 95% patents in Figure 1. Panels (a) and (b) compare citation counts of influential and bottom 95% patents over patent age. Panel (a) plots the average number of citations for each group using different scales for visual clarity. We note in Panel (a) that citations of both influential and bottom 95% patents grow rapidly in the initial five years of patent life. The average number of citations for the bottom 95% patents reaches its peak at the age of five and remains relatively constant thereafter. In contrast, the average number of citations for influential patents continues to grow, reaching its peak much later at the patent age of 16. Influential patents also receive a significantly larger number of citations at every point in the patent's age (*e.g.,* 3 vs. 0.5 at the age of five). Consistent with this observation, Panel (b) shows that the cumulative number of citations for influential patents grows at a much faster rate compared to that of the bottom 95%

[1] The approach using percentile distributions of cumulative citations is well accepted in the literature. The right-tail parametrization has been shown to highly correlate with the value and commercializability of a patent (Trajtenberg, 1990; Sampat and Ziedonis, 2004). Also, see Brogaard, Engelberg, and Van Wesep (2018) as an example of similar convex value-distributions for academic citations.

patents. In fact, by the patent age of 20, influential patents have accumulated more than five times the cumulative citations of the bottom 95% patents.

[Insert Figure 1 Here]

We then explore different paths to success by examining the time it takes to become an influential patent. Panel (c) of Figure 1 plots the distribution of the years it takes for a patent to be recognized as an influential patent.[2] The average (median) time until an influential patent reaches the top 5% of the cumulative citation distribution within the same grant year and CPC class is five (three) years. We use the 90th percentile cutoff (14 years) in this distribution to further characterize patents that take a significantly longer time to become an influential patent and classify them as "late-bloomer patents." We call the remaining influential patents "early-bloomer patents."

Figure 2 contrasts the divergent paths to the success point between early- and late-bloomer patents more precisely. Late-bloomer patents receive a small number of citations near the grant year but accumulate a large number of citations later in their lives. A marked increase in cumulative citation growth around the patent age of 10 in Panel (a) confirms this point. The box plot in Panel (b) also shows this convexity of late-bloomer patents in cumulative citations over the patent age. In contrast, early-bloomer patents start to accumulate citations early on as shown in Panel (c), and the box plot in Panel (d) shows their concavity of cumulative citations over the patent age.

[Insert Figure 2 Here]

Next, we examine the differences in patent characteristics at the time of patent grant. We present the summary statistics that compare different groups of patents in Table 1. Panel A presents the descriptive statistics of influential and bottom 95% patents. Influential patents tend to be broader, spanning a larger number of CPC class categories and containing

[2] We denote the year when a given patent becomes an influential patent "influence-year." We use this variable extensively in our subsequent analyses when we analyze the timing of the patent-hunting benefits.

more patent claims. They are more likely to be assigned to public corporations, make more backward citations, and experience positive market responses on the grant date as measured by the KPSS (Kogan et al., 2017) patent value metric. When we compare early-bloomer patents to late-bloomer patents in Panel B, we see a number of patterns that develop as the patents age. In particular, confirming Figure 2, early-bloomer patents surge to more citations early in their lives: having significantly more cumulative citations at age 5, 10, and even 15 years. However, by the end of the patents' respective terms, the lead flips, and late-bloomer patents actually end up with significantly more citations on average than early-bloomer patents (70 vs. 53).

[Insert Table 1 Here]

At the time of grant, however, the distinction between early- and late-bloomer patents is less apparent. The KPSS patent values for both types are statistically and economically indistinguishable, suggesting that market participants cannot initially differentiate between them. This insight implies that a patent's ultimate impact may be significantly influenced by its subsequent users and citations. These findings motivate a deeper examination of the characteristics of patents citing early- and late-bloomers, as well as the firms and inventors that utilize them.

Panel C presents the descriptive statistics of citing patents' characteristics. Compared to the focal patents in Panels A and B, we first note that citing patents appear to be less successful than focal patents in terms of the number of forward citations that they receive. In contrast, the citing patents make a substantially broader search of patents (*e.g.,* make significantly more backward citations). It is possible that this result is partly driven by the fact that those citing patents relative to focal patents are granted later in time, which increases the size of the entire patent pool for backward citations. However, we find that within citing patents, those that cite late-bloomer patents particularly make a larger number of backward citations in comparison to those that cite early-bloomer patents. In our analyses later, we further focus on this aspect—the breadth of patent users—of late-bloomer patents to investigate how once-neglected inventions evolve into highly successful innovations.

6 Results

6.1 Late-bloomer Patent Writers and Users

We begin our analysis by contrasting firms that write late-bloomer patents and firms that use those late-bloomer patents. In theory, users could utilize these patents in a number of ways, such as purchasing, licensing, or citing the late-bloomer patents explored. Due to the opacity in defining and precisely delineating the idea-space encompassed by any given patent's commercializable right (*e.g.,* the smartphone is associated with over 250,000 patents and growing (Yang, 2014)), it may make the perceived need or pressure to license or purchase a given patent less pressing. Consistent with this notion, patent licensing and reassignments (purchases) are empirically rare, affecting fewer than 3% of patents (Berman, 2002).[3] Patent citations thus represent by far the most common and empirically observed method through which firms utilize and build upon innovation. Given this, we focus on contrasting firms that write late-bloomer patents and firms that cite (*i.e.,* use) those late-bloomer patents. Because firms can both write and use late-bloomer patents, we sharpen the definition of writers and users exclusively for comparative statistics that are shown in Table 2. We define here users of late-bloomer patents to be more exclusive as firms that have never written a late-bloomer patent themselves during the sample period while they cite at least one. Writers of late-bloomer patents are defined as firms that have produced at least one late-bloomer patent during the sample period regardless of whether they have ever cited late-bloomer patents of other firms.

In Internet Appendix Table A2, we list the top 20 firms for each group of late-bloomer patent writers and users during our sample period. The first thing we note from the list is

[3] This statistic is also consistent with what we observe in our sample, where only roughly 2.5% of patents are licensed.

that writers appear to be much older (*e.g.,* U.S. Surgical Corp, Johnson & Johnson, and IBM) than users, *i.e.,* log(age) of writers is 2.17 whereas that of users is 1.56 (so nearly double at 9 years vs. roughly 4.8). Since the age gap can drive large differences in many financial variables mechanically regardless of the writer or user identity, we report summary statistics using a more refined age-matched sample. Specifically, the users in the refined sample are the five nearest neighbors of each writer. For this analysis, we use public firms that have ever written or cited a late-bloomer patent during our sample period. Panel A of Table 2 presents the summary statistics.

[Insert Table 2 Here]

We first examine various patenting characteristics. From Table 2, the average number of granted patents in a year is significantly larger for writers than users (29.71 vs. 2.84). Furthermore, the average number of patents in a year that eventually become influential patents is tenfold greater for writers than users (4.73 vs. 0.51). By construction, users have no late-bloomer patents while writers generate 0.62 late-bloomer patents per year, and 13.2% of their patents are identified as late-bloomer patents. Regarding patent impacts, we find that late-bloomer writers receive 63.91 external citations per year while users receive 4.13 external citations per year. Because writers file significantly more patents per year, we normalize the citation counts per year by the total number of patents per year. We still find that writers' normalized citation counts are greater, *i.e.,* 2.45 for writers vs. 1.52 for users, and the difference is statistically significant. These patterns collectively suggest that late-bloomer writers have a relative advantage in patenting, both in quantity and quality. Despite this conclusion, we find an interesting pattern regarding patent claims and commercialization. We note that the average number of claims is larger for users than writers (17.37 vs. 16.58), and the difference is statistically significant. The larger number of patent claims implies a broader patent applicability, and, hence, a higher chance of bringing the invention to markets as a new product. For a proxy for patent commercialization, we use the total number of new products (Mukherjee et al., 2022) per year divided by the total number of new patents per year. Based on this measure, we find that the commercialization rate is also significantly higher for users relative to writers (25.6% vs. 18.1%). Both results are consistent with the interpretation that late-bloomer users (patent hunters) particularly stand out as market players that are relatively more capable of commercializing new inventions.

We then compare financial variables between writers and users. We confirm that the average age of firms in this age-matched sample is 5 years in both groups and no longer shows the large age gap after the nearest neighbor matching. Even after the age matching, we find that the firm size, measured by all aspects—including book assets, market assets, and total revenues—is significantly larger for writers than users. Consistent with this result, we find that CAPX investment rates are higher for writers than for users and that writers are more likely to be mature firms paying dividends to shareholders. We also observe that writers invest more in R&D than users with the investment rate difference at around 1.6%. However, it is important to highlight that users also make significant investments in R&D, indicating they are still innovative firms but differ in their approach to innovation. We do not find any other differences in the remaining financial variables, including profitability, leverage, and advertisement. Lastly, we use the 2002 Input-Output Accounts from the Bureau of Economic Analysis (BEA) and create a variable for industry consumer dependence. We define an industry as more consumer-dependent if the industry's production percentage for "Personal consumption expenditures" in the BEA Input-Output Accounts is in the top tercile. We find that user firms are significantly more likely to be in consumer-dependent industries.

In sum, writers of late-bloomer patents tend to be older, large value firms with a more extensive stock of patents and citations and greater R&D spending. These firms appear to have enough resources given their size and investment scales, however do not commercialize every good innovation they generate. Conversely, users of late-bloomer patents are relatively smaller in size and younger, however generate significantly more new products per patent, and are firms more likely to serve consumers directly with their products.

6.2 Persistence in Being Late-bloomer Writers or Users

We note that our analysis in Panel A of Table 2 does not allow for switching between the two groups of writers and users throughout the period. In this section, we relax our classification scheme and allow firms to switch between the writer and user groups. Panel B of Table 2 presents a transition matrix where we examine the likelihood of firms changing their writer or user identity in the next period. Again, we use the sample of public firms that have ever written or cited a late-bloomer patent during our sample period. We define four sets of write/user status. $Strict\ Writer$ is a firm year where the firm produces a late-bloomer patent based on its grant year but does not cite any late-bloomer patent. $Flexible\ Writer/User$ is a firm year where the firm produces a late-bloomer patent based on its grant year and also cites late-bloomer patents. $Strict\ User$ is a firm year where the firm cites late-bloomer patents but does not produce any late-bloomer patents. Lastly, $Idle$ indicates a firm year where the firm neither produces nor cites a late-bloomer patent. We present both the number and percentage of observations for each category.

First, we find that the number of $Strict\ Writer$ is relatively modest in any given year in row (a) at only 5.7% (861 out of 14,946) of firm-years. However, when $Flexible\ Writer/User$ are added (row (b)), the percentage of writers rises to nearly 25% of the 14,946. Every other category (rows (b) - (d)), has a sizable amount of persistence, in that the most likely outcome is staying in the same category, as seen in that the diagonal contains the largest value (transition probability). In comparison, there exist 4.6 times more $Strict\ Users$ than $Strict\ Writer$, on average. Moreover, patent hunting appears to be a persistent firm characteristic. Roughly 72% (89%)[4] of current $Strict\ Users$ ($Flexible\ Users$) continue to patent hunt in some form in the next year.

These transition probabilities, coupled with the descriptive statistics in Panel A, reveal distinct innovation styles between late-bloomer writers and users. While existing literature

[4] 72% is calculated as the sum of next year either again being a Strict User (50.82%) added to being a Flexible Writer/User (20.94%). Equivalently, this summed transition probability for $Flexible\ Writer/User$ at t is 61.12%+28.18% 89%.

has primarily focused on patent writers and their innovation outcomes, our findings underscore the distinct role of patent users in the innovative process, suggesting a need for a more nuanced understanding of innovation dynamics beyond traditional categorizations.

6.3 Benefits of Patent Hunting

As discussed in the previous section, late-bloomer patent writers and users have distinct innovation styles and firm characteristics. In this section, we examine the incentives of being a user of late-bloomer patents (*i.e.,* of patent hunting). We first conduct firm-level analyses to explore the firm benefits of patent hunting. We estimate the following regression model:

$$Y_{j\{t,t+4\}} = a + b_1 log(1 + LBhunting)_{jt} + b_2 log(1 + EBhunting)_{jt} + b_3\theta_{jt} + \gamma_j + \eta_t, \quad (1)$$

where j is user firm and t is year. $log(1+LBhunting)_{jt}$ is the log of one plus the total number of late-bloomer patents that user firm j cites in year t. To compare the benefits of citing late-bloomer patents to those of citing early-bloomer patents that have a comparable patent impact, we also include $log(1+EBhunting)_{jt}$ in the model. $log(1+EBhunting)_{jt}$ is the log of one plus the total number of early-bloomer patents that user firm j cites in year t. The dependent variable for firm benefits is sales growth or firm value as measured by Tobin's Q in the subsequent five years. $Sales\ growth_{j\{t,t+4\}}$ is computed as $(sales_{j\{t+4\}}/sales_{j\{t\}}) - 1$, and $Avg\ Tobin's\ Q_{j\{t,t+4\}}$ is the five-year firm average of Tobin's Q. The regression is at the user firm-year level and includes firm-level control variables (θ_{jt}), firm fixed effects (γ_j), and year fixed effects (η_t). Standard errors are clustered at the firm level. Table 3 presents the results from the estimation of Eq.(1).

[Insert Table 3 Here]

We find in columns 1 and 2 that the intensity of patent hunting (finding and building upon late-bloomer patents) is associated with significantly higher future sales growth and firm value. The coefficients translate into a 6.4% increase in sales growth and a 2.2% increase in

Tobin's Q associated with doubling the number of late-bloomer patents hunted. In contrast, we find coefficients on $log(1+EBhunting)$ in both columns are significantly negative. While it is highly probable that a firm cites both late- and early-bloomer patents in a year or even within its single patent, the incremental benefits of searching, finding, and building upon early-bloomer patents do not appear to provide incremental benefit in the same manner as late-bloomer patents.

In columns 3 and 4, we examine the differential benefits of citing late-bloomer patents early vs. later by replacing $log(1+LBhunting)$ with $log(1+earlyLBhunting)$ and $log(1+laterLBhunting)$. $earlyLBhunting$ is the number of late-bloomer patents that a user cites, where the citation occurs among the first three citations on a given late-bloomer patent since its grant and prior to its influence-year. $laterLBhunting$ is the number of the rest of late-bloomer patents that the user cites. We find that the benefits of early hunting are roughly 10-50% larger in point estimate for sales growth and firm value, respectively than those for later hunting. These findings are consistent with those hunters being early finders, adopters, and innovating upon, neglected ideas reaping greater benefits relative to later followers and adopters.[5]

Overall, the firm-level analyses provide initial evidence consistent with benefits associated with patent hunting in terms of future firm growth and value. The benefits appear to be exclusive to hunting late-bloomer patents and do not extend to early-bloomer patents. In Section 8.1, we explore an exogenous event shocking the intensity of patent hunting at a given firm, and show evidence consistent with Table 3 on the causal benefits of patent hunting on firm sales and value.

[5] In Internet Appendix Table A3, we consider analogous tests for patent writers. The coefficient estimates in all columns are standardized for ease of comparison. In the first two columns, we examine the association between writer firms' financial benefits and the total number of late bloomers written in a year. In the last two columns, we present the corresponding results of users as in the first two columns of Table 3, again using standardized measures for ease of comparison with the coefficients on writer benefits in columns 1 and 2. In columns 1 and 2, we find writing late-bloomer patents is positively and significantly related to writers' future firm value while there is no relation to sales growth. However, such benefits are significantly smaller than those of users shown in columns 3 and 4, and the firm value increase with writing late-bloomer patents is only about one-third of writing early-bloomer patents in column 2.

To further elucidate the benefits of patent hunting, we employ a patent-level difference-in-differences analysis. This approach compares the sales growth and firm value of patent writers and users across three patent categories: late-bloomers, early-bloomers, and bottom 95% patents. For late- and early-bloomers, we examine changes before and after the year they reach the top 5% of cumulative citations within their grant year and CPC class (the influence-year). For the bottom 95% patents, we use the peak year of their cumulative forward citations. This methodology allows us to isolate the specific impacts of different patent types on firm performance. We estimate the following model:

$$Y_{ijpt} = a + b_1 user_{ijp} + b_2 iyear^{post}_{pt} + b_3 user_{ijp} \times iyear^{post}_{pt} + b_4 \theta_{it} + b_5 \delta_{jt} + \gamma_p + \eta_t, \qquad (2)$$

where j is user firm, p is focal patent created by writer firm i, and t is year. $user_{ijp}$ is an indicator that equals one for user firm j of patent p created by firm i and zero for writer firm i of the same patent. We consider writer-user pairs within the 20 years of the focal patent's grant. The dependent variable for firm benefits is sales growth or by Tobin's Q.

For late-bloomer and early-bloomer patents, $iyear^{post}_{pt}$ equals one if t is after their influence-year and zero otherwise. For a bottom 95% patent, $iyear^{post}_{pt}$ equals one if t is after patent p's peak year in cumulative forward citations and zero otherwise. The regression is at the focal patent-firm-year level and includes firm-level control variables (θ_{it}, δ_{jt}), focal patent fixed effects (γ_p), and year fixed effects (η_t). Standard errors are clustered at the focal patent-by-year level.

Using the above approach, Internet Appendix Table A4 documents that late-bloomer users (patent hunters), on average, have larger sales growth and firm value relative to the writers of the patents themselves. We do not observe similar patterns in early-bloomer patents and the bottom 95% patents. These patent-level analyses provide support for the idea that these benefits appear to be exclusive to hunting late-bloomer patents and do not extend to early-bloomer patents, suggesting a distinct value in identifying and leveraging previously overlooked innovations.

Internet Appendix Table A5 re-estimates the patent-level results in Table A4 by dropping citations possibly driven by business relationships (Fadeev, 2024). We find that our results are still economically large and significant in this sample with no business relationships. That is, patent hunting benefits do not appear to be driven by business relationships. To identify business relationships between citing and cited firms, we obtain business relationship data from FactSet Revere and Compustat customer segment data between 2003 and 2017. Due to the availability of data, the sample period for this analysis is restricted to 2003-2017. Following Fadeev (2024), we identify firm pairs categorized as partners (licensing, research collaboration, joint venture), suppliers, or customers as relationship-based citations and exclude them from the analysis. These results reinforce our conclusion that there are unique benefits in searching out, finding, and building upon, overlooked late-bloomer patents, unrelated to business relationships.

6.4 Patent Hunting and New Product Creation

In this section, we examine the relationship between patent hunting and new product creation. We use a measure of new product launches developed by Mukherjee et al. (2022), based on media articles announcing new product introductions.[6] When no significant product launch is reported, we assign a zero value. We then adapt Eqs.(1) and (2) by replacing the dependent variable with the logarithm of one plus the total number of new products in a given year. Table 4 presents our findings. Column 1 shows results from the modified Eq.(1). Columns 2 and 3 offer a comparison between late-bloomer and early-bloomer patents, while column 4 focuses on the bottom 95% patents. This analysis allows us to investigate whether patent hunting, particularly of late-bloomer patents, translates into tangible product innovations.

[Insert Table 4 Here]

[6] We thank the authors for sharing their data.

First, we find from the firm-level analysis in column 1 that patent hunting is significantly and positively associated with new product launches in the subsequent five years since hunting. The coefficient on $log(1 + LBhunting)$ translates into a 2.2% increase in the number of new product launches when doubling the number of late-bloomer patent citations. We also find that the number of early-bloomer citing is related to an increase in new product launches during the next five years, but the effect is significantly lower (p<0.01 on the F-test of coefficient equivalence), and less than one-half in magnitude compared to the effect for late-bloomer citing.

Results at the patent level—comparing writers and subsequent citers of the same patent—reported in column 2 show that patent hunters significantly have more new product launches than late-bloomer writers in general and that their influence-years are particularly associated with a 4.8% incremental increase in new product launches. In contrast, we do not observe that late-bloomer writers show similar increases in new products after their influence-years based on the significantly negative coefficient estimate for stand-alone $iyear^{post}$. Column 3 shows the results for early-bloomer patents. The coefficient for standalone $iyear^{post}$ is significantly negative, indicating that early-bloomer writers also face a decrease in the number of new products after their written patents' influence-years. However, the interaction term between $user$ and $iyear^{post}$ is statistically insignificant and close to zero.

Lastly, when we examine the bottom 95% patents in column 4, users have significantly fewer new product launches than writers, regardless of the influence-year timing. Centrally, we note that the benefits of commercialization as measured by new product launches typically accrue to the writer of a patent based on the results for the bottom 95% patents (and the positive and significant coefficient on $iyear^{post}$). The results in Table 4 provide strong evidence that late-bloomer patents are distinguished by their users (not writers) that appear to reap the greater benefits from developing new products. Such patent hunter benefits from creating new products likely contribute to the observed greater sales growth and firm value for patent hunters compared to writers.

7 Mechanism underlying Patent Hunting

7.1 Patent Writers' Incentives

Thus far, our results suggest that late-bloomer patents are unique in that they provide greater benefits to users than writers, with benefits associated with the creation of new products and commercialization. If so, one might wonder what incentivizes patent writers, and so founds their existence, ex-ante. In this section, we explore precisely this—reasons why late-bloomer writers produce (and continue to produce) late-bloomer patents.

First, we examine the characteristics of late-bloomer patents within the writer firms' patent portfolios. To do so, we regress an indicator variable for a late-bloomer patent on several patent and firm characteristics at the time of patent grant. For those characteristics, we particularly focus on the following three constraints under which writers may optimally delay implementing (or overlook) ideas and patents: (i) capacity constraints, (ii) a competitive threat, and (iii) financial constraints.

For measures of capacity constraints, we consider *tech-class weight* and *tech-class dist to core*. *tech-class weight* is the fraction of the writer's patents in a specific CPC class of a given patent over the entire sample period. This measure captures the importance of a particular technology class to the patent's writer. *tech-class dist to core* is the class-to-class proximity between the CPC class of a given patent and the core CPC class of its writer. The core CPC class is the CPC class with the highest *tech-class weight* within a firm. The class-to-class proximity measure is the distance between two CPC classes computed using the vector of how many patents with other CPC classes cite the patents in a given CPC class.

For a measure of competitive threat, we use *log(competing patent stock)* which is the log of the total number of patents from U.S. public firms with the same CPC class up to the grant year of a given patent. This measure captures how many other players exist in the same technology space at the time a given patent is produced. Lastly, for financial constraints, we consider the KZ index (Kaplan and Zingales, 1997), firm size and age following Hadlock and Pierce (2010), and the textual analysis-based measure of financial constraints by Linn and Weagley (2021) (referred to as the LW index).[7]

In addition to these constraint measures, we control for firm size, age, profitability, CAPX investment, and R&D investment. The regression is at the patent level (one observation per patent) by taking the averages of relevant variables when a patent has multiple CPC classes and using the grant-year data when control variables are time-varying. The regression includes both the writer and grant year fixed effects. Table 5 presents the results.

[Insert Table 5 Here]

In Table 5, columns 1 and 2 employ *tech-class weight* as a measure of intellectual capacity constraints, while columns 3 and 4 use *tech-class dist to core*. The KZ index (columns 1 and 3) and the LW index (columns 2 and 4) measure firm financial constraints. We also include firm size and age as additional proxies for financial constraints, following Hadlock and Pierce (2010). Our findings consistently show that a patent is more likely to become a late-bloomer when: i) it falls outside the writer's primary technology space, or ii) it is more distant from the firm's core technological competencies. Additionally, patents in relatively novel technology spaces (with fewer competing patents) are more likely to become late-bloomers. Interestingly, financial constraints do not appear to significantly influence the timing of late-bloomer patent commercialization. These results align with the characteristics of late-bloomer patent writers—typically older, larger firms with sufficient cash reserves—suggesting that these firms possess the financial flexibility to invest in innovations that may not require immediate commercialization.

More broadly, these dynamics suggest a rational response by writers, in that they have attenuated incentives to rapidly commercialize a new idea when the development cost is

[7]While Hoberg and Maksimovic (2015) pioneered the use of textual analysis to measure financial constraints, we opt for the approach developed by Linn and Weagley (2021) due to our extended sample period. Linn and Weagley (2021) employ a machine learning algorithm, specifically random decision forests, to estimate the relationship between firm-level accounting variables and financial constraints. Their method covers the period from 1972 to 2021, aligning well with our study's time frame and providing a more comprehensive measure of financial constraints for our analysis.

high—due to increased learning costs or opportunity costs from entering a new space—or when they are facing a less competitive threat. The results in Table 5 are also consistent with an intermediate solution in which late-bloomer patents represent ideas that the writers have attempted to commercialize but ultimately did not succeed, or they did not continue to pursue for similar reasons.

7.2 Patent-Complexity and Competition

In this section, we investigate the factors that may affect the value of patent hunting, focusing on two key aspects: (i) the complexity of the patent itself, which likely affects the cost of processing and integrating it into one's own innovation, and (ii) the intensity of hunting competition for a given patent. While patent complexity is often unknown before searching, high ex-post complexity could potentially reduce the overall benefits due to increased processing and integration costs. Similarly, although the level of competition in a technology space is somewhat uncertain ex-ante, intense ex-post competition (i.e., multiple firms vying to integrate the same idea) might be expected to diminish the potential benefits for any individual hunter. We explore both of these dimensions in Table 6, aiming to provide insights into how these factors shape the effectiveness and value of patent hunting strategies.

[Insert Table 6 Here]

The sample for this analysis consists of all firms in our sample that hunt late-bloomer patents. The dependent variables are late-bloomer users' next five-year sales growth and average Tobin's Q. In columns 1 and 2, we consider search costs using the Gunning fog index as a proxy for the complexity of patent text. The index quantifies document readability, with higher values indicating greater complexity. Consistent with the above dynamics, the results in columns 1 and 2 suggest that the benefits of using late-bloomer patents decline with the complexity of the targeted patent. This effect is particularly pronounced for firm value that incorporates the cost components of patent hunting. The point-estimate for sales growth is negative but not statistically significant. Columns 3 and 4 explore competition dynamics using the number of competing patent hunters in the same technology class. In both columns, patent hunter benefits decrease significantly in both sales growth and firm value when competition is more intensive to exploit similar technologies. Consistent with the search cost model in Section 3, this evidence suggests higher costs of patent hunting reduce the net benefits associated with the search strategy of hunting.

7.3 Inventor Component of Patent Hunting

Although we are measuring patent hunting at the firm level, of course it is not the entire firm that joins together to patent hunt. Instead, it is likely the R&D Team, and in fact only a subset of the R&D Team, that engage in patent hunting. Thus, in this section, we collect micro-level data on inventors, and investigate the role of individual inventors in patent hunting. In particular, we examine whether hunting benefits will be larger when the original inventors of a late-bloomer patent join a firm that hunts the patent. We also examine whether inventors who have patent-hunted in the past are more likely to do so again, regardless of who they work for. This test aims to answer if patent hunting can be also initiated by inventors who move from one firm to the other. If so, it suggests that patent hunting can be influenced by inventor-level attributes in addition to those of firms.

Table 7 presents our analysis of whether patent hunting benefits are enhanced when user firms hire the original inventors of the late-bloomer patents they cite. We extend the regression specification in Eq.(2) to include a triple interaction term with an indicator for inventor moves, focusing exclusively on late-bloomer patents. The variable *inventor move* is binary, equaling one if an inventor joins a firm that cites their patent, and zero otherwise. On average, inventors of 3.4% of late-bloomer patents subsequently move to user firms.

Our findings in columns 1 and 2 suggest that user benefits in sales growth and firm value post-influence year are significantly amplified when the user firm hires the inventors of late-bloomer patents they cite. While user benefits remain economically and statistically significant even without inventor moves, the magnitude of these benefits increases 7-8 fold when inventors do move. Furthermore, columns 3 and 4 demonstrate that these incremental benefits are even more pronounced when we focus on the sub-sample of inventors who move before their patents become influential. This timing allows the new firms to capture a larger share of the benefits from both the late-bloomer patents and the associated human capital. These results underscore the importance of not only identifying valuable patents but also acquiring the human capital behind them, suggesting a powerful synergy between patent hunting and strategic hiring in driving firm performance.

[Insert Table 7 Here]

While the results thus far show that inventors contribute to generating patent-hunting benefits, it does not necessarily suggest that they initiate patent hunting. To explore the potential for this, we turn our focus to individual inventors and examine whether an inventor who patent hunts in her current firm is also more likely to patent hunt in subsequent firms. If we find a positive association of inventors' hunting behaviors across workplaces, it is further evidence of a potential inventor-level component. We explore this possibility in Table 8.

[Insert Table 8 Here]

In the first two columns, we only consider an inventor's subsequent employer. In the last two columns, we take the average of up to three subsequent firms. The dependent variable in columns 1 and 3 is an indicator for whether the inventor is involved in any patent hunting during the subsequent employment period, while the dependent variable in columns 2 and 4 is the average number of late-bloomer patents that the inventor hunts during the subsequent employment period. We take into account the inventor's general citing behavior by controlling for their use of influential patents that include both early- and late-bloomer patents in addition to individual characteristics including gender, total number of invented patents, and total number of firms worked for during the sample period, along with their current employer's characteristics. The variables of interest are the indicator for whether the inventor's current employment is involved with any patent hunting and the total number of late-bloomer patents that the inventor's current employment has hunted.

Throughout all four columns, we find that patent hunting of an inventor in her current employment is positively and significantly associated with patent hunting of the inventor in subsequent firms. The estimated association is around 7% for the extensive margin and 14-16% for the intensive margin. We also find that when an inventor cites any influential patents in the current employment, the inventor is also more likely to hunt late-bloomer patents in subsequent workplaces. However, this association is much weaker than the association with solely late-bloomer patents by one-fourth in the magnitude in columns 1 and 3, for example.

Stepping back, these results are consistent with two interpretations of what is driving the strong relationships at the individual inventor-level: both firm- and inventor-driven. First, at the inventor level, they are consistent with patent hunting having a key inventor-level value and skill component that both follows inventors, and for whose value is increased by inventors when they implement their original ideas at subsequent firms. Second, it could be that firms specializing in patent-hunting are more likely to search out, and hire, inventors from other patent-hunting firms. Further, the more they believe in an idea, the more they invest into it, including advertising, product development, and hiring the original inventor. Although it is empirically difficult to distinguish these two explanations perfectly, the results in these inventor-level analyses highlight the important role of individual inventors in the late-bloomer patent space, along with adding patent-hunting skills and value to the firms which they join.

8 Is Patent Hunting a Sideshow?

Our central finding suggests that patent hunting is associated with sales growth and increased value for user firms, primarily by facilitating the successful commercialization of new inventions. However, a legitimate concern arises: could patent hunters like Nvidia possess inherent capabilities that predispose them to pioneering new product markets with exceptional innovations, regardless of their citation of late-bloomer patents? If so, citing these patents might be merely a formality—a courteous acknowledgment of earlier work or a compulsory addition required by patent examiners during the approval process. Alternatively, in the natural flow of technological progress, certain firms might inevitably encounter new technologies at opportune moments, enabling commercialization where their predecessors lacked the necessary complementary technologies or market demand. In both scenarios, patent hunting would coexist with the observed benefits, new product creation, and other success metrics, but might not be a necessary catalyst for this evolution. This section aims to disentangle these possibilities and establish whether patent hunting is indeed a critical driver of innovation and growth, or merely coincidental to firms' inherent innovative capabilities.

To address these potential concerns and establish the causal role of patent hunting in driving innovation and firm performance, we employ a two-pronged approach. First, we use an identification strategy shocking the supply of patent hunters available in a region, and examine this effect on the innovative process and outcomes. Second, we explore a learning curve of patent hunting, along with using a number of detailed characteristics of the patents themselves. These include Natural Language Processing (NLP) techniques to explore how the hunted patent is cited in the text of the hunter's patent itself; along with whether the patent hunter organically discovered and cited the patent versus being forced by a patent examiner to do so.

8.1 Shocks to the Supply of Patent Hunting Inventors

Despite the positive realizations associated with patent hunting, one might be concerned that these results are driven instead by some unobserved firm characteristics affecting both sales growth and patent hunting behavior, or even perhaps that expected future growth in sales or firm value (that on average materialized) prompt and allow slack to hunt for neglected ideas. To address such endogeneity concerns, we consider an identification strategy based on an instrumental variable (IV) approach that exploits our findings on hunting inventors in Section 7.3.

In Section 7.3, we show that patent hunting is partly an inherent inventor trait persisting across workplaces. This suggests that we can exploit the forced moves of patent-hunting inventors due to neighboring firms' bankruptcies as an exogenous shock to focal firms' patent hunting intensity. Doing so leverages the quasi-random nature of bankruptcies and the subsequent redistribution of talent to nearby firms. To address potential confounding effects of local bankruptcies on firm sales (e.g., reduced competition or changes in local economic conditions), we narrow our comparison to a tight set of firms with bankrupt neighbors. This ensures that all firms in our analysis are exposed to similar local economic shocks, isolating the effect of changes in patent hunting talent. Crucially, we use the relative patent hunting intensity of the bankrupting firms as our instrument, capturing the differential shock to the patent-hunter supply in nearby firms. This approach allows us to measure the varying degrees of exposure to patent hunting talent across firms, even within the same local area. Firms located near bankrupting companies with higher concentrations of patent hunters are likely to experience a larger influx of this specialized talent.

This strategy effectively orthogonalizes our analysis against concerns about the exclusion restriction related to bankrupt neighbors. By focusing solely on differences in liberated patent hunting talent, rather than the occurrence of bankruptcies themselves, we can evaluate the true impact of patent hunting on firm innovation and performance. The validity of our instrument rests on two key assumptions. First, the bankruptcy of neighboring firms is exogenous to the focal firm's innovation strategy. Second, the only channel through which these bankruptcies affect the focal firm's patent hunting activity is through the movement of patent hunting inventors. By carefully constructing our sample and instrument, we believe these assumptions are reasonably satisfied. This approach allows us to isolate the causal effect of changes in patent hunting supply, providing robust evidence on whether patent hunting truly drives innovation and firm performance, or if it is merely coincidental to other firm characteristics or broader economic trends.

In this analysis, we only consider firms in our sample located within a 100-mile radius of bankrupt neighbors and examine inventor spillovers from the bankrupt neighbors to nearby firms. The idea behind this approach is that when a neighboring firm goes bankrupt, inventors from the bankrupt firm are highly likely to join nearby firms and continue hunting late-bloomer patents at the new firms if they have previously engaged in such activities as shown in Table 8. In support of this logic, we find that approximately 20% of inventor moves occur within a 100-mile radius, both unconditionally and in association with bankruptcies. Thus, we use the hunting intensity of the bankrupt neighbor *prior to* bankruptcy as our instrument. We expect a positive correlation between the bankrupt neighbor's hunting intensity and the subsequent hunting activities of nearby firms, given the propensity of inventors from bankrupt firms to move to nearby firms. Conversely, we do not expect the bankrupt neighbor's hunting intensity to be correlated with the sales growth of nearby firms, except through the effects of inventor moves. Again, it is not that nearby bankruptcy affects sales growth. By restricting the sample for the IV analysis to solely those firms with bankrupting neighbors, we alleviate this concern and solely vary the extent to which patenting intensity of the nearby bankrupting firm has an impact on neighboring firms. We estimate the following first-stage regression for each IV and present the results in the odd-numbered columns of Table 9:

$$log(1 + LBhunting)_{jkt} = a + b_1 \ bankrupt \ neighbor \ hunting \ intensity_{jk\{t-3\}} + \gamma_j + \eta_t, \quad (3)$$

where j is user firm, k is the bankrupt neighbor, and t is year. To account for the fact that it takes time for the moving inventor to develop new patents at a new firm, we lag the instruments by three years ($t - 3$). The dependent variable is $log(1 + LBhunting)$ of firms near bankrupt neighbors. The regressions are at the user firm-year level consistent with Table 3 and include firm fixed effects (γ_i) and year fixed effects (η_t). Standard errors are clustered by the bankruptcy-area-by-year level to account for the correlation within the area (within 100 miles from the bankrupt neighbor) and year.

[Insert Table 9 Here]

We measure the bankrupt neighbor's hunting intensity in two different ways. The first two columns use the average of inventor-level hunting intensities within firm as the instrument, while the last two columns use the firm-level fraction of hunting inventors as the instrument. *bankrupt neighbor's hunting intensity* is the past three-year average before bankruptcy, considering that inventors may not file patents every year. A bankrupt neighbor's hunting intensity is set to zero when no bankruptcy occurs. When multiple bankruptcies occur within the same 100-mile radius at the same time, we use the average hunting intensity of all bankrupt firms.

We find a strong and significant first-stage effect of the IV on nearby firms' patent hunting with significance at the 1%-level in the first stage in columns 1 and 3. The F-statistics also well exceed the threshold of 10 suggested by Stock and Yogo (2005) for this test, indicating a strong instrument. We then estimate the following second-stage regression in the even-numbered columns (columns 2 and 4):

$$Sales\ growth_{j\{t,t+4\}} = a + b_1 log(1 + \widehat{LBhunting})_{jt} + \gamma_j + \eta_t. \tag{4}$$

The coefficient estimates for the instrumented patent hunting are positive and statistically significant, with stronger results when patent hunting intensity is defined at the individual inventor level rather than the firm level. An increase in patent hunting, driven exclusively by the spillover of hunting-inventors from bankrupt neighbors, significantly boosts nearby firms' future sales growth. Specifically, a one standard deviation increase in the instrumented Log(1+LBhunting) leads to a remarkable 64.3% and 40.3% increase in sales growth in columns 2 and 4, respectively. To put this in perspective, the mean five-year sales growth in our sample is 93.5%. These substantial effects underscore the economic significance of patent hunting activities. By leveraging the exogenous variation in patent hunting talent caused by neighboring bankruptcies, our IV analysis reinforces our main findings and provides robust evidence that the benefits accrued from hunting overlooked patents are indeed causal, not merely correlational. This causal link strengthens the argument for patent hunting as a critical driver of innovation and firm growth.

8.2 Valuable Patent Hunting: Learning Component, Technological Distance, and Self vs. Examiner Added Citations

In this subsection, we explore inference using a complementary set of tests, based on the following three approaches. First, we examine the learning component of patent hunting. We expect that patent hunting is indeed a skill, not a mere formality, if more experienced patent hunting leads to greater benefits. Second, we examine the technology distance between patent hunters' own patents and the late-bloomer patents that they cite. We expect that patent hunting is more technologically meaningful if the distance to cited late-bloomer patents is smaller than the distance to other cited patents. Third, using NLP of the text of the patent itself, we examine in-text mentions of late-bloomer patents. Patents mentioned in the text of the patent are both more likely to have been searched by the inventor, along with being more pivotal to their innovation. Further, we contrast the likelihood that the late-bloomer patents being added by the patent examiner versus the hunters themselves. Table 10 presents the results of these analyses.

[Insert Table 10 Here]

In Panel A, we compare the financial benefits of patent hunting by experienced and less-experienced patent hunters. We estimate the patent-level specification in Eq.(2) for subgroups of experienced users (columns 1 and 2) and less-experienced users (columns 3 and 4). Experienced users are firm-years in the top 10% distribution in the firm-level average of the late-bloomer fraction among all patents that the user cited in the past 5 years from a given citing year. The remaining firm-years are less experienced users. We find, based on the coefficient estimates on the interaction term between *user* and $iyear^{post}$, that financial benefits are significantly greater for experienced users in both sales growth and firm value. For experienced users, sales growth benefits are four times larger, and Tobin's Q benefits are significantly positive, while they are negative for less-experienced users. These results are consistent with patent hunting having a firm-level learning component.

In Panel B, we analyze the technological proximity between patent hunters' own patents and cited late bloomers compared to other cited patents by the same users. If citing a late-bloomer patent lacks intentional consideration, we expect to observe no significant difference in the distance between hunters' own patents and late-bloomer patents compared to other cited patents. However, column 1 shows that the technological distance between user patents and the cited late bloomer patents is notably closer than the distance to other cited non-late bloomer patents. We find that cited late bloomers are much closer to users' own patents than cited the bottom 95% patents in column 3, albeit slightly farther than cited early bloomers in column 2, with the difference being significantly smaller than that in column 3.

Lastly, in Panel C, we examine (i) the likelihood of being mentioned in the text part of citing patents (i.e., in-text citation rather than front-page citation), (ii) the number of in-text mentions, (iii) the sentiment of neighboring words around text mentions, and (iv) the likelihood of being cited by examiners. We compare these statistics of cited patents across late-bloomer and non-late-bloomer groups. In column 3, we find that late-bloomers, when compared to non-late-bloomer patents, are more likely to be referenced in the text part of users' own patents, have more positive sentiment when mentioned in the text part, and are significantly less likely to be added by examiners.

In Internet Appendix Figure A3, we explore further dimensions of cited late-bloomer patents including their overall impact on the broader innovation landscape after being cited and the impact of patent hunters' own patents after citing them. Panel (a) plots the growth in forward citations that late bloomers, early bloomers, and bottom 95% patents receive after they are cited by patent hunters. If patent hunters' citation of a hunted patent merely served to acknowledge an earlier work, we would not observe notably distinct subsequent impacts of these patents on the entire innovation landscape compared to other cited patents. The figure shows that hunted patents experience a 20% increase in forward citations on average during the five years after patent hunters cite them. In stark contrast, early bloomers exhibit a decline in citation growth, with the bottom 95% patents showing an even more pronounced negative trajectory. These findings suggest that citing late bloomers carries distinctive importance and should not be equated with citing other patents. Panel (b) plots the subsequent citations of patent hunters' own patents, distinguishing between their organic patents and those that use hunted patents. Organic patents are defined as patents with greater backward self-citations. The figure shows that patents hunting late bloomers in their backward citations receive markedly more forward citations over the following ten years from their grant. These findings within patent hunting firm (*e.g.,* within Nvidia's own set of patents) exhibit significantly greater influence across the wider innovation landscape for hunted patents versus their other innovative outputs.

9 Conclusion

We use the universe of patents granted over the past five decades to provide new insight into the fundamental chain of experimentation, search, and implementation that underlies the innovation process. Namely, we document large sample evidence of the importance of patent hunters—agents in the later stages of the innovation chain that engage in directed innovative search of seeking out, developing, and commercializing overlooked patents—in the eventual life-cycle of influential patents. We show that amongst all influential patents, a sizable number are characterized by "late-blooming" on which patent hunters play a key role. These late-blooming patents, even though initially overlooked, on average are more influential than early-blooming patents, and open up significant new product markets.

Patent hunters amass substantial benefits from detecting neglected patents in terms of sales growth, Tobin's Q (market value), and new products. The patents that they search out tend to be closer to their core (and are more peripheral to the writers), along with being in—at that moment they are patented—less competitive idea and innovation spaces. Patent hunting is persistent at the firm level and appears to have a learning component, as the benefits increase with patent hunting activities within firm. Patent hunting also appears to have inventor-level components, as hunted patents are more valuable when tied with original inventors, along with patent-hunting inventors continuing their hunting across workplaces.

The patent-hunting process also appears to have spillovers for the system in terms of creating more attention, innovation, and new product development in the hunted patents' idea spaces. Taken together, the results represent a new understanding of the latter stages of the innovation process, an area that is less well-understood and has received relatively less attention. Future research should continue to explore these dynamics, along with other important agents and factors that underlie ultimately successful (and unsuccessful) realizations following the initial idea and patenting stages of innovation.

References

Berman, Bruce, 2002, in *From Ideas to Assets: Investing Wisely in Intellectual Property* (John Wiley & Sons).

Bloom, Nicholas, Mark Schankerman, and John Van Reenen, 2013, Identifying technology spillovers and product market rivalry, *Econometrica* 81, 1347–1393.

Brogaard, Jonathan, Joseph Engelberg, and Edward Van Wesep, 2018, Do economists swing for the fences after tenure?, *Journal of Economic Perspectives* 32, 179–194.

Chamley, Christophe, and Douglas Gale, 1994, Information revelation and strategic delay in a model of investment, *Econometrica: Journal of the Econometric Society* 1065–1085.

Escolar, Emerson G, Yasuaki Hiraoka, Mitsuru Igami, and Yasin Ozcan, 2023, Mapping firms' locations in technological space: A topological analysis of patent statistics, *Research Policy* 52, 104821.

Fadeev, Evgenii, 2024, Creative construction: Knowledge sharing and cooperation between firms, *Working paper* .

Furman, Jeffrey L., Markus Nagler, and Martin Watzinger, 2021, Disclosure and subsequent innovation: Evidence from the patent depository library program, *American Economic Journal: Economic Policy* 13, 239–270.

Gal-Or, Esther, 1987, First mover disadvantages with private information, *The Review of Economic Studies* 54, 279–292.

Garfield, Eugene, 1970, Would Mendel's work have been ignored if the Science Citation Index was available 100 years ago?, *Essays of an Information Scientist* 1, 69–70.

Garfield, Eugene, 1980, Premature discovery or delayed recognition–why?, *Essays of an Information Scientist* 4, 488–493.

Garfield, Eugene, 1989a, Delayed recognition in scientific discovery-citation frequency-analysis aids the search for case-histories, *Current Contents* 23, 3–9.

Garfield, Eugene, 1989b, More delayed recognition. Part 1. Examples from the genetics of color blindness, the entropy of short-term memory, phosphoinositides, and polymer rheology, *Current Contents* 38, 3–8.

Garfield, Eugene, 1990, More delayed recognition. Part 2. From Inhibin to Scanning electron microscopy, *Current Contents* 9, 3–9.

Glänzel, Wolfgang, Balázs Schlemmer, and Bart Thijs, 2003, Better late than never? On the chance to become highly cited only beyond the standard bibliometric time horizon, *Scientometrics* 58, 571–586.

Glode, Vincent, and Guillermo Ordonez, 2023, Technological progress and rent seeking, *The University of Pennsylvania Working Paper Series* .

Graham, Stuart, and Deepak Hegde, 2015, Disclosing patents' secrets, *Science* 347, 236–237.

Hadlock, Charles J, and Joshua R Pierce, 2010, New evidence on measuring financial constraints: Moving beyond the kz index, *The review of financial studies* 23, 1909–1940.

Hegde, Deepak, Kyle Herkenhoff, and Chenqi Zhu, 2015, Patent publication and innovation, *National Bureau of Economic Research* .

Hoberg, Gerard, and Vojislav Maksimovic, 2015, Redefining financial constraints: A text-based analysis, *The Review of Financial Studies* 28, 1312–1352.

Hou, Jianhua, and Xiucai Yang, 2019, Patent sleeping beauties: Evolutionary trajectories and identification methods, *Scientometrics* 120, 187–215.

Kaplan, Steven N, and Luigi Zingales, 1997, Do investment-cash flow sensitivities provide useful measures of financing constraints?, *Quarterly journal of economics* 112, 169–215.

Ke, Qing, Emilio Ferrara, Filippo Radicchi, and Alessandro Flammini, 2015, Defining and identifying sleeping beauties in science, *Proceedings of the National Academy of Sciences* 112, 187–215.

Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy, 2021, Measuring technological innovation over the long run, *AER: Insights* 3, 303–320.

Kim, Jinhwan, and Kristen Valentine, 2021, The innovation consequences of mandatory patent disclosures, *Journal of Accounting and Economics* 71, 101381.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman, 2017, Technological innovation, resource allocation, and growth, *Quarterly Journal of Economics* 132, 665–712.

Lerner, Josh, and Amit Seru, 2022, The use and misuse of patent data: Issues for corporate finance and beyond, *Review of Financial Studies* 35, 1–61.

Linn, Matthew, and Daniel Weagley, 2021, Uncovering financial constraints, *Journal of Financial and Quantitative Analysis* 1–36.

Mukherjee, Abhiroop, Tomas Thornquist, and Alminas Žaldokas, 2022, New products, *Working paper* .

Pezzoni, Michele, Reinhilde Veugelers, and Fabiana Visentin, 2022, How fast is this novel technology going to be a hit? antecedents predicting follow-on inventions, *Research Policy* 51, 104454.

Sampat, Bhaven N, and Arvids A Ziedonis, 2004, Patent citations and the economic value of patents: A preliminary assessment, in *Handbook of quantitative science and technology research: the use of publication and patent statistics in studies of S&T Systems*, 277–298 (Springer).

Stock, James, and Motohiro Yogo, 2005, Testing for Weak Instrument in Linear IV Regression. In D. W. K. Andrews and J.H. Stock (eds.), *Identification and Inference for*

Econometric Models: Essays in Honor of Thomas Rothenberg pp. 80–108. Cambridge, UK: Cambridge University Press.

Strumsky, Deborah, and José Lobo, 2015, Identifying the sources of technological novelty in the process of invention, *Research Policy* 44, 1445–1461.

Trajtenberg, Manuel, 1990, A penny for your quotes: Patent citations and the value of innovations, *RAND Journal of Economics* 172–187.

Uzzi, Brian, Satyam Mukherjee, Michael Stringer, and Ben Jones, 2013, Atypical combinations and scientific impact, *Science* 342, 468–472.

Van Raan, Anthony F. J., 2004, Sleeping beauties in science, *Scientometrics* 59, 467–472.

Van Raan, Anthony FJ, 2017, Sleeping beauties cited in patents: Is there also a dormitory of inventions?, *Scientometrics* 110, 1123–1156.

Van Raan, Anthony FJ, and Jos J Winnink, 2019, The occurrence of 'sleeping beauty' publications in medical research: Their scientific impact and technological relevance, *PLoS One* 14, e0223373.

Verhoeven, Dennis, Jurriën Bakker, and Reinhilde Veugelers, 2016, Measuring technological novelty with patent-based indicators, *Research policy* 45, 707–723.

Weitzman, Martin L, 1998, Recombinant growth, *The Quarterly Journal of Economics* 113, 331–360.

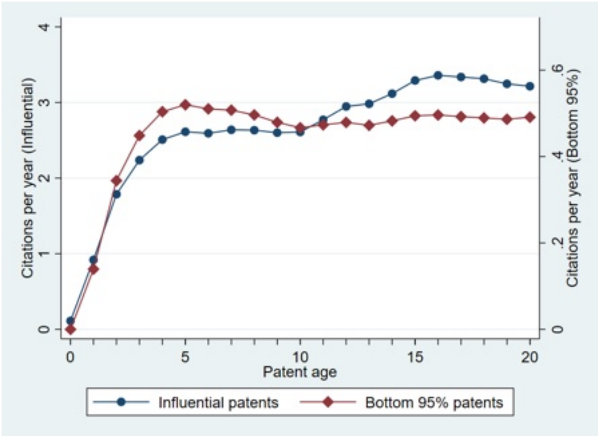
Whited, Toni M, and Guojun Wu, 2006, Financial constraints risk, *Review of Financial Studies* 19, 531–559.

Williams, Heidi L, 2017, How do patents affect research investments?, *Annual Review of Economics* 9, 441–469.

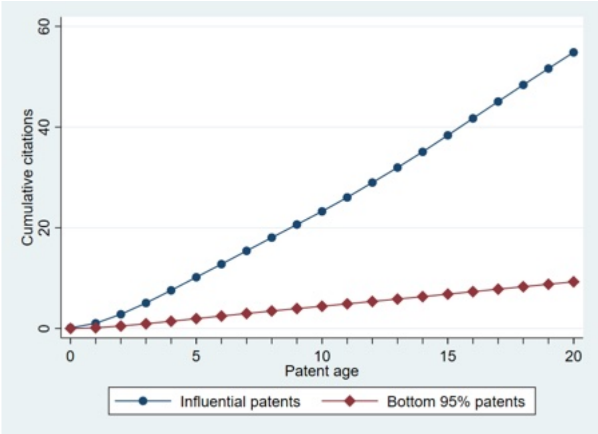
Yang, Jessie, 2014, The use and abuse of patents in the samrtphone wars: A need for change, *Journal of Law, Technology & the Internet* 5, 239–258.

Figure 1: Influential patent forward citations over time

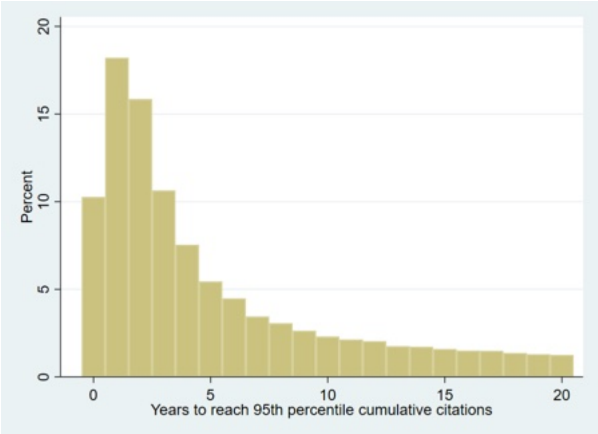
A patent, granted between 1976 and 1999, is classified as an influential patent if the cumulative citations within the CPC class and grant year cohort are in the 95th percentile or higher at any point in time during the patent term of 20 years since its grant. Panel (a) plots the average number of citations each year, excluding self-citations, over the patent age. Panel (b) plots the cumulative number of citations. Panel (c) shows the distribution of time until becoming an influential patent.



(a) Average citations



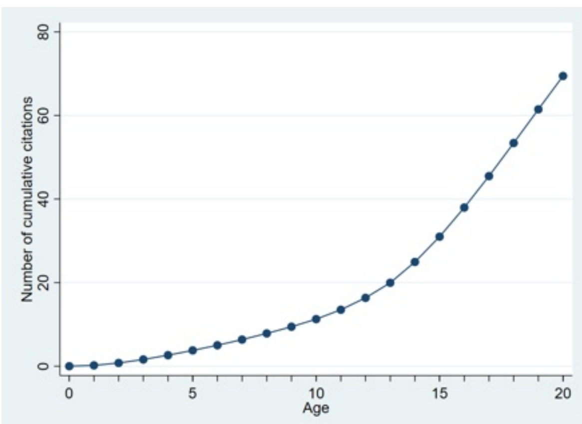
(b) Cumulative citations



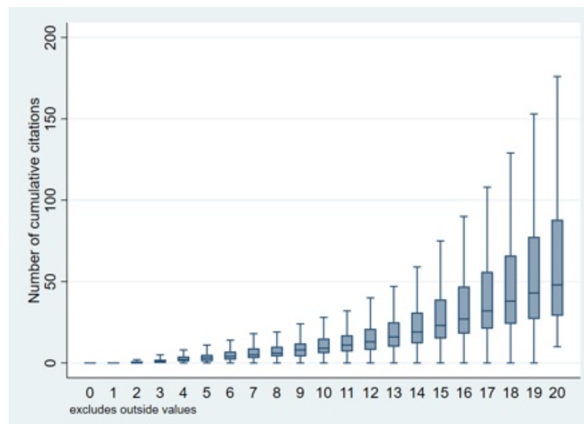
(c) Time until becoming an influential patent

Figure 2: Late-bloomer vs. early-bloomer patent cumulative citations

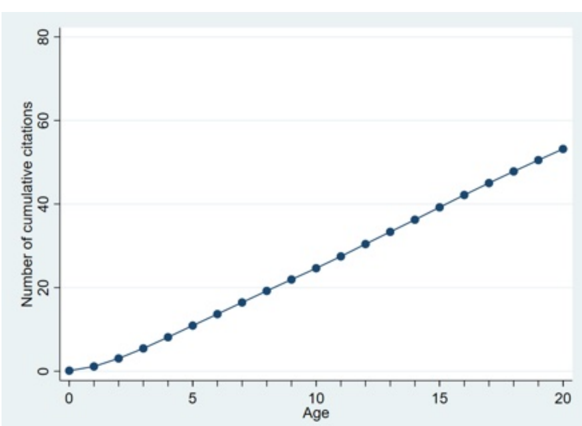
The sample consists of influential patents granted between 1976 and 1999. The figures on the left panel (a and c) plot the average number of cumulative forward citations over patent age. The figures on the right panel (b and d) present the box plots of the number of cumulative forward citations, where the mid-line and upper/lower hinge represent the median and the interquartile range between the 25th and 75th percentiles, respectively. An influential patent is a patent that has ever reached the 95th percentile of the distribution of cumulative forward citations (net of self-citations) within its cohort of the same CPC class and grant year during 20 years since its grant. A late-bloomer patent is a patent that takes an excessively long time period before it becomes an influential patent. We use the 90th percentile point in the time-to-influence distribution (14 years) to define the excessively long time period. An early bloomer is an influential patent that is not classified as a late-bloomer patent.



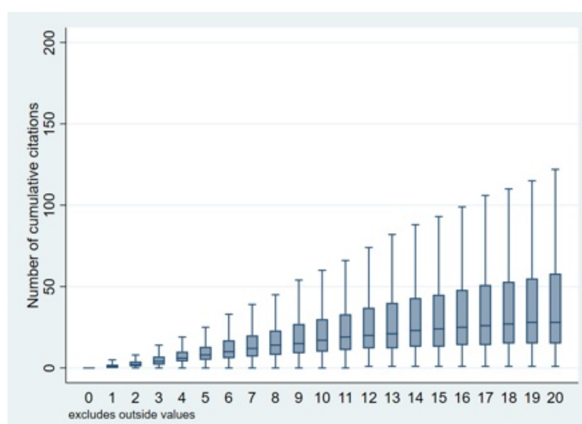
(a) Late-bloomer patents



(b) Late-bloomer patents



(c) Early-bloomer patents



(d) Early-bloomer patents

Table 1: Summary statistics

The table presents summary statistics of patent characteristics. Panel A compares influential patents and the bottom 95% patents granted between 1976 and 1999. Panel B compares late-bloomer and early-bloomer patents using the sample of only influential patents. An influential patent is a patent that has ever reached the 95th percentile of the distribution of cumulative forward citations (net of self-citations) within its cohort of the same CPC class and grant year during 20 years since its grant. A late-bloomer patent is a patent that takes an excessively long time period before it becomes an influential patent. We use the 90th percentile point in the time-to-influence distribution (14 years) to define the excessively long time period. An early bloomer is an influential patent that is not classified as a late-bloomer patent. Panel C compares citing patents of early-bloomer and late-bloomer patents. See Internet Appendix A for other variable definitions in detail. ***p<0.01, **p<0.05, *p<0.1.

Panel A: Influential vs. bottom 95% patents

	Influential patents			Bottom 95% patents			
	mean	p 50	sd	mean	p50	sd	difference
issue year	1989.90	1991.00	6.84	1989.61	1991.00	6.95	0.29***
cum. citations at age 5	10.17	8.00	10.00	1.96	1.00	2.39	8.20***
cum. citations at age 10	23.12	16.00	24.40	4.43	3.00	5.07	18.69***
cum. citations at age 15	38.04	23.00	47.16	6.84	4.00	8.23	31.20***
cum. citations at age 20	54.26	30.00	76.22	9.31	6.00	12.17	44.95***
count class	2.06	2.00	1.29	1.83	2.00	1.07	0.23***
count claims	15.85	12.00	13.98	12.27	10	10.17	3.58***
avg. claim word count	77.70	62.75	56.40	76.45	61.44	57.10	1.25***
two-examiners	0.42	0.00	0.49	0.39	0	0.49	0.03***
backward citation	12.13	8.00	15.82	9.39	7	10.52	2.74***
individual inventor	0.02	0.00	0.15	0.02	0	0.15	-0.00***
public	0.46	0.00	0.50	0.39	0	0.49	0.07***
KPSS value	11.28	3.90	31.45	9.08	3.26	23.56	2.19***
Unique number of patents		213,772			1,499,277		

Panel B: Late-bloomers vs. early-bloomers

	Late-bloomers			Early-bloomers			
	mean	p 50	sd	mean	p50	sd	difference
issue year	1989.44	1991.00	7.06	1989.95	1991.00	6.82	-0.51***
cum. citations at age 5	3.80	3.00	3.32	10.88	8.00	10.24	-7.08***
cum. citations at age 10	11.29	9.00	8.33	24.44	17.00	25.23	-13.15***
cum. citations at age 15	31.19	23.00	24.52	38.80	23.00	48.97	-7.61***
cum. citations at age 20	69.80	49.00	64.73	52.53	28.00	77.20	17.27***
count class	2.17	2.00	1.39	2.05	2.00	1.28	0.12***
count claims	16.02	13.00	14.13	15.83	12.00	13.96	0.19**
avg. claim word count	74.01	60.12	53.84	78.11	63.00	56.66	-4.09***
two examiners	0.41	0.00	0.49	0.42	0.00	0.49	-0.01**
backward citation	12.41	8.00	17.64	12.09	8.00	15.61	0.32***
individual inventor	0.02	0.00	0.14	0.02	0.00	0.15	-0.00*
public	0.46	0.00	0.50	0.46	0.00	0.50	-0.01**
KPSS value	11.15	4.31	29.61	11.29	3.85	31.65	-0.14
Unique number of patents		21,960			191,812		

Panel C: Citing patents of late-bloomers vs. and early-bloomers

	Late-bloomer citing patents			Early-bloomer citing patents			
	mean	p50	sd	mean	p50	sd	difference
issue year	2006.42	2008.00	9.20	2003.70	2004.00	10.35	2.72***
cum. citations at age 5	5.05	2.00	11.71	4.04	2.00	8.47	1.01***
cum. citations at age 10	15.39	7.00	29.49	11.05	5.00	21.08	4.34***
cum. citations at age 15	27.43	12.00	51.16	18.50	8.00	36.63	8.94***
cum. citations at age 20	35.99	16.00	69.11	23.21	10.00	47.37	12.77***
count class	2.22	2.00	1.54	2.00	2.00	1.31	0.22***
count claims	19.56	17.00	15.54	17.26	15.00	13.43	2.30***
avg. claim word count	64.91	53.45	109.04	70.34	57.30	73.48	-5.43***
two-examiners	0.37	0.00	0.48	0.39	0.00	0.49	-0.02***
backward citation	96.89	31.00	195.04	43.96	16.00	113.79	52.94***
individual inventor	0.01	0.00	0.08	0.01	0.00	0.08	-0.00**
public	0.40	0.00	0.49	0.41	0.00	0.49	-0.01***
KPSS value	16.34	5.81	42.47	13.65	4.51	38.77	2.69***
Observations		790,936			2,797,100		

Table 2: Late-bloomer writers vs. users and their persistence

Panel A compares firm-level patenting and financial characteristics of late-bloomer writers and users. Writers and users are matched by firm age. Late-bloomer writers are the firms that have produced at least one late-bloomer patent during the sample period. Late-bloomer users are the firms that have never produced a late-bloomer patent but cited at least one late-bloomer patent during the sample period. The sample consists of 3,097 firms in total with 1,892 late-bloomer writers and 1,205 late-bloomer users that are mutually exclusive. We use five nearest neighbor users for each late-bloomer writer based on the firm age. ATE stands for the average treatment effect, and SE stands for the standard error. Panel B presents the transition matrix of patent hunter status from year t to year $t+1$. Strict Writer is one for a firm year where the firm produces a late-bloomer patent but does not cite any late-bloomer patent, and zero otherwise. Flexible Writer/User is one for a firm year where the firm produces a late-bloomer patent and also cites late-bloomer patents, and zero otherwise. Strict User is one for a firm year where the firm cites late-bloomer patents but does not produce any late-bloomer patents, and zero otherwise. Idle is one for a firm year where the firm neither produces nor cites a late-bloomer patent. In each status (a) to (d), the value is the number of observations with its corresponding percentage in the brackets. See Internet Appendix A for other variable definitions in detail. ***p<0.01, **p<0.05, *p<0.1.

Panel A: Cross-sectional firm characteristics

	Late-bloomer writers	Late-bloomer Users	ATE	SE
	(1)	(2)	(3)	(4)
Log(age) without matching	2.176	1.557	0.619***	0.030
No. patents per year	29.71	2.840	26.87***	3.234
No. superstars per year	4.739	0.512	4.227***	0.485
No. late bloomers per year	0.620	0	0.620***	0.0409
No. late bloomers/no. patents	0.132	0	0.132***	0.00798
No. external cites per year	63.91	4.135	59.78***	6.330
No. external cites/no. patents	2.450	1.517	0.933***	0.112
No. claims/no. patents	16.58	17.37	-0.795**	0.346
No. new products/no. patents	0.181	0.256	-0.0749***	0.0271
Log(asset)	5.212	4.665	0.546***	0.0827
Log(sale)	4.930	4.383	0.547***	0.0968
Log(age)	1.799	1.798	0.000116	0.000170
Tobinq	2.495	2.523	-0.0281	0.0698
Salegr	0.167	0.157	0.00998	0.0102
Roa	-0.0686	-0.0651	-0.00343	0.0110
Leverage	0.192	0.196	-0.00421	0.00666
Ppe_asset	0.479	0.476	0.00335	0.0118
Rnd_sale	0.507	0.437	0.0697	0.0559
Capx_sale	0.187	0.163	0.0238	0.0174
Adv_sale	0.0113	0.0111	0.000140	0.00110
D_dv	0.425	0.384	0.0414***	0.0143
Consumer dependent	0.231	0.256	-0.0250**	0.0116

Panel B: Persistence of patent hunting

	status at $t+1$				
	Strict Writer	Flexible Writer/User	Strict User	Idle	total
status at t	(1)	(2)	(3)	(4)	
(a) Strict Writer	113 [13.12]	148 [17.19]	184 [21.37]	416 [48.32]	861 [100]
(b) Flexible Writer/User	46 [1.65]	1,709 [61.12]	788 [28.18]	253 [9.05]	2,796 [100]
(c) Strict User	118 [2.97]	832 [20.94]	2,019 [50.82]	1,004 [25.27]	3,973 [100]
(d) Idle	444 [6.07]	379 [5.18]	1,308 [17.88]	5,185 [70.87]	7,316 [100]
total	721 [4.82]	3,068 [20.53]	4,299 [28.76]	6,858 [45.89]	14,946 [100]

Table 3: Benefits from patent hunting

The table presents results from the regressions that examine financial benefits to patent users. The observations are at the firm-year level for the period of 1976 to 2020. *Sales growth* is the five-year sales growth, computed as $(sales_{t+4}/sales_t) - 1$, and $Avg\ Tobin's\ Q$ is the five-year average of Tobin's Q over t to $t+4$. *LBhunting* and *EBhunting* are the numbers of cited late-bloomer and early-bloomer patents in a given year, respectively. *earlyLBhunting* is the number of cited late-bloomer patents where the citation occurs among the first three citations on a given late-bloomer patent since its grant and also prior to its influence-year. An influence-year is the year when a given patent becomes an influential patent by reaching the 95th percentile of the distribution of cumulative forward citations (net of self-citations) within its cohort of the same CPC class and grant year during 20 years since its grant. *laterLBhunting* is the number of remaining late-bloomer patents cited. The regressions include the firm fixed effects and year fixed effects. Standard errors in parentheses are clustered at the firm level. ***p<0.01, **p<0.05, *p<0.1. See Internet Appendix A for variable definitions in detail.

	Sales growth	Avg Tobin's Q	Sales growth	Avg Tobin's Q
	(1)	(2)	(3)	(4)
log(1+LBhunting)	0.0831***	0.0658***		
	(0.0275)	(0.0168)		
log(1+EBhunting)	-0.0962***	-0.0217**	-0.0955***	-0.0250***
	(0.0181)	(0.00866)	(0.0180)	(0.00852)
log(1+earlyLBhunting)			0.0749**	0.0848***
			(0.0305)	(0.0238)
log(1+laterLBhunting)			0.0658**	0.0576***
			(0.0290)	(0.0179)
log(asset)	-0.757***	-0.309***	-0.757***	-0.309***
	(0.0366)	(0.0139)	(0.0366)	(0.0139)
log(age)	-0.443***	-0.225***	-0.443***	-0.226***
	(0.0350)	(0.0155)	(0.0351)	(0.0155)
roa	-1.837***	-0.251***	-1.837***	-0.251***
	(0.149)	(0.0435)	(0.149)	(0.0436)
leverage	-0.940***	0.00403	-0.939***	0.00523
	(0.139)	(0.0557)	(0.139)	(0.0557)
Mean	0.901	2.080	0.901	2.080
$H_0 : LB = EB$ (p-value)	0.000	0.000		
$H_0 : earlyLB = laterLB$ (p-value)			0.847	0.417
Firm FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Observations	75589	98776	75589	98776
Adjusted R^2	0.350	0.719	0.350	0.719

Table 4: Commercialization of technology

The table presents results from the regressions that examine the commercialization of technology by patent hunters. The observations are at the firm-year level in column 1 and at the focal patent-user firm-year level in columns 2, 3, and 4 from 1990 to 2015. The sample period is shorter due to the availability of the new product data. *New product count* is the total number of new products from Mukherjee et al. (2022). *Avg log*$(1+$ *new products*$)$ is the five-year average number of new products over t to $t+4$. *LBhunting* and *EBhunting* are the numbers of cited late-bloomer and early-bloomer patents in a given year, respectively. The sample in columns 2, 3, and 4 consists of late-bloomer, early-bloomer, and bottom 95% patents and their citing patents, respectively. The bottom 95% patents comprise 100,000 randomly selected patents that are not influential patents, for a comparable sample size with influential patents. *User* is an indicator variable that is one if the firm cites a focal patent and zero otherwise. *iyear* is the influence-year, defined as the year when a given patent becomes an influential patent by reaching the 95th percentile of the distribution of cumulative forward citations (net of self-citations) within its cohort of the same CPC class and grant year during 20 years since its grant. For the bottom 95% patents, *iyear* is the year with the largest number of citations during 20 years since its grant. iyear$_{post}$ is an indicator variable equal to one if the year is after *iyear* of a given patent and zero otherwise. The regression in column 1 includes the firm fixed effects and year fixed effects. The regressions in columns 2, 3, and 4 include cited patent fixed effects and year fixed effects. Standard errors are clustered at the firm level in column 1 and focal patent-by-year level in columns 2, 3, and 4. ***p<0.01, **p<0.05, *p<0.1. See Internet Appendix A for variable definitions in detail.

	Avg log(1+new products)	Log(1+new products)		
	Late-bloomers	Late-bloomers	Early-bloomers	Bottom 95%
	(1)	(2)	(3)	(4)
log(1+LBhunting)	0.0313***			
	(0.0105)			
log(1+EBhunting)	0.0124***			
	(0.00463)			
user × iyear$_{post}$		0.0472***	0.000301	-0.0286***
		(0.00842)	(0.00386)	(0.00570)
user		0.0359***	0.00362	-0.0456***
		(0.00569)	(0.00346)	(0.00495)
iyear$_{post}$		-0.0404***	-0.0106***	0.0196***
		(0.00877)	(0.00391)	(0.00513)
Control variables	Y	Y	Y	Y
Firm FE	Y	N	N	N
Cited patent FE	N	Y	Y	Y
Year FE	Y	Y	Y	Y
Observations	66636	1201198	10045142	640206
Adjusted R^2	0.706	0.626	0.898	0.669

Table 5: Constraints for late-bloomer patent writers

The table presents results from the regressions that examine potential constraints for late-bloomer writers. The sample consists of influential patents only, and the analysis compares late-bloomer and early-bloomer patents. Observations are at the patent level (one observation per patent) by taking the averages of relevant variables when a patent has multiple CPC tech classes. The dependent variable is $\mathbb{1}$(Late-bloomer) equal to one if a focal patent is a late-bloomer and zero otherwise. We consider (i) intellectual capacity constraints measured by *tech-class weight* or *tech class dist to core*, (ii) competitive threat measured by *log(competing patent stock)*, and (iii) financial constraints measured by *fin_const (KZ)*, firm age and size, or *equity_const/debt_const (LW)*. *tech-class weight* is the fraction of the patents in the CPC tech class of a given patent in all patents of its assignee over the entire sample period. *tech class dist to core* is the class-to-class proximity between the CPC tech class of a given patent and the core CPC tech class of its assignee. *log(competing patent stock)* is the log of the number of all patents from U.S. public firms with the same CPC tech class up to the grant year of a given patent. The regressions include the writer-firm fixed effects. Standard errors in parentheses are clustered at the firm and grant year levels. ***p<0.01, **p<0.05, *p<0.1. See Internet Appendix A for variable definitions in detail.

	$\mathbb{1}$(Late-bloomer)			
	(1)	(2)	(3)	(4)
tech-class weight	-0.0432***	-0.0455**		
	(0.0145)	(0.0167)		
tech-class dist to core			0.0143**	0.0142**
			(0.00534)	(0.00551)
ln(competing patent stock)	-0.00556***	-0.00534***	-0.00561***	-0.00543***
	(0.00134)	(0.00138)	(0.00131)	(0.00135)
fin_const (KZ)	-0.00701**		-0.00695**	
	(0.00296)		(0.00293)	
equity_const (LW)		-0.00913		-0.00936
		(0.00776)		(0.00776)
debt_const (LW)		-0.00177		-0.00165
		(0.00770)		(0.00770)
log_asset	0.00782*	0.00618	0.00769	0.00604
	(0.00455)	(0.00464)	(0.00454)	(0.00464)
log_age	0.000186	0.00214	0.000170	0.00221
	(0.00575)	(0.00769)	(0.00577)	(0.00768)
Control variables	Y	Y	Y	Y
Writer FE	Y	Y	Y	Y
Grant year FE	Y	Y	Y	Y
Observations	94889	86801	94889	86801
Adjusted R^2	0.033	0.033	0.033	0.033

Table 6: Costs of patent hunting

The table presents results from the cross-sectional regressions that examine the effects of costs of patent hunting on hunter benefits. The sample consists of all late-bloomer citing firms. The observations are at the focal patent-user firm-citing year level for the period of 1976 to 2020. *Sales growth* is the five-year sales growth, computed as $(sales_{t+4}/sales_t) - 1$. *Avg Tobin's Q* is the five-year average of Tobin's Q. *complexity* is the Gunning Fog Index of the text of each cited late-bloomer patent (focal patent). *competition* is the average number of competitors with overlapping CPC tech classes among all users that cite the focal patent. The regressions include the focal patent class fixed effects and year fixed effects. Standard errors in parentheses are clustered at the focal patent level. ***p<0.01, **p<0.05, *p<0.1. See Internet Appendix A for variable definitions in detail.

	Sales growth (1)	Avg Tobin's Q (2)	Sales growth (3)	Avg Tobin's Q (4)
complexity	-0.000675	-0.0191***		
	(0.00174)	(0.00387)		
competition			-0.00251***	-0.00795***
			(0.000925)	(0.00209)
log_asset	-0.0531***	-0.155***	-0.0480***	-0.143***
	(0.00305)	(0.00602)	(0.00287)	(0.00566)
log_age	-0.126***	0.0156*	-0.141***	-0.0194**
	(0.00537)	(0.00937)	(0.00528)	(0.00904)
leverage_b	-0.231***	-1.316***	-0.291***	-1.309***
	(0.0275)	(0.0566)	(0.0266)	(0.0558)
Mean	0.316	2.267	0.316	2.267
Focal patent class FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Observations	95841	116692	108953	135942
Adjusted R^2	0.117	0.220	0.117	0.212

Table 7: Hiring hunted patent inventors

The table presents results from the difference-in-differences regressions that examine the incremental financial benefits to patent users that hire cited late-bloomer patent inventors. The observations are at the focal patent-user firm-year level for the period of 1976 to 2020. *inventor move* is an indicator variable that is one if the late-bloomer inventor moves to the user firm and zero otherwise. *user* is an indicator variable that is one if the firm cites a focal patent and zero otherwise. *iyear* is the influence-year when a given patent becomes an influential patent by reaching the 95th percentile of the distribution of cumulative forward citations (net of self-citations) within its cohort of the same CPC class and grant year during 20 years since its grant. For a bottom 95% patent, *iyear* is the year with the greatest number of citations during 20 years since its grant. iyear_{post} is an indicator variable equal to one if the year is after *iyear* of a given patent and zero otherwise. The regressions include the cited-patent fixed effects and year fixed effects. Standard errors in parentheses are clustered at the cited patent-by-year level. ***p<0.01, **p<0.05, *p<0.1. See Internet Appendix A for variable definitions in detail.

	All inventor moves		Moves before influence-year	
	Sales growth	Tobin's Q	Sales growth	Tobin's Q
	(1)	(2)	(3)	(4)
inventor move × user × iyear_{post}	0.0371***	0.115***	0.0473***	0.241***
	(0.00567)	(0.0388)	(0.00658)	(0.0471)
user × iyear_{post}	0.00505***	0.0135*	0.00580***	0.0139
	(0.00126)	(0.00814)	(0.00153)	(0.00923)
iyear_{post}	-0.000185	0.0000323	-0.000185	-0.000779
	(0.00146)	(0.00628)	(0.00146)	(0.00629)
inventor move × iyear_{post}	-0.0193***	-0.0581***	-0.0263***	-0.0864***
	(0.00340)	(0.0187)	(0.00388)	(0.0224)
inventor move × user	-0.00167	0.0670***	-0.00194	0.0522*
	(0.00401)	(0.0243)	(0.00460)	(0.0277)
inventor move	0.00961***	0.0164	0.0110***	0.0257*
	(0.00232)	(0.0126)	(0.00261)	(0.0146)
user	0.0137***	0.0632***	0.0137***	0.0627***
	(0.000934)	(0.00525)	(0.000934)	(0.00525)
Control variables	Y	Y	Y	Y
Cited patent FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Observations	1501145	1510968	1485859	1495647
Adjusted R^2	0.228	0.397	0.228	0.398

Table 8: Persistent inventor-level patent hunting

The table presents results from the regressions that examine whether patent hunting has an inventor-specific component. The observations are at the inventor level. The sample consists of a universe of patent inventors, whose patents are granted between the 1976-1999 period, and their public firm employers. We only consider inventors who change jobs during the sample period. We drop the years where an inventor works for more than one employer at the same time. The dependent variable is either an indicator for whether the inventor cites late-bloomer patents at least once in the next employment ($\mathbb{1}$(late-bloomers)) or the number of late-bloomer patents that the inventor cites in the next employment (no.(late-bloomers)). Columns 1 and 2 consider only the subsequent employment, and columns 3 and 4 consider the averages of up to three subsequent employments. $\mathbb{1}$(late-bloomers)) and no.(late-bloomers) on the right-hand side of regressions are the indicators for whether a given inventor cites a late-bloomer patent in the current employment and the number of total late-bloomer patents that the inventor cites in the current employment, respectively. $\mathbb{1}$(influential) and no.(influential) are the indicators for whether a given inventor cites an influential patent in the current employment and the number of total influential patents that the inventor cites in the current employment, respectively. The regressions control for a given inventor's gender, the total number of patents that the inventor has produced, the total number of firms that the inventor works for during the sample period, and the current employer's financial characteristics. The regressions include the current employer-fixed effects and the inventor's work-start-year fixed effects. Standard errors in parentheses are clustered at the inventor level. ***p<0.01, **p<0.05, *p<0.1. See Internet Appendix A for variable definitions in detail.

	Next firm		Next three firms	
	$\mathbb{1}$(late-bloomers)	no.(late-bloomers)	$\mathbb{1}$(late-bloomers)	no.(late-bloomers)
	(1)	(2)	(3)	(4)
$\mathbb{1}$(late-bloomers)	0.0681***		0.0738***	
	(0.00606)		(0.00580)	
$\mathbb{1}$(influential)	0.0189***		0.0181***	
	(0.00453)		(0.00432)	
no.(late-bloomers)		0.136***		0.155***
		(0.0224)		(0.0230)
no.(influential)		0.00672***		0.00776***
		(0.00212)		(0.00217)
inv_gender	-0.00696	0.0130	-0.0107	0.00915
	(0.00844)	(0.0156)	(0.00903)	(0.0172)
inv_npat	0.00236***	0.00131***	0.00257***	0.00137***
	(0.000291)	(0.000230)	(0.000298)	(0.000249)
inv_nfirms	-0.0178***	-0.0108***	-0.0154***	-0.00475*
	(0.00212)	(0.00226)	(0.00223)	(0.00273)
Control variables	Y	Y	Y	Y
Current employment FE	Y	Y	Y	Y
Work start year FE	Y	Y	Y	Y
Observations	51544	51544	51544	51544
Adjusted R-squared	0.053	0.062	0.062	0.062

Table 9: Benefits from patent hunting with instrumental variable analysis

The table presents results from the instrumental variable regressions that examine sales growth benefits to patent users where we mitigate potential endogeneity issues. The observations are at the firm-year level for the period of 1987 to 2020. The sample period in this analysis is shorter because our bankruptcy data from Audit Analytics starts in 1987. We use the hunting intensity of bankrupt neighbors within a 100-mile radius as an instrument for nearby firms' late-bloomer hunting activities. Columns 1 and 2 use the average of inventor-level hunting intensities (the fraction of late-bloomer patents) within the bankrupt neighboring firm as the instrument. Columns (3) and (4) use the firm-level hunting intensity (the fraction of hunting inventors) at the bankrupt neighboring firm as the instrument. *bankrupt neighbor's hunting intensity* is the past three-year average before bankruptcy, considering that inventors may not file patents every year. A bankrupt neighbor's hunting intensity is set to zero when no bankruptcy occurs. When multiple bankruptcies occur within the same 100-mile radius at the same time, we use the average hunting intensity of all bankrupt firms. Columns 1 and 3 are the first-stage results from the regressions of patent hunting on the instrument. To account for the fact that it takes time for the moving inventor to develop new patents at a new firm, we lag the instrument by three years. We use the Cragg-Donald Wald F-stat for the weak instrument test and the Kleibergen-Paap rk statistic for the underidentification test. Columns 2 and 4 are the second-stage results from the regressions of sales growth on instrumented patent hunting. *Sales growth* is the five-year sales growth, computed as $(sales_{t+4}/sales_t) - 1$. *LBhunting* is the number of hunted late-bloomer patents in a given year. The regressions include the firm fixed effects and year fixed effects. Standard errors in parentheses are clustered at the bankrupt firm neighbor by year level. ***p<0.01, **p<0.05, *p<0.1. See Internet Appendix A for variable definitions in detail.

	Inventor-level hunting intensity		Firm-level hunting intensity	
	First stage	Second stage	First stage	Second stage
	log(1+LBhunting)	Sales growth	log(1+LBhunting)	Sales growth
	(1)	(2)	(3)	(4)
bankrupt neighbor hunting intensity	0.454***		0.236***	
	(0.0982)		(0.0539)	
instrumented log(1+LBhunting)		2.475***		1.555*
		(0.955)		(0.888)
log(asset)	0.117***	-0.865***	0.117***	-0.757***
	(0.00695)	(0.119)	(0.00695)	(0.112)
log(age)	0.0859***	-0.168*	0.0855***	-0.0882
	(0.0110)	(0.0965)	(0.0110)	(0.0912)
roa	-0.0807***	-1.592***	-0.0809***	-1.665***
	(0.0167)	(0.190)	(0.0167)	(0.187)
leverage	-0.0320	-0.736***	-0.0320	-0.762***
	(0.0266)	(0.180)	(0.0266)	(0.172)
First-stage F-stat	21.41		19.16	
Weak instrument test	21.96		18.47	
Underidentification test	19.71		16.11	
Firm FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Observations	25874	25874	25874	25874
Adjusted R-squared	0.776	0.135	0.776	0.253

Table 10: Valuable patent hunting

The table examines skills and deliberateness of patent hunting. Panel A presents results from the difference-in-differences regressions regressions that examine financial benefits to patent users by the users' hunting experience. We estimate the specifications in Eq.(2) for subgroups of experienced users (columns 1 and 2) and less-experienced users (columns 3 and 4). *Experienced* users are firm-years in the top 10% distribution in the firm-level average of the late-bloomer fraction among all patents that the user cited in the past 5 years from a given citing year. The remaining firm-years are *less experienced* users. The regressions include the cited patent fixed effects and year fixed effects. Standard errors are clustered at the patent-by-year level. Panel B examines the technology proximity between the user's citing patent and the cited patents, comparing cited late-bloomer patents to all the other cited patents (early-bloomer and bottom 95% patents). Panel C compares the following statistics for the groups of cited late-bloomer patents and all the other cited patents: (i) the likelihood of being referenced in the text part of citing patents, (ii) the number of text mentions, (iii) the sentiment of neighboring words around the text mentions, and (iv) the likelihood of being added by examiners. The measures (i), (ii), and (iii) are from Moon, Suh, and Zhou (2024). ***p<0.01, **p<0.05, *p<0.1. See Internet Appendix A for variable definitions in detail.

Panel A: Experienced hunting

	Experienced		Less experienced	
	Sales growth	Tobin's Q	Sales growth	Tobin's Q
	(1)	(2)	(3)	(4)
user × iyear$_{post}$	0.0160***	0.0993***	0.00405***	-0.0384***
	(0.00404)	(0.0282)	(0.00123)	(0.00734)
user	0.0184***	0.312***	0.0134***	0.0406***
	(0.00365)	(0.0243)	(0.000910)	(0.00483)
iyear$_{post}$	-0.0162***	-0.0586***	0.00113	0.0203***
	(0.00367)	(0.0189)	(0.00143)	(0.00596)
Control variables	Y	Y	Y	Y
Cited patent FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Observations	241937	247010	1281769	1287061
Adjusted R^2	0.222	0.410	0.241	0.424

Panel B: Proximity between hunters' and hunted patents

	Technology Proximity		
	(1)	(2)	(3)
LB	0.00795***	-0.00282*	0.0253***
	(0.00192)	(0.00159)	(0.00293)
patage_f	-0.00213***	-0.00205***	-0.00146***
	(0.0000930)	(0.000161)	(0.0000897)
Comparison group	EB, bottom 95%	EB only	bottom 95% only
Citing patent FE	Y	Y	Y
Tech class FE	Y	Y	Y
Observations	2938358	1812704	1546350
Adjusted R^2	0.523	0.563	0.508

Panel C: Deliberate hunting

	Late-bloomers	Non-late-bloomers	(1)-(2)
	(1)	(2)	(3)
$\mathbb{1}$(in-text cited)	0.0552	0.0446	0.0106***
No.(in-text mentions)	0.0676	0.0558	0.0117***
Sent(in-text mentions)	0.373	0.329	0.0438***
$\mathbb{1}$(examiner cited)	0.178	0.362	-0.184***

Internet Appendix to "Patent Hunters"

Lauren Cohen, Umit G. Gurun, S. Katie Moon, and Paula Suh

A Variable Definition

Variable Name	Definition
Influential patent	A patent that has ever reached the 95th percentile cumulative forward citations (net of self-citations) within the CPC class-grant year cohort.
Late-bloomer patent	A patent that takes more than 14 years (the 90th percentile in time-to-influence distribution) to become an influential patent.
Early-bloomer patent	An influential patent that is not a late-bloomer patent.
Cum. citations	The cumulative number of forward citations net of self-citations.
Count class	The number of unique technology classes.
Count claims	The number of claims.
Avg. claim word count	The average number of words in claims.
Two-examiners	An indicator variable equal to one if a patent was reviewed by two examiners and zero otherwise.
Backward citation	The number of backward citations.
Individual inventor	An indicator variable equal to one if the patent is assigned to an individual and zero otherwise.
Public	An indicator variable equal to one if the patent is assigned to a public firm and zero otherwise.
KPSS value	Kogan et al. (2017) value of patent.
no. patents per year	The number of granted patents of the firm in a year.
no. influential patents per year	The number of granted patents of the firm that become influential patents in a year.
no. late-bloomers per year	The number of granted patents of the firm that become late-bloomer patents in a year.
no. external cites per year	The number of citations received in a year net of self-citations.
no. external cites/no. patents	The total number of citations received (net of self-citations) scaled by the total number of patents.
no. claims/no. patents	The total number of claims scaled by the total number of patents.
no. new products/no. patents	The total number of new product launches scaled by the total number of patents.
log(asset)	The logarithm of total assets.
log(sale)	The logarithm of total assets.
log(age)	The logarithm of firm age.
tobin's Q	The Tobin's Q ratio, calculated as the market value of a company divided by the total assets.
sales growth	Logarithm of the total revenues divided by the previous year's total revenues.
roa	Return on assets, calculated as the net income divided by the total assets.
leverage	The debt-to-assets ratio, calculated as total debt divided by total assets.
ppe_asset	Tangible fixed assets (Property, Plant, and Equipment) scaled by the total assets.
rnd_asset	R&D expense scaled by the total assets.
capx_asset	Capital expenditure scaled by the total assets.
adv_asset	Advertising expense scaled by the total assets.
d_dv	An indicator variable equal to one if the firm pays dividends.
consumer dependent	An indicator for consumer-dependent industries whose production percentage for "Personal consumption expenditures" in the 2002 Input-Output Accounts from the Bureau of Economic Analysis is in the top tercile.
log(totalpat)	The logarithm of the total number of U.S. public firm patents in a given year.
cumulative products	The cumulative number of new product launches since the beginning of the Mukherjee et al. (2022) data set up to $t-1$.
fin const (KZ)	The Kaplan-Zingales index based on the five-factor model in Kaplan and Zingales (1997).
fin const (WW)	The Whited-Wu index from Whited and Wu (2006).
inventor move	An indicator variable that is one if the late-bloomer inventor moves to the user firm and zero otherwise.
ninv	Number of inventors in the citing firm.
inv_gender	An indicator variable equal to one for male inventors and zero for female inventors.
inv_npat	The total number of the patents that the inventor produced during the sample period.
inv_nfirms	The total number of firms that the inventor worked for during the sample period.
dlog(asset)	The difference in the 10-year log(asset) since the citing year between firms citing and writing the focal patent.
dlog(age)	The difference in the 10-year log(age) since the citing year between firms citing and writing the focal patent.
droa	The difference in the 10-year roa since the citing year between firms citing and writing the focal patent.
dleverage	The difference in the 10-year leverage since the citing year between firms citing and writing the focal patent.

2

Figure A1: Late-Bloomer patent example

The figures show Texas Instrument(TI)'s patent (#5,025,407) and its relation to Nvidia's late-bloomer patent hunting benefits. Panel (a) shows the front page bibliographic data and some exhibits from TI's patent grant. It contains the invention title, assignee names, backward citations, and an abstract. Panel (b) illustrates the forward citations that the patent #5,025,407 has received over the 20-year patent term since its grant year. Each mark represents a forward citation, and the letter inside refers to citing firms' initials. A dotted circle is a foreign assignee/foreign assignee's U.S. subsidiary/university. A square and circle each denote a citation made by a late-bloomer writer and user, respectively. A circle without a letter denotes a citation made specifically by Nvidia, who is a patent hunter (*i.e.*, late-bloomer user). 2006 is the influence-year, defined as the year when the patent's cumulative citations reach the 95th percentile within its cohort of the same CPC class and grant year. Panel (c) presents Nvidia's stock prices over the 2000-2017 period. Panel (d) shows the video game industry revenues by segment over a similar period between 2002 and 2019.

United States Patent [19]
Gulley et al.

[11] **Patent Number:** **5,025,407**
[45] **Date of Patent:** **Jun. 18, 1991**

[54] **GRAPHICS FLOATING POINT COPROCESSOR HAVING MATRIX CAPABILITIES**

[75] Inventors: **David W. Gulley; Jerry R. Van Aken,** both of Sugar Land, Tex.

[73] Assignee: **Texas Instruments Incorporated,** Dallas, Tex.

[21] Appl. No.: **387,459**

[22] Filed: **Jul. 28, 1989**

[51] **Int. Cl.**[5] **G06F 7/52**
[52] **U.S. Cl.** **364/754;** 364/736
[58] **Field of Search** 364/754, 736, 748, 518

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,763,365	10/1973	Seitz	364/754
4,493,048	1/1985	Kung et al.	364/754
4,697,247	9/1987	Grinberg et al.	364/754
4,719,588	1/1988	Tatemichi et al.	364/754
4,878,190	10/1989	Darley et al.	364/752
4,916,651	4/1990	Gill et al.	364/736

OTHER PUBLICATIONS

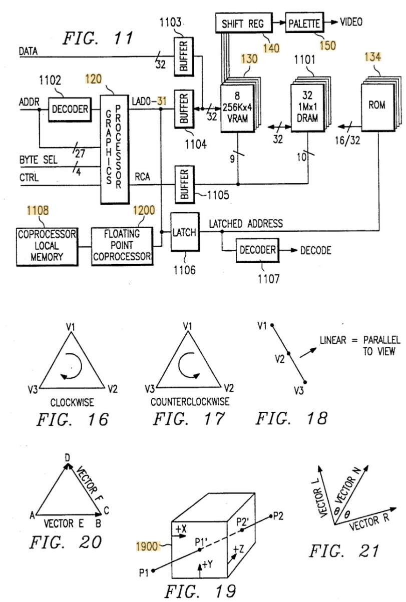
Mokhoff, N., Graphics Chips Forge High-Res Boards for PCs, Workstations, *Electronic Design,* Mar. 17, 1988, Design Trade-Offs in High-End Graphics Board, *Electronic Design,* Mar. 17, 1988, pp. 77-84.
Foley, J. D., and A. Van Dam, *Fundamentals of Interactive Computer Graphics,* Reading, Mass: Addison-Wesley, 1982, pp. 245-265, 274-279, 297-302.
Newman, W. M., and R. F. Sproull, *Principles of Interactive Computer Graphics,* 2nd ed., New York: McGraw-Hill, 1979, pp. 57-60, 333-351, 491-501.

Primary Examiner—Dale M. Shaw
Assistant Examiner—Long T. Nguyen
Attorney, Agent, or Firm—James F. Hollander; James T. Comfort; Melvin Sharp

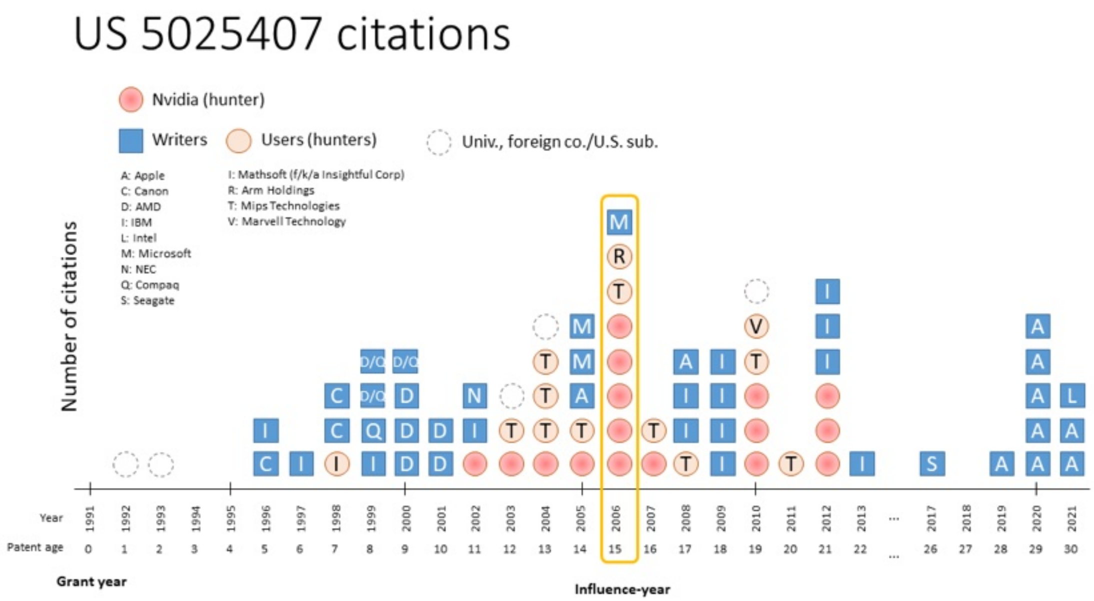
[57] **ABSTRACT**

A graphics coprocessor designed to work in conjunction with a host graphic processor in a graphics system. The coprocessor is adapted to perform arithmetic calculations including matrix calculations. The matrix size is such that the intermediate results require more registers than are practical to include in the coprocessor. This has been solved by arranging for certain selected ones of the intermediate results to continue within the program execution from stage to stage and avoiding intermediate storage.

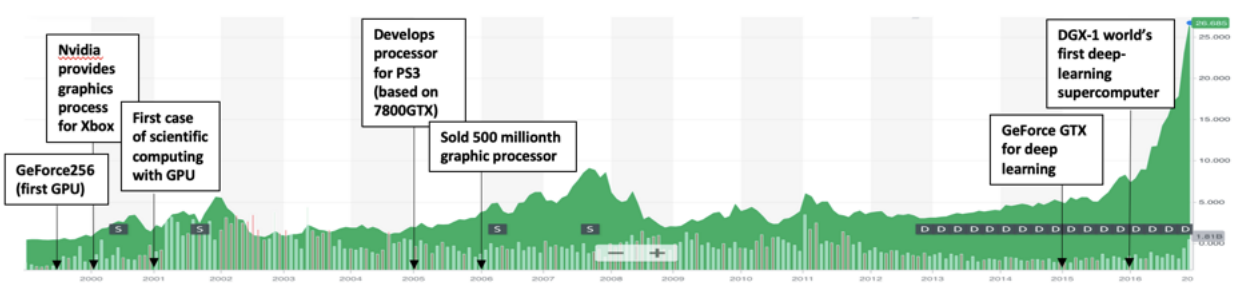
U.S. Patent June 18, 1991 Sheet 4 of 8 **5,025,407**



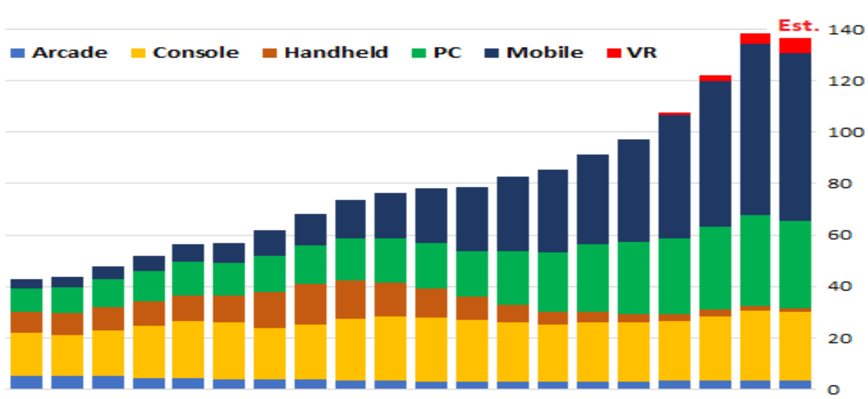
(a) Texas Instrument's Patent #5,025,407



(b) Patents Citing #5,025,407



(c) Nvidia Stock Prices (2000-2017)



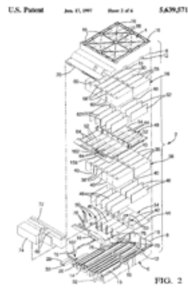
(d) Video Game Industry Revenues ($bn, 2002-2019)

Figure A2: Other hunting examples

Battery pack

DirecTV 1997, No. 5639571
Tesla citing extensively from 2012

A battery pack for easy access to, and uniform cooling/heating of, the individual battery modules thereof. The pack comprises stackable housing parts (i.e., top and bottom) housing multiple tiers of battery modules supported by underlying trays having openings/holes therein aligned with gaps/spaces between adjacent battery modules through which cooling/heating air is uniformly flowed in parallel between the modules from an underlying plenum. The battery modules are compressively immobilized in the housing by resilient foam pads which bear down on the tops of the modules.

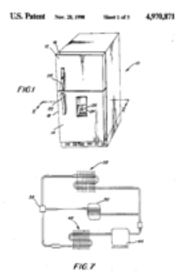


(a) Tesla cites DirecTV

Carbonator refrigeration system

Coca-Cola 1990, No. 4970871
Whirlpool citing extensively from 2000

A carbonator refrigeration system for use in a conventional refrigerator for dispensing a chilled carbonated liquid such as water or a beverage from the front door of the refrigerator. The system includes a compressor, an evaporator, a condenser, a carbonator and a valve member wherein the valve member is responsive to conditions detected within the refrigerator for selectively directing a source of cooling fluid to or away from a heat exchange device provided in connection with the carbonator. The carbonator refrigeration system enables cooling of the carbonator for home dispensing use in a time-share manner with the remaining mechanical refrigeration components.

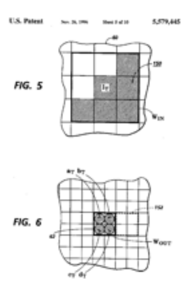


(b) Whirlpool cites Coca-Cola

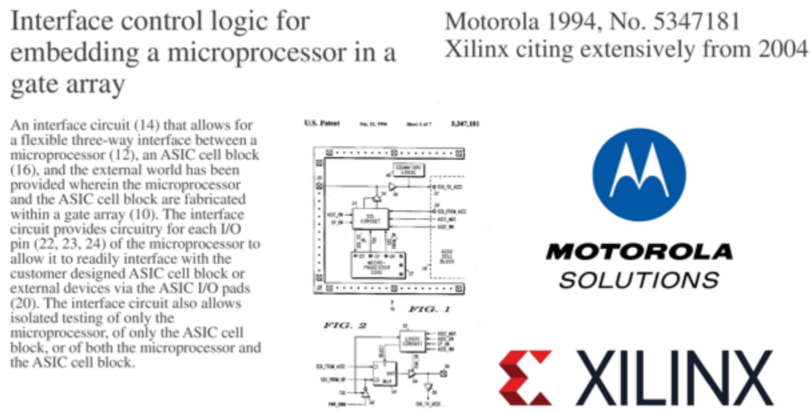
Image resolution conversion method that employs statistically generated multiple morphological filters

Xerox 1996, No. 5579445
Adobe citing extensively from 2012

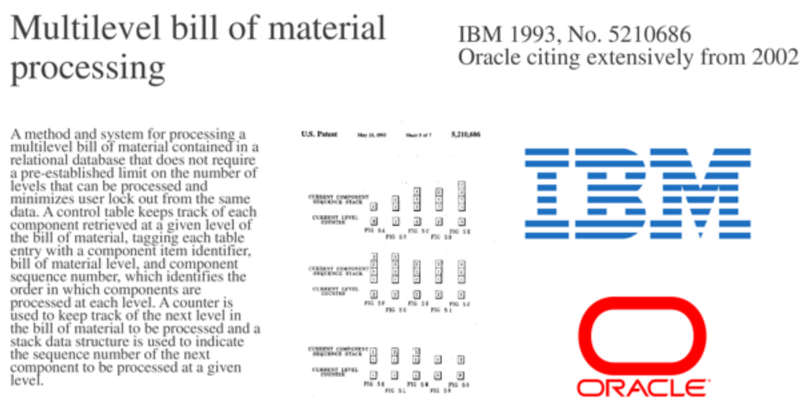
A method and apparatus for automating the design of morphological or template-based filters for print quality enhancement. A plurality of different phase, but same resolution, subsampled images are generated from training documents. Statistical data derived therefrom is then employed in an automated process to generate filters. The filters may be used for resolution enhancement and/or conversion of bitmap images. Furthermore, the statistical data is used to produce filters that are intended to not only optimize image structure, but image density as well.



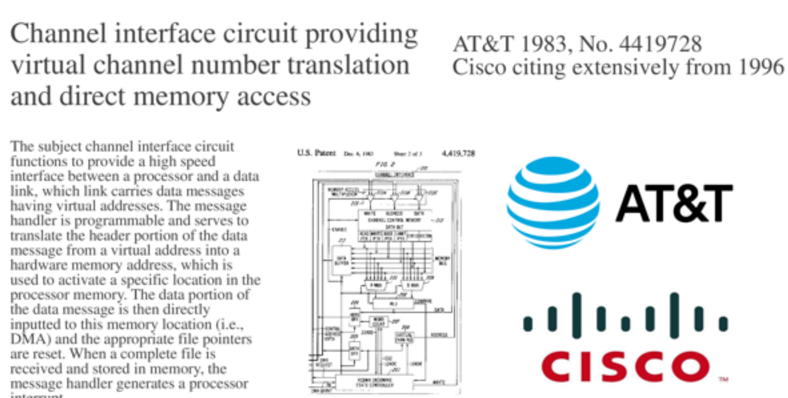
(c) Adobe cites Xerox



(d) Xilinx cites Motorola

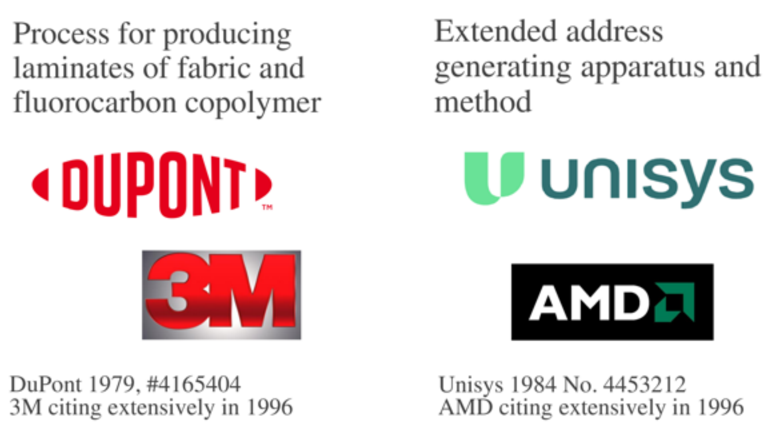


(e) Oracle cites IBM



(f) Cisco cites AT&T

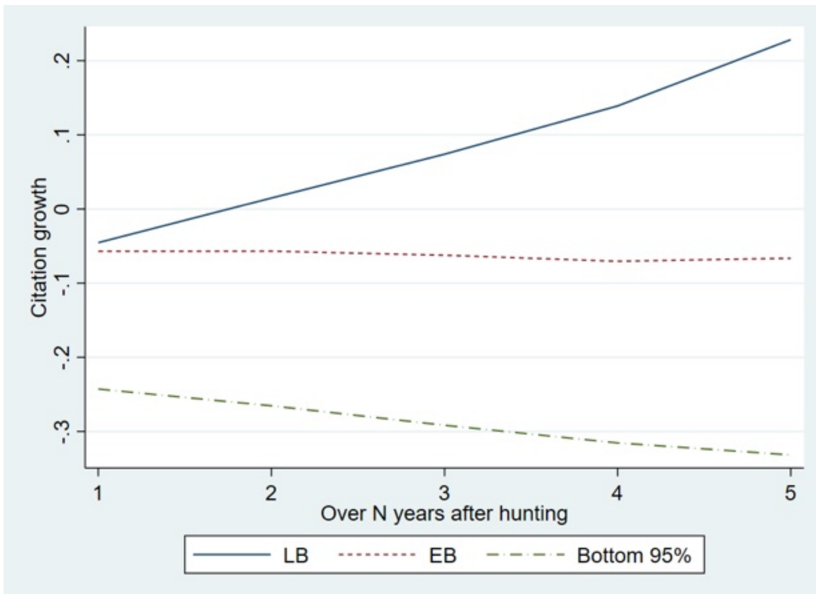
(g) Clorox/Ecolab cite P&G



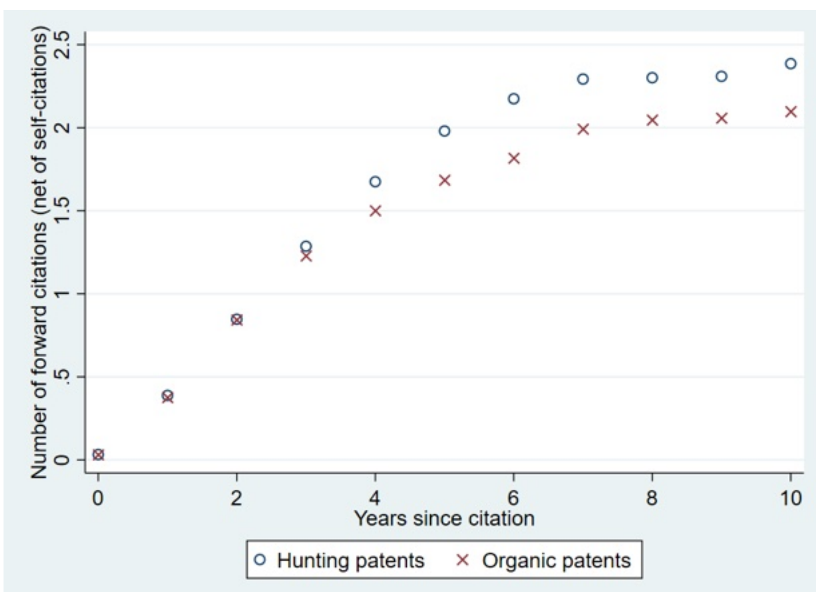
(h) 3M/AMD cite DuPont/Unisys

Figure A3: Citation growth of hunted patents and hunters' own patents

The figures show the forward citation growth of the hunted patents and hunters' own patents since the citing year. In Panel A, the sample consists of all user-cited patent pairs in our sample between 1976 and 2020. Citation growth is measured over the N years from each user's citation year. In Panel B, the sample consists of all hunters' own patents that cite at least one late-bloomer patent. The figure plots the subsequent citations of patent hunters' own patents, comparing their organic patents and those that use late-bloomer patents (*i.e.,* hunting patents). Organic patents are defined as patents with greater backward self-citations compared to the number of late-bloomer citations.



(a) Citation growth of hunted patents



(b) Citations of hunters' organic patents vs. hunting patents

Table A1: Sample selection

Our base patent sample consists of all USPTO patents granted between 1976 and 1999. The table describes our sample-selection procedure and method of classifying patents into influential patents, late-bloomer patents, and early-bloomer patents with the number of observations in each group.

	Number of patents	Description
Base patent sample	1,712,247	All USPTO patents granted between 1976 and 1999. The sample period starts in 1976 due to the availability of data on patent assignees, inventors, claims, and other information from the PatentsView database. The sample period ends in 1999 as identifying an influential patent requires 20 years since each patent's grant year. We also exclude approximately 0.45% of the remaining patents from the sample when they have no CPC information.
Influential patents	213,772	An influential patent is the patent that has ever reached the 95th percentile of the distribution of cumulative forward citations (net of self-citations) within its cohort of the same CPC class and grant year at any point in time during 20 years since its grant.
Late-bloomers	21,960	A late-bloomer patent is a patent that takes an excessively long time period before it becomes an influential patent. We use the 90th percentile point in the time-to-influence distribution (14 years) to define the excessively long time period.
Early-bloomers	191,812	An early bloomer is an influential patent that is not classified as a late-bloomer patent.

Table A2: Late-bloomer patent writers and users

The table shows the list of the top 20 late-bloomer patent writers and users. The late-bloomer writer and user definitions are consistent with Table 2. Late-bloomer writers are the firms that have produced at least one late-bloomer patent during the sample period. Late-bloomer users are the firms that have never produced a late-bloomer patent but cited at least one late-bloomer patent during the sample period.

Rank	Writers	Users
1	U S SURGICAL CORP	PARKERVISION INC
2	JOHNSON & JOHNSON	WEATHERFORD INTL PLC
3	INTL BUSINESS MACHINES CORP	JPMORGAN CHASE & CO
4	MOTOROLA SOLUTIONS INC	TIVO CORP
5	3M CO	AMKOR TECHNOLOGY INC
6	HITACHI LTD	IMMERSION CORP
7	GENERAL ELECTRIC CO	SMITH & NEPHEW PLC
8	AT&T CORP	BROADCOM CORP
9	CANON INC	LENNOX INTERNATIONAL INC
10	MEDTRONIC PLC	BLACKBERRY LTD
11	EASTMAN KODAK CO	FAIRCHILD SEMICONDUCTOR INTL
12	TEXAS INSTRUMENTS INC	RESMED INC
13	PANASONIC HOLDINGS CORP	ARTHROCARE CORP
14	HP INC	SANDISK CORP
15	SONY GROUP CORPORATION	NIKE INC
16	DONNELLY CORP	ICU MEDICAL INC
17	NEC CORP	LIFE TECHNOLOGIES CORP
18	PROCTER & GAMBLE CO	AFFYMETRIX INC
19	XEROX HOLDINGS CORP	NETAPP INC
20	APPLE INC	TYCO INTERNATIONAL PLC

Table A3: Benefits from patent writing

The table presents results from the regressions that examine financial benefits to patent writers. The observations are at the firm-year level for the period of 1976 to 2020. *Sales growth* is the five-year sales growth, computed as $(sales_{t+4}/sales_t) - 1$, and $Avg\ Tobin's\ Q$ is the five-year average of Tobin's Q over t to $t+4$. *LBwriting* and *EBwriting* are the numbers of written late-bloomer and early-bloomer patents in a given year, respectively. *log*(*LBwriting*), *log*(*EBwriting*), *log*(*LBhunting*), and *log*(*EBhunting*) are standardized for ease of comparisons. The regressions include the firm fixed effects and year fixed effects. Standard errors in parentheses are clustered at the firm level. ***p<0.01, **p<0.05, *p<0.1. See Appendix A for variable definitions in detail.

	Writer benefits		User benefits	
	Sales growth	Avg Tobin's Q	Sales growth	Avg Tobin's Q
	(1)	(2)	(3)	(4)
log(1+LBwriting)	0.00841	0.0199***		
	(0.0104)	(0.00754)		
log(1+EBwriting)	-0.00150	0.0707***		
	(0.0211)	(0.0126)		
log(1+LBhunting)			0.0553***	0.0437***
			(0.0183)	(0.0112)
log(1+EBhunting)			-0.129***	-0.0292**
			(0.0243)	(0.0116)
log(asset)	-1.036***	-0.286***	-0.757***	-0.309***
	(0.0511)	(0.0173)	(0.0366)	(0.0139)
log(age)	-0.475***	-0.219***	-0.443***	-0.225***
	(0.0511)	(0.0184)	(0.0350)	(0.0155)
roa	-2.072***	-0.415***	-1.837***	-0.251***
	(0.217)	(0.0522)	(0.149)	(0.0435)
leverage	-1.108***	-0.0693	-0.940***	0.00403
	(0.182)	(0.0625)	(0.139)	(0.0557)
Mean	0.950	1.924	0.901	2.080
$H_0: LB = EB$ (p-value)	0.676	0.000	0.000	0.000
Firm FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Observations	45749	58030	75589	98776
Adjusted R^2	0.441	0.815	0.350	0.719

Table A4: Benefits from patent hunting (patent level)

The table presents results from the difference-in-differences regressions that examine financial benefits to patent users. The observations are at the focal patent-user firm-year level for the period of 1976 to 2020. The samples in columns 1 and 2, 3 and 4, and 5 and 6 consist of late-bloomer, early-bloomer, and bottom 95% patents and their citing patents, respectively. *user* is an indicator variable that is one if the firm cites a focal patent and zero otherwise. *iyear* is the influence-year, defined as the year when a given patent becomes an influential patent by reaching the 95th percentile of the distribution of cumulative forward citations (net of self-citations) within its cohort of the same CPC class and grant year during 20 years since its grant. For the bottom 95% patents, *iyear* is the year with the largest number of citations during 20 years since its grant. iyear_{post} is an indicator variable equal to one if the year is after *iyear* of a given patent and zero otherwise. The bottom 95% patents comprise 100,000 randomly selected patents that are not influential patents, for a comparable sample size with influential patents. The regressions include the cited-patent fixed effects and year fixed effects. Standard errors in parentheses are clustered at the patent-by-year level. ***p<0.01, **p<0.05, *p<0.1. See Appendix A for variable definitions in detail.

	Late-bloomers		Early-bloomers		Bottom 95% patents	
	Sales growth	Tobin's Q	Sales growth	Tobin's Q	Sales growth	Tobin's Q
	(1)	(2)	(3)	(4)	(5)	(6)
user × iyear_{post}	0.00730***	0.0179**	0.000159	-0.0193***	0.00146	-0.0482***
	(0.00125)	(0.00816)	(0.000570)	(0.00363)	(0.000922)	(0.00542)
user	0.0143***	0.0644***	0.00732***	0.0191***	0.0146***	0.0308***
	(0.000923)	(0.00525)	(0.000511)	(0.00327)	(0.000810)	(0.00468)
iyear_{post}	-0.00146	-0.000481	-0.00408***	0.0157***	-0.000298	0.0128***
	(0.00145)	(0.00625)	(0.000573)	(0.00301)	(0.000856)	(0.00384)
Mean	0.047	2.075	0.057	2.020	0.046	2.022
Control variables	Y	Y	Y	Y	Y	Y
Cited patent FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Observations	1523717	1534074	11017213	11086629	759482	763142
Adjusted R^2	0.226	0.386	0.287	0.452	0.261	0.490

Table A5: Hunting benefits without business relationship

The table presents a re-estimation of the patent-level difference-in-differences regressions in Table A4, which examines financial benefits to late-bloomer users, by additionally considering business relationships. The sample consists of only non-relationship-based citation pairs by excluding citation pairs of firms in any business relationships of partners, suppliers, and customers. The firm relationship data is obtained from FactSet Revere and Compustat customer segment data between 2003 and 2017. The observations are at the focal patent-user firm-year level for the citations made between 2003 and 2017 due to the availability of the business relationship data. *User* is an indicator variable that is one if the firm cites a focal patent and zero otherwise. *iyear* is the influence-year, defined as the year when a given patent becomes an influential patent by reaching the 95th percentile of the distribution of cumulative forward citations (net of self-citations) within its cohort of the same CPC class and grant year during 20 years since its grant. iyear_{post} is an indicator variable equal to one if the year is after *iyear* of a given patent and zero otherwise. The regressions include the cited patent fixed effects and year fixed effects. Standard errors in parentheses are clustered at the patent-by-year level. ***p<0.01, **p<0.05, *p<0.1. See Appendix A for variable definitions in detail.

	Sales growth (1)	Tobin's Q (2)
user × iyear_{post}	0.00956***	0.0691***
	(0.00196)	(0.0132)
user	0.0156***	0.0525***
	(0.00156)	(0.00908)
iyear_{post}	-0.00356	-0.0165*
	(0.00221)	(0.00925)
Control variables	Y	Y
Cited patent FE	Y	Y
Year FE	Y	Y
Observations	639088	645331
Adjusted R^2	0.216	0.379