Search Speed and Theory Discovery: A Model for Innovation

Arnaldo Camuffo Alfonso Gambardella Andrea Pignataro *

November 25, 2024

Abstract

We develop a model for analyzing firm innovation strategies, distinguishing between the exploitation of established firm-level paradigms (which we call "theories") and the exploration of new paradigms. Central to the model is the role of speed. The empirical analysis of a sample of all US public firms with at least one patent between 1980 and 2021, reveals that the speed at which firms exploit innovations within their existing paradigms is positively correlated with their exploration of new technological areas, an increase in patent output relative to R&D, and an increase in firm size over time. The model identifies four key drivers of innovation success: search speed within existing theories, the probability of discovering new theories, the potential to generate new paradigms, and the scale of firm resources. High search speed boosts early-stage productivity and resource accumulation, increasing the likelihood of discovering new theories and sustaining growth. Among other things, our model provides an explanation for the hyper-growth of many high-tech companies today. Policymakers should consider interventions that accelerate search-enhancing technologies, and foster theory generation across industries, promoting equitable growth and reducing disparities in innovation capacity.

JEL: L21, L26, M13, M21

Keywords: Innovation, exploration, AI, speed, firm strategies, growth

*Camuffo: Department of Management & Technology, Bocconi University and ION Management Science Lab, *arnaldo.camuffo@unibocconi.it*. Gambardella: Department of Management & Technology and ION Management Science Lab, Bocconi University, *alfonso.gambardella@unibocconi.it*. Pignataro: ION Group, *ap@iongroup.com*.

1 Introduction

Innovation is a key driver of economic growth. At the micro-level, there is extensive literature on innovation (Hall & Rosenberg , 2010; Hall & Helmers , 2024). At the macro-level, the new growth literature, stemming from Romer (1990) and Aghion & Howitt (1992), places innovation at the center of the growth processes for countries, industries, and firms.

Surprisingly, however, neither the micro nor macro literature has thoroughly explored how changes in the speed of innovation impacts firms, industries, and economies. The issue is relevant. For instance, Rosenberg (1992) showed that technical advances in scientific instrumentation increased the speed of innovation in industries, enhancing innovation opportunities. Similarly, several technologies, such as computers, have accelerated the pace of activities, including innovation search and the generation of innovations. Today's artificial intelligence (AI) holds the promise of enhancing these processes to a considerable extent.

This paper introduces a framework for analyzing firm innovation strategies through the lens of search processes. We combine speed with a fundamental feature of innovation search: firms either search for innovations within established theories or seek to discover new theories before searching within them. Building on Dosi's (1982) paradigm shift concept, which parallels Kuhn's process of scientific revolutions, we view theories as firm-level paradigms. In our dynamic model, firms search for innovations in a staged process, allocating a given amount of resources between searching within current theories and exploring new ones. Discovering a new theory resets the process, while searching within a theory gradually exhausts innovations, prompting firms to shift resources toward discovering new theories. Higher search speeds exhaust theories more rapidly, leading to earlier and larger resource allocations toward new theory discovery.

The model predicts that firm size and innovation are influenced by search speed, the propensity for basic research, and the resources that they can deploy into this process. It highlights how firms of different sizes prioritize basic versus applied research, thereby shaping their growth strategies. Firms with high search speed, more basic research propensity, and size tend to favor radical innovations. Additionally, the model identifies nonlinear effects of search speed. At low speeds, focusing on basic research can be inefficient, as firms risk moving on from current theories too early. However, at high speeds, rapidly finding new theories becomes essential, as existing opportunities are quickly depleted, maximizing returns.

Our model examines a single firm's knowledge production, setting aside competition, spillovers, and other elements that could enrich understanding of innovation speed. We also simplify by assuming that more knowledge yields higher economic returns, without detailing the conversion process. While these factors are important, our analysis concentrates on the primary mechanisms of speed, reserving broader considerations for future research.

Section 2 reviews the literature. Section 3 provides motivating evidence. Sections 4 and 5 present our model and its implications, while Sections 6 and 7 offer discussion and conclusions. The Appendices include derivations, proofs, and the construction of our measure of patent clusters.

2 Related Literature

The framework of our paper sheds light on several puzzles in innovation and economic growth. We address Christensen's (1997) paradox: Despite efficient resource allocation, established firms struggle with disruptive innovations due to difficulty in identifying new opportunities (selection problem) and the threat to existing businesses (cannibalization problem). The model underscores the significance of search speed. A slower search process pushes firms to focus on exploiting current theories, limiting their capacity to explore new avenues. On the other hand, a faster search pace rapidly exhausts available opportunities, making it advantageous for firms to pursue new theoretical frameworks.

This finding aligns with Levinthal & March's (1993) concept of "myopia of learning," where firms specializing in a slow search environment within a bounded domain, experience diminishing returns but adopt this focus as an optimal strategy under the constraints of slow search. Slow search restricts firms to exploiting existing knowledge, hindering the exploration of new areas.

Bloom et al. (2020) showed that "ideas are harder to find" despite significant innovation resources, consistently with our model's predictions under slow search technology. Notably, our model suggests that it is not a lack of inherent capability that hinders new ideas but rather the optimal resource allocation towards existing trajectories when search is slow. This implies that slow search allows for continued investment within an established framework, even as the innovation rate declines, delaying the need to explore entirely new theories.

A growing body of literature explores AI's impact on innovation and growth (Furman & Seamans, 2019; Agrawal et al., 2019; McElheran et al., 2024). AI, characterized as a general-purpose technology (Cockburn et al., 2019; Trajtenberg, 2019) has the potential to significantly increase search speed (Agrawal et al., 2024). AI's ability to dramatically accelerate search speed creates conditions for a paradigm shift. Additionally, Ludwig & Mullainathan (2024) argue that generative AI enhances the generation of novel representations, extending AI's influence beyond accelerating search speed to improving the ability to envision new theories.

We contribute to the growth literature (Romer, 1990; Aghion & Howitt, 1992) by exploring the micro-foundations of knowledge as an asset and its implications for growth. Building on Weitzman's (1979 & 1998) work on knowledge formation and recombination, our model addresses the broader concept of theories and the interplay between search within and across them, particularly how search speed influences this dynamic. This approach aligns with Ludwig & Mullainathan's (2024) view of theory generation as a collaborative process involving humans and AI. Our focus on search speed and its impact on discovery mechanisms illuminates a critical and previously unexplored dimension of the innovation process and its connection to economic growth. While Acemoglu & Restrepo (2018 & 2020) focus on AI and task replacement, we emphasize the interplay between established ("old") and emerging ("new") theories driving innovation and growth. This interplay extends beyond AI's role in task creation.

Arrow (1962) noted that spillovers hinder private investment in basic research. Our framework supports this in the context of slow search speeds, where established theories retain long-term value. However, at high speeds, theory exhaustion necessitates new discoveries for competitive advantage, making basic research an endogenous driver for envisioning new frameworks. This aligns with high-tech firms' significant investments in basic research today. Our work expands on classic justifications for basic research investment (Nelson, 1959; Rosenberg, 1990). We show that at high search speeds, even non-diversified firms benefit from basic research, extending beyond internalizing spillovers (Nelson, 1959) or accessing external knowledge (Rosenberg, 1990). We identify broader conditions under which firms invest in basic research, aligning with Arora et al.'s (2021a) findings of declining basic research investment during periods of slow search speeds. Our model suggests that the rise of AI, potentially leading to faster search speeds, could induce increased investments in basic research, especially by larger firms seeking sustained competitive advantages.

3 Descriptive Evidence

3.1 Firm Size, Research Activities, and Innovation

The interplay between exploration strategies, research activities, and firm size is important for understanding innovation dynamics and competitive advantage. Extensive literature reveals that both large and small firms innovate, though larger firms often face diminishing R&D returns, while smaller firms, despite limited resources, often achieve higher innovation productivity. The balance between incremental and radical innovation has also been examined, with younger managers and competitive environments fostering more disruptive changes.

This section fills gaps in the literature by exploring the role of search speed, showing that faster speeds boost exploration and growth, particularly in top-performing firms. Before showing our evidence, we cast it into what we currently know about firm size and innovation.

The innovation literature has explored extensively the optimal firm size for innovation, revealing that both large and small firms can be highly innovative and invest significantly in basic research (Cohen et al., 1987; Cohen & Klepper, 1996a; Cohen & Klepper, 1996b; Acs & Audretsch; 1988, 1990; Baumol, 2002). However, while R&D investment scales with firm size, R&D productivity often decreases, leading to the ”size contingent” theory, which suggests that R&D returns vary by firm size (Mansfield, 1981; Rosen, 1991; Cohen & Klepper, 1996a).

The relationship between firm size and basic research is more complex (Cohen, 2010). Firms often underinvest in basic research due to its broadly shared societal benefits (Nelson, 1959 & 2006). Smaller firms, despite limited resources, often achieve higher patent productivity relative to their size, indicating diminishing returns for larger firms (Scherer, 1965). Basic research, particularly in the context of General Purpose Technologies (Rosenberg, 1994; Bresnahan & Trajtenberg, 1995), plays a crucial role in boosting productivity and driving economic growth across industries. Larger firms tend to focus on process innovations, leveraging economies of scale (Cohen & Klepper, 1996a) and internal research to prevent competitors from exploiting their discoveries (Arora et al., 2021a).

The choice between incremental and radical innovation is another key factor, with incremental innovations building on existing technologies and radical innovations creating new technology clusters crucial for long-term growth. Firms with open corporate cultures and young managers are more likely to pursue radical innovations, reflecting a culture open to disruptive change (Acemoglu et al., 2022). In competitive environments, firms often focus on internal research to avoid benefiting rivals, while regulations can push firms toward more radical, high-risk strategies (Arora et al., 2021a). Additionally, while regulations typically reduce overall innovation by limiting incremental efforts, they can also push firms toward more radical, high-risk strategies (Aghion et al., 2023).

Small and medium-sized enterprises often excel in niche markets, achieving global leadership in specific areas through continuous innovation (Acs & Audretsch, 1990; Noteboom, 1994)). The German Mittelstand and Southern European SMEs in low-tech industries exemplify this success, thriving through incremental innovations despite resource constraints (De Massis et al., 2018). Different entrepreneurship models lead to varying growth rates and firm sizes (Lehmann et al., 2019), with innovative startups poised for growth differing structurally from ”optimally stable” small firms that maintain their size over time

(Gimenez-Fernandez et al., 2020).

3.2 Impact of Search Speed

We explore the relationship between search speed, firm size, and the allocation of resources between incremental and radical innovation. A plausible hypothesis is that rapid search depletes the available innovations within a specific theoretical paradigm. Consequently, firms with a higher capacity to discover new theories are more likely to pivot to different paradigms and engage in greater exploration. This dynamic should enhance their overall innovation capabilities, as they quickly exhaust existing theories and continually discover new ones, revitalizing their ability to generate innovations.

Meanwhile, firms with more resources can afford to invest in both deepening their current innovations and pursuing new theoretical discoveries. This dual approach positions them to explore more and achieve accelerated growth, thanks to increased innovation productivity. However, the impact will vary, with the most innovation-driven firms maximizing the opportunities presented by faster search processes.

We provide initial suggestive evidence about these conjectures, by employing the DISCERN2 database (Arora et al., 2024a & 2024b; https://zenodo.org/records/13619821), which includes comprehensive data on all U.S. public firms with at least one patent from 1980 to 2021. This database links firms by ownership structure, ensuring that patents filed under different names are correctly attributed to their ultimate owners. Additionally, DISCERN2 integrates patent data with Compustat financial data, providing a robust dataset for our analysis. It also provides the information whether in any year the firm acquired other firms (one or more). Likewise, data for acquired firms are available up to the year in which they are acquired. To our knowledge, this is the most accurate and comprehensive dataset that we could use for our purpose. Table 1 provides definitions and descriptions of all variables used in our study.

Table 1: Description of Variables

Variables	Description
ΔSpeed	Change in speed: Ratio between: a) firm's patents in year t; b) firm's patents in year $t+1$ in CPC subclasses in which the firm patented in year t or before (Our computations from DISCERN data)
% Patent Clusters	Firm's share of patent clusters that include firm's patents in year t (Our computations from DISCERN data. See Appendix A5)
Patents/R&D	Firm's patents over million R&D expenditures (DISCERN2 & Compustat)
Size Index	Average firm's percentile position in total assets, net income, invested capital, and sales (Compustat)
MA	Dummy for whether the firm has acquired one or more firms in the year
Dividends/Assets	Firm dividends over total assets (Compustat)

We measure the change in speed, which we label as $\Delta Speed$, as the ratio between patents in one year and the patents in the following year excluding patents in CPC patent subclasses in which the firm never patented in the past. We use patents as a proxy for firm's ideas. We focus on the change in speed because speed will be affected by the scale of the firm or other similar features. As we will also see in our model in the next section, our measure $\Delta Speed$, which is a ratio between patents in consecutive years, provides a clean measure of the change in speed. We posit that, within a given domain, faster firms generate more patents in the previous year compared to the subsequent year. For instance, given two firms producing 100 patents in three years, a faster firm produces 40, 50, and 10 patents vis-à-vis 30, 45, and 25 patents of the slower firm.

Additionally, we calculated this variable using patents in the same CPC classes or sections, with no much variation in the results. However, subclasses fit well our analysis as they represent well-defined technologies and activities, while more aggregate classes are more heterogeneous. For example, class C07 refers to Organic Chemistry, while subclass C07K refers to Peptides; similarly, class G06 is related to Computing, Calculating, and Counting, while subclass G06F pertains specifically to Electric Digital Data Processing.

To measure exploration, we developed a metric based on patent clusters. These clusters aggregate patents from all firms in our database that share similar combinations of CPC classifications, with details of the clustering algorithm provided in Appendix A5. Our firm-year variable, *% Patent Clusters*, reflects the proportion of clusters that a firm's patents fall into in a given year. We reapplied the clustering algorithm annually, as using a single clustering approach across all years failed to converge due to the large volume of patents. Even the annual clustering required five days to stabilize. Additionally, a single clustering scheme for the entire 1980–2021 period would not capture evolving technological trends, which our measure seeks to reflect. We also calculated a 5-year moving average of clusters, yielding similar results.

The redefinition of clusters every year is not a problem because clusters rearrange randomly using the same criterion (similar combinations of CPC subclasses). Thus, the share of firm patents falling in the clusters is consistent over time. Moreover, we do want the updates of clusters to take into account the evolution of technology – as a matter of fact, the number of clusters increases over time. The share of clusters of the firm's patents then represents a measure of the extent of exploration relatively to the evolution of technology.

The *Patents/R&D* ratio covers all the patents of the firm in any given year, including the patents in new CPC subclasses in which the firm patented for the first time in the year. It is then a measure of the innovation productivity of the firm's R&D including patents in both old and new CPC subclasses.

We constructed the *Size Index* by calculating the percentile position of each firm-year's total assets, net income, invested capital, and sales. The *Size Index* is the average of these four percentile positions. In cases where total assets, net income, or invested capital are negative for some firms in certain years, these negative values lower the firm's percentile ranking accordingly. Our *Size Index* accounts for these variations.

The variable *MA* is a binary indicator representing whether a firm has acquired one or more companies in a given year. We use this dummy to account for changes in firm-level variables that result from additions related to the acquired targets. Using a specific dummy for the exact number of acquisitions reduces our sample size, as many of these instances correspond to singleton observations that are excluded from the analysis. Although the results remain consistent, we prefer to use a single dummy variable to control for the average impact of acquiring one or more firms. Finally, as we will see, we employ *Dividends/Assets* as a measure of the extent to which firms distribute surplus as opposed to reinvesting available assets.. We assume that all firms with zero-dividends throughout the period (1980-2024) do not distribute dividends and we set them as missing observations for this variable.

We focus on the 38,992 observations in DISCERN2 for which we have non-missing values for all our variables, but *Dividends/Assets*. As we said, *Dividends/Assets* is only available for firms that distribute dividends. For the analyses using this variable, we focus on the sample in which this variable is available. Table 2 provides descriptive statistics.

Figure 1 shows the 1981-2021 trends of Δ*Speed*, *% Patent Clusters*, and *Patents/R&D*. They are the annual averages across firms of the differences of the variable from the firm average over the years. This absorbs the firm-specific effects (along with the *MA* dummy) so that we can focus on the average within-firm trends. We can then interpret the yearly values in these figures as the dynamics of an aggregate representative firm in our sample.

Table 2: Descriptive Statistics

Variables	mean	sd	median	min	max	N
ΔSpeed	1.090	1.153	0.922	0	30	38992
% Patent Clusters	0.015	0.040	0.004	0	0.595	38992
Patents/R&D	1.350	13.6	0.337	-83.3	1666.7	38992
Assets (*)	4701.6	25008.3	315.0	0.001	797769	38992
Net Income (*)	238.3	1951.3	3.637	-98696	104821	38992
Invested Capital (*)	2768.2	14034.4	218.4	-54554	502274	38992
Sales	3456.1	15484.1	256.4	-21.8	469822	38992
Size Index	64.6	26.8	70	1	100	38992
MA	0.077	0.266	0	0	1	38992
Dividends/Assets	0.019	0.415	0.007	0	63.5	23781

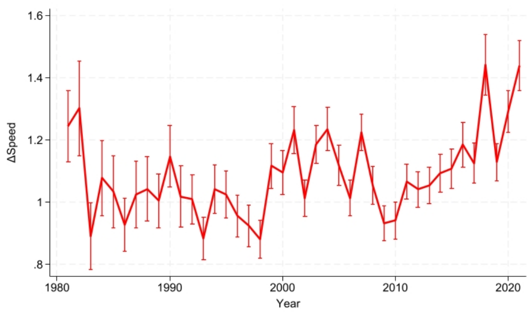
(*) million USD. All observations available for all the variables employed in the analyses of this section, but Dividends/Assets where we use the subset of observations for firms with at least one positive value of the variable during 1980-2021 (as a proxy for excluding non-public firms that do not distribute dividends). Three observations for R&D and eight observations for sales are negative. We chose not to modify the original DISCERN2 data.

Interestingly, all three variables follow a U-shape. They decline till about the start of the millennium, and increase again. Table 3 confirms this trends. It reports the OLS regressions of the three variables using firm fixed effects, the *MA* dummy, and the variables *year* and $year^2$. In all three regressions the coefficients of *year* and $year^2$ are, respectively, negative and positive, suggesting a U-shape, and they are statistically significant. They estimate turning points in years 1997, 2004, and 2009.

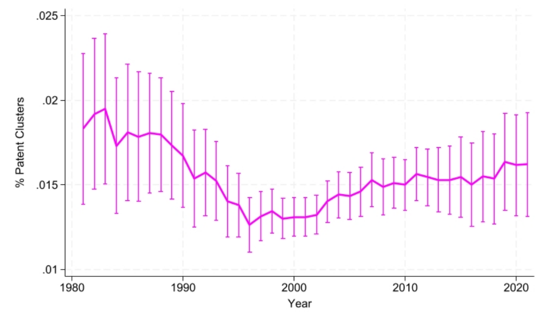
We interpret these trends in three ways. First, the change in speed declined up to the start of the millennium, and has risen again thereafter. Second, exploration, as measured by *% Patent Clusters*, follows the same pattern. Thus, as we conjectured, a decrease in the change in speed is associated with a decrease in exploration, and vice versa. Third, change in speed and exploration are associated with increases in innovation productivity, as measured by *Patents/R&D*.

We also want to assess the conjecture that the relation between *% Patent Clusters* and Δ*Speed* is moderated by the scale of resources. Figure 2 relates the annual averages of

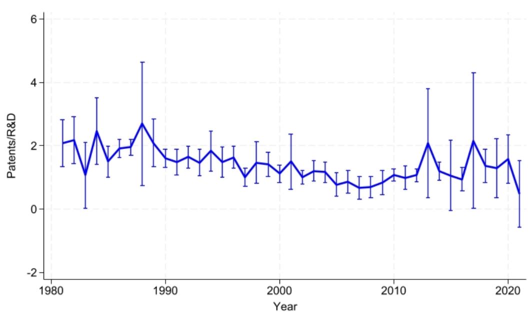
Figure 1: ΔSpeed, % Patent Clusters, Patents/R&D, 1981-2021



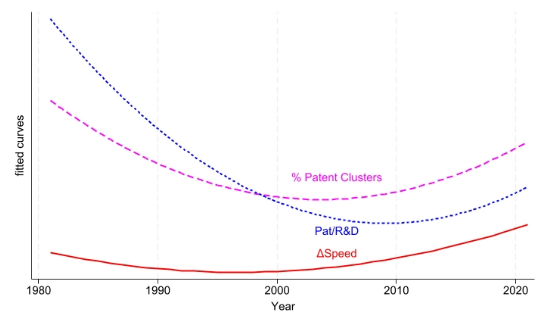
(a) ΔSpeed



(b) % Patent Clusters



(c) Patents/R&D



(d) Fitted Curves

Annual averages of differences of variables from firm average over time (with 95% confidence intervals, including absorption of *MA*). Fitted curves are predicted values in Table 3, rescaled for ease of comparison.

the differences of these two variables from the firm average over time (absorbing *MA* as well), distinguishing the firms in the top and bottom three quartiles of *Size Index*. Again, we can think of the annual dots as the representative firm in the top and bottom three quartiles.

Figure 2 first shows that, for given Δ*Speed*, *%Patent Cluster* is systematically higher for the firms in the top quartile, which are the larger firms. While this may seem a natural implication of the firm's scale, it could also be the very consequence of our conjecture that knowledge theories exhaust. A larger scale of available resources is not invested repeatedly in the same theory, but eventually redirected towards new theories where firms can search

Table 3: Fitted Curves

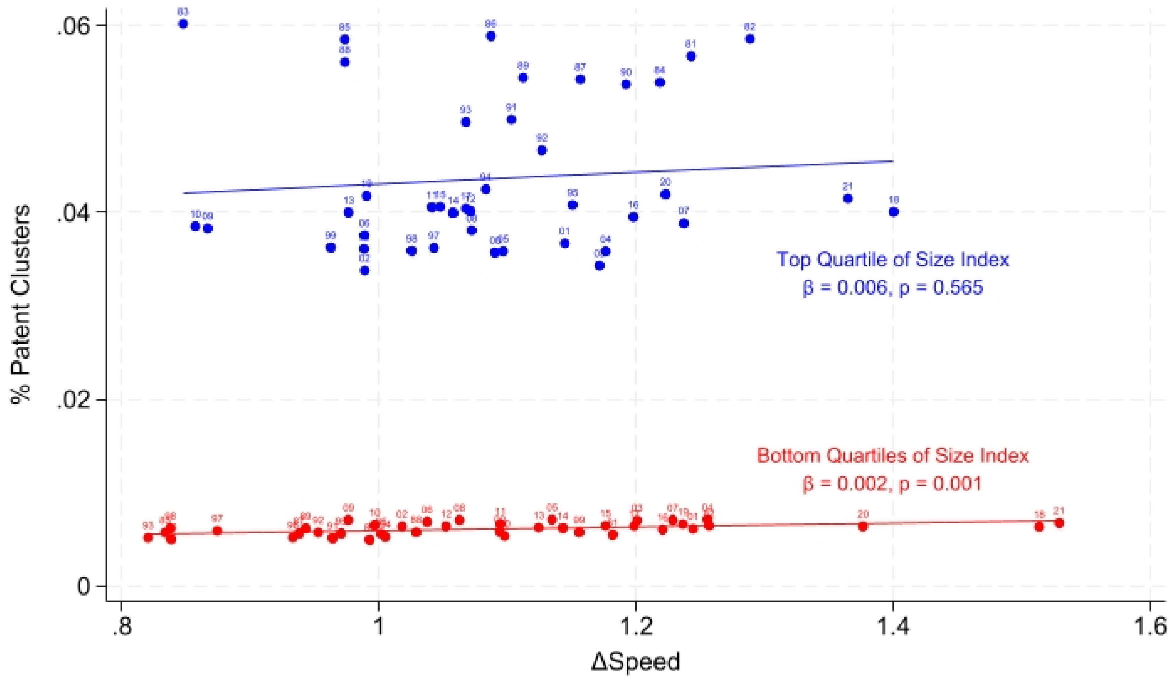
Variables	Speed	% Patent Clusters	Patents/R&D
year	-1.794***	-0.043***	-5.799**
	(0.000)	(0.007)	(0.017)
year^2	0.000***	0.000***	0.001**
	(0.000)	(0.007)	(0.017)
Constant	1,792.377***	43.192***	5,826.514**
	(0.000)	(0.007)	(0.016)
Observations	38,992	38,992	38,992
R-squared	0.124	0.794	0.371
Fixed effcts	firm, MA	firm, MA	firm, MA
Clustered errors	firm	firm	firm
log likelihood	-58317	101454	-148180

Robust p-values in parenthesis. *** $p < 0.001$; ** $p < 0.05$; * $p < 0.10$. Based on the estimated parameters, turning years are, respectively, 1997, 2004, 2009.

in fresh new domains. Second, Figure 2 shows that Δ*Speed* increases *% Patent Clusters* faster in the top quartile of size. This could suggests that resources help to increase exploration as speed accelerates. However, this faster increase is statistically less robust than the smaller increase in the bottom three quartiles. This implies that there is quite some variability across firms in the top quartile of size in responding to stronger changes in speed with greater exploration.

Figure 3 studies whether exploration matters. The figure shows whether the representative firm in the top quartile of exploration (*% Patent Clusters*) grows faster (higher increase in the *Size Index*) when the change in speed is higher. As a matter of fact, we now find that: a) for the top quartile of *% Patent Clusters*, the *Size Index* grows faster with Δ*Speed* than for the bottom quartiles – moreover, the former is statistically more significant than the latter; b) the top quartile of *% Patent Clusters* corresponds to a systematically higher *Size Index*, which ranges from around the 70^{th} percentile to above the 80^{th}. Simply put,

Figure 2: % Patent Clusters vs ΔSpeed, by Quartile of Size Index



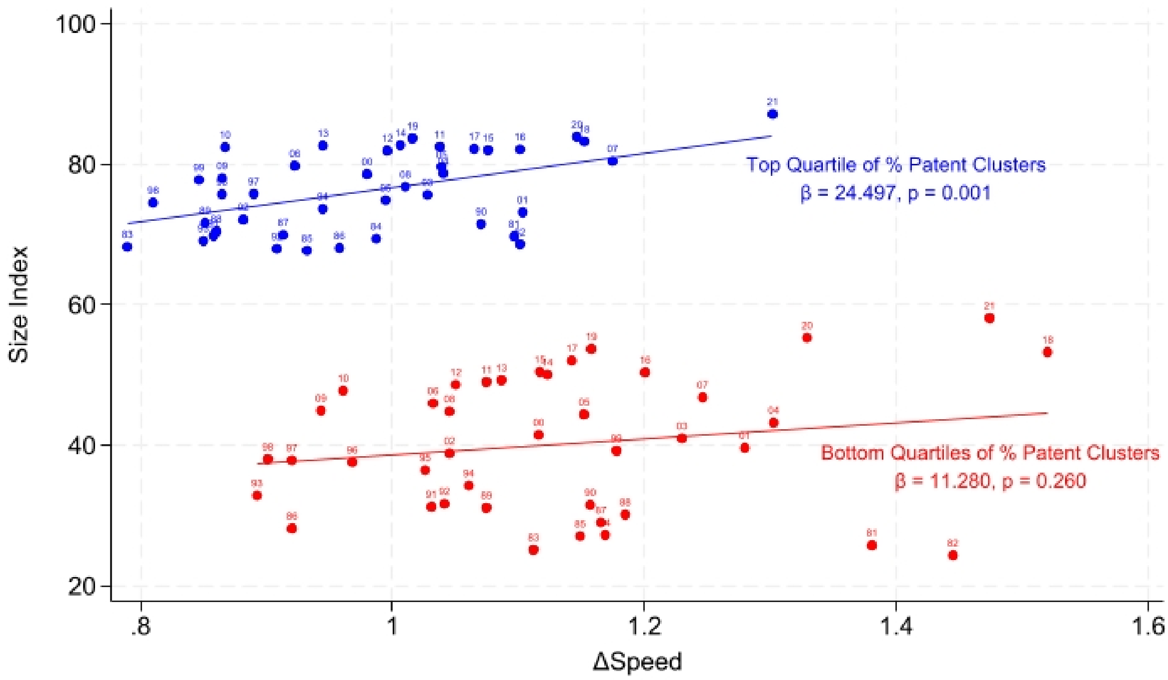
Annual averages of differences of variables from firm average over time, including absorption of *MA*.

as $\Delta Speed$ increases, larger firms are associated with larger shares of patent clusters and larger increases in size.

All this raises the question why large firms grow by investing in new theories, as implied by the stronger association between $\Delta Speed$ and *Size Index* for the top quartile of *% Patent Clusters*, and some of these firms do not invest in new theories, as implied by the high variance of the positive association between $\Delta Speed$ and *% Patent Clusters* for firms in top quartile of the *Size Index*, as shown in Figure 2.

Figure 4 sheds some light on this question by showing that, for the top quartile of firms by *Size Index*, *% Patent Clusters* grows with Δ *Speed* only for firms with *Dividends/Assets* below the median of the sample in each year. Interestingly, they are also the firms with a systematically lower share of patent clusters. This suggests that firms that do not

Figure 3: Size Index vs ΔSpeed, by Quartile of % Patent Clusters

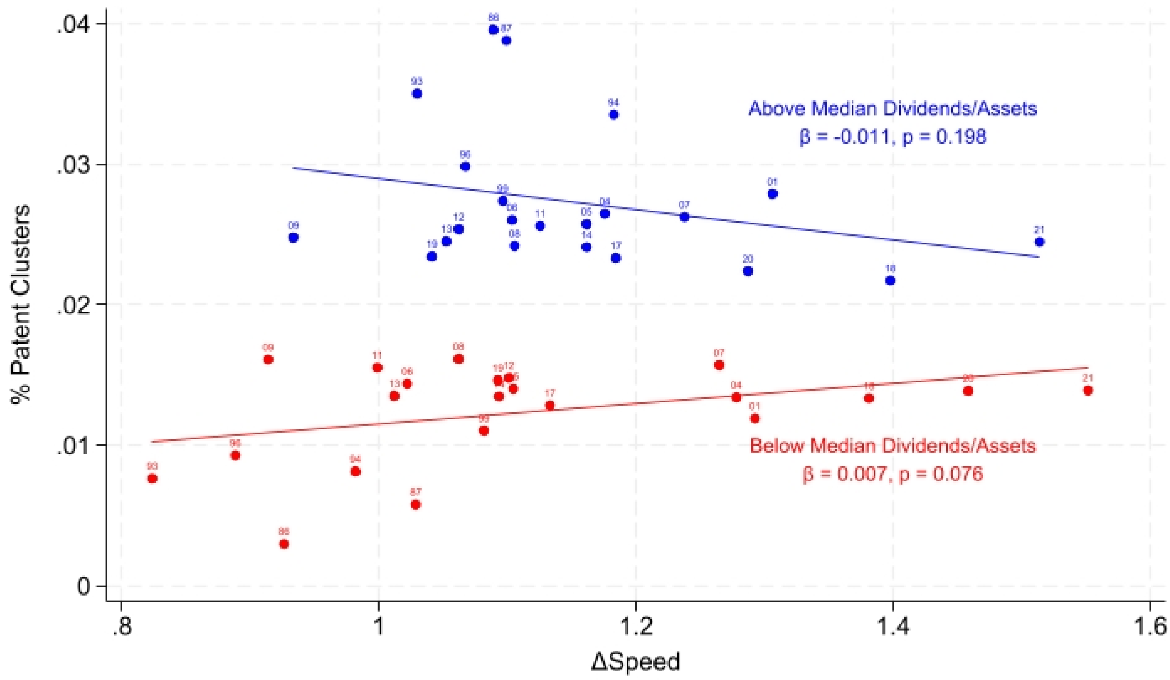


Annual averages of differences of variables from firm average over time, including absorption of *MA*.

explore much respond to higher $\Delta Speed$ by replacing the distribution of dividends with investments in exploration. Firms that exhibit high *% Patent Clusters* distribute, instead, dividends, and do not respond to $\Delta Speed$ by increasing exploration.

Finally, Figure 5 shows that the firms in the top quartile of *Size Index* grow faster with $\Delta Speed$ when they are below the median of Dividends/Assets for the year. This suggests that alternative uses of dividends, such as for exploration, contribute to faster growth.

Figure 4: % Patent Clusters vs ΔSpeed by Quartile of Dividends/Assets
(Top Quartile of Size Index)



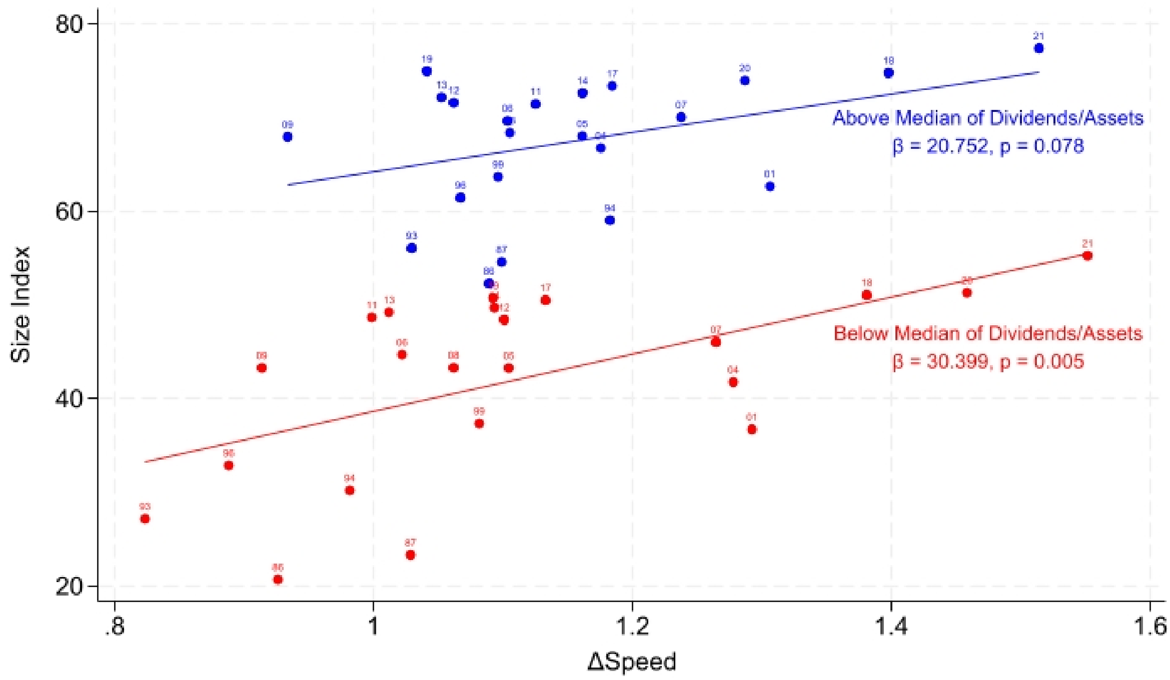
Annual averages of differences of variables from firm average over time, including absorption of *MA*.

4 Model

4.1 Set-Up

In this section, we present a model that can explain the patterns observed in the data. We conceptualize the innovation process as a search within bounded spaces, where firms progressively deplete opportunities. This process exhibits diminishing returns as exploitable opportunities within a given bounded space are gradually exhausted. Drawing on Camuffo et al. (2024), we refer to these bounded spaces as "theories"—collections of business problems, solutions, conjectures, and logical connections that shape what companies search for and why. Discontinuous change, like discovering new problems or spaces, re- vamps

Figure 5: Size Index vs ΔSpeed by Quartile of Dividends/Assets
(Top Quartile of Size Index)



Annual averages of differences of variables from firm average over time, including absorption of *MA*.

the search within a new domain, leading to fresh new opportunities. Theories are deemed new when the firm realizes it must pivot to new areas, encompassing different problems, solutions, conjectures, and logical association.

We assume that firms search by following a staged approach. Let $t_i \in (0, T)$ be the time allocated in stage $i = 1, 2, ...\infty$ to search for new ideas in the current theory, and T the time endowment in each stage. This time endowment can differ because firms can use more resources (e.g. more researchers) or more productive researchers (e.g. because of more qualified researchers or better technologies). During each stage, they invest time t_i to uncover new ideas within the current theory and they dedicate the remaining time $(T - t_i)$ to discover new theories.

Let $\lambda_i t_i^{\alpha}$, with $\alpha \in (0,1)$, be the expected arrival rate of ideas in current theory in stage i. We posit that firms explore a space s_i which is the space of what they know in stage i. A theory defines this space. Firms can invest time t_i to learn about the theory – that is, about its different ramifications, potentials, and applications – and they can expand the theory according to $s_i = t_i^{\alpha}$. Thus, the space of what they know is not fixed, and firms can expand it through exploration. However, this exploration within a theory cannot change the structure, or the paradigm, of the theory, which is determined by the sequence of λ_i.

We use a specific functional form for λ_i, that is, $\lambda_i = \lambda_1 i e^{-\gamma(i-1)}$. This functional form allows for the possibility that the productivity parameter increases in the initial stages and then decreases. In this way, we capture the fact that at the initial stages firms may not yet experience a reduction in the productivity of search, but eventually they do. Let $\lambda_1 = \lambda(1-e^{-\gamma})^2$ so that $\sum\limits_{i=1}^{\infty} \lambda_i = \lambda$. The parameter λ is the expected total number of ideas of the theory.[1]

The theory's richness or potential is then λ, the total number of ideas that it can generate. The exploration within the theory helps to find these λ ideas by exploring the spaces, or parts of the theory, s_i, which can be larger or smaller depending on the time t_i invested in the search within the theory. But, as we discussed earlier, the shift to a new theory with a different λ occurs because they spend time $T - t_i$ to discover a new theory.

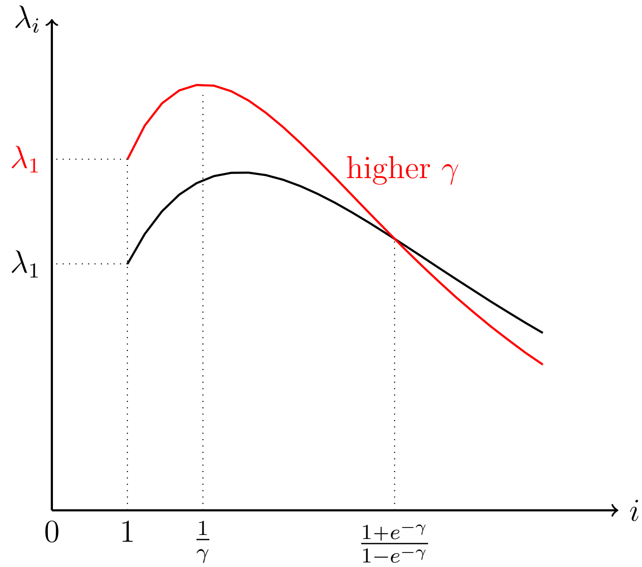
Therefore, the arrival rate of new ideas depends on a measure of the theory's potential to generate new ideas (λ), the efficiency of search parameter $\gamma > 0$ that represents the rate of decay of ideas across stages in the current theory, and the resources t_i allocated to the search in the current theory. We interpret γ as our measure of change in speed. A faster search raises the productivity of search in the earlier stages of the current theory

[1] Rewrite $\sum\limits_{i=1}^{\infty} \lambda_i = \lambda_1 \left[\sum\limits_{i=1}^{\infty} e^{-\gamma(i-1)} - \frac{\partial}{\partial\gamma} \sum\limits_{i=2}^{\infty} e^{-\gamma(i-1)} \right] = \lambda$

and lowers it in later stages. This depends on the fact that the space is bounded: if firms discover more ideas earlier, they discover fewer ideas later because they have already discovered a larger fraction of ideas within the theory.

The parameter γ shifts the curve λ_i upward in early stages, and lowers it for larger i, as shown in Figure 6. Moreover, λ_i starts decreasing at $i = \frac{1}{\gamma}$, and when γ increases, the higher-γ curve falls below the lower-γ curve at $i = \frac{1+e^{-\gamma}}{1-e^{-\gamma}}$. Since for $\gamma > 0$, $\frac{1+e^{-\gamma}}{1-e^{-\gamma}} > \frac{1}{\gamma}$, the higher-$\gamma$ curve always falls below the lower-γ curve in the declining portion of the curve.[2] Also, a higher γ shifts the tilt of the curve, which occurs at $i = \dfrac{1}{\gamma}$, to an earlier stage i. Figure 6 illustrates these patterns.[3]

Figure 6: Effect of Higher Speed (γ) Within Theories



[2]The sign of $\frac{1+e^{-\gamma}}{1-e^{-\gamma}} - \frac{1}{\gamma}$ is the same as the sign of $\gamma\left(1+e^{-\gamma}\right) - \left(1-e^{-\gamma}\right)$. The minimum of this expression, which occurs at $\gamma = 1$, is $2e^{-1} > 0$, which implies $\frac{1+e^{-\gamma}}{1-e^{-\gamma}} > \frac{1}{\gamma}$ $\forall \gamma > 0$.

[3]If $\gamma > 1$, the λ_i curve tilts at $i < 1$, and thus it will be always declining when $i \geqslant 1$. This does not affect the implications of our model. This is a direct consequence of speed: High speed rapidly depletes opportunities, causing the phase of diminishing productivity to begin immediately.

The firm's dynamic program for knowledge accumulation K is

$$K = \sum_{i=1}^{j-1} e^{-\delta i} \lambda_i T^\alpha + \sum_{i=j}^{\infty} e^{-\delta i} e^{-\pi(i-j)} \left[\lambda_i t_i^\alpha + \pi \left(1 + g_i\right) K\right] \tag{1}$$

where δ is a discount factor, $\alpha \in (0, 1)$ is a parameter that measures the productivity of effort, and if $j = 1$ the first term of this expression disappears.

Firms wait $j - 1$ stages before exploring a new theory. Till then $t_i = T$. At stage j they invest $t_j \leqslant T$ in the current theory and start exploring a new theory. If they do not find a new theory in stage j, which occurs with probability $e^{-\pi}$, firms invest $t_{j+1} < T$ in the current theory, and look for a new theory, which they find with probability π. If they find the new theory, they start exploiting it in the next stage, and obtain K. In the generic stage $i > j$, they have not found a new theory in the previous $i - j$ stages with probability $e^{-\pi(i-j)}$, exploit the current theory by investing t_i, and find a new theory with probability π.

The term $e^{-\pi(i-j)}$ is the probability of no occurrences in $i - j$ independent Poisson processes. We approximate the probability of at least one occurrence, which is $1 - e^{-\pi}$, with π. *De facto* we are assuming that π is small enough that $1 - e^{-\pi} \approx \pi$ – that is, in the time span of one stage firms obtain at most one occurrence. This assumption is harmless, and amounts to say that the time span of each stage is relatively short.

The term g_i is the growth rate of knowledge in stage i, which we approximate by

$$1 + g_i = T - t_i \tag{2}$$

Thus, firms spend their exploration time to identify richer theories (i.e., with higher potential for ideas generation λ). A lower t_i, which raises $T - t_i$, increases $(1 + g_i)\lambda$, which is the expected total number of ideas of the new theory. However, the new theory is feasible only with probability π.

This set-up relies on some simplifying assumptions. We assume that firms cannot find new theories exogenously. If they do not spend time to look for them, that is $T-t_i = 0$, they do not find theories with a positive expected number of ideas. Following Cohen & Levinthal (1989) and Rosenberg (1990), they must make at least some minimal investments in absorptive capacity to understand external scientific knowledge. Also, the probability π of finding new feasible theories does not depend on the search for new theories in previous stages. Similarly, the expected number of ideas of the new theory, which depends on $1 + g_i$, only depends on the effort in the current stage and not on the efforts in the previous stages. These extensions complicate the analysis without providing major new insights.

4.2 Solving the Dynamic Program

In Section A1 of the Appendix we solve the dynamic program (1) by choosing the sequence of time allocations t_i, $i \geqslant 1$. The first order condition implies that in stage i firms allocate t_i till the additional unit of t_i allocated to further exploration in the current theory is equal to what firms expect to obtain if they allocate that unit of time to the search of a new theory.

This yields the following optimal t_i, for $i > j$, as a function of the initial condition t_j.

$$t_i = \Gamma_{ij}^{\frac{1}{1-\alpha}} t_j \tag{3}$$

where $\Gamma_{ij} \equiv \dfrac{i}{j} e^{-\gamma(i-j)}$

To determine the initial condition t_j, we make the following assumption.

Assumption $(\mathcal{A})$**.** *The optimal* $t_1 \geqslant T$.

Under this assumption, Lemma 1 below establishes that the initial condition occurs in the declining portion of the productivity (λ_i) curve, and the initial condition is such that $t_j \approx T$.

Lemma 1. $\mathcal{A} \implies j \geqslant \frac{1}{\gamma}$ *&* $t_j \approx T$.

Proof. See Appendix. ■

As shown in Section A1 of the Appendix, we solve the first order condition for t_j by approximating the optimal $t_j = T$. This amounts to solving for the optimal j. In turn, this implies that for $i \leqslant j$, $t_i = T$, and for $i > j$, t_i is determined by (3).

We obtain the following first order condition

$$\frac{\alpha j e^{-\gamma(j-1)}}{\pi T} - \frac{e^{-\delta}}{\left[1 - e^{-(\delta+\gamma)}\right]^2}(1 + \Omega) = 0 \tag{4}$$

where

$$\Omega \equiv j e^{-(\delta+\gamma)(j-1)} \left[1 - e^{-(\delta+\gamma)}\right]^2 \Psi_j \tag{5}$$

and $\Psi_j \equiv \sum_{i=j}^{\infty} e^{-\delta(i-j)} \psi_i$, with $\psi_i \equiv e^{-\pi(i-j)} \left[\alpha + (1-\alpha)\, \Gamma_{ij}^{\frac{1}{1-\alpha}}\right] - \Gamma_{ij}$.

Lemma 2. $\mathcal{A} \implies \exists$ *an optimal* j.

Proof. See Appendix. ■

Section A1 of the Appendix shows that the optimal K is

$$K = \frac{\alpha \lambda_1 j e^{-\gamma(j-1)}}{\pi T^{1-\alpha}} \tag{6}$$

After replacing $\dfrac{\alpha j e^{-\gamma(j-1)}}{\pi T}$ from (4) obtain

$$K = K_0 \left(1 + \Omega\right) \tag{7}$$

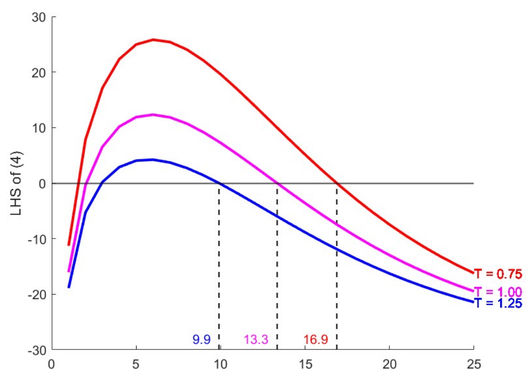
where $K_0 \equiv \sum_{i=j}^{\infty} e^{-\delta i} \lambda_i T^{\alpha} = e^{-\delta} \lambda T^{\alpha} \left[\dfrac{1 - e^{-\gamma}}{1 - e^{-(\delta+\gamma)}} \right]^2$ is K under no switch ($\pi = 0$), and we replaced $\lambda_1 = \lambda \left(1 - e^{-\gamma}\right)^2$.

Lemma 3. *The optimal j declines with T and π*

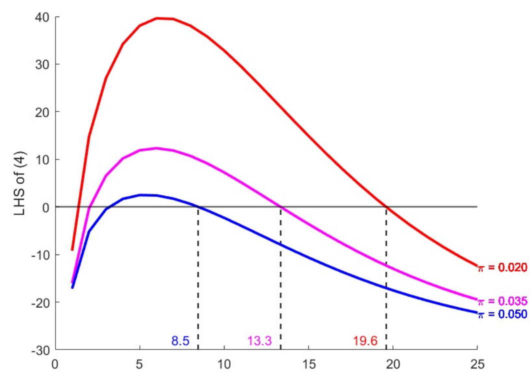
Proof. See Appendix. ■

As an illustration of the optimal solution simplied by *foc* (4) (Lemma 2) and the comparative static implied by Lemma 3, Figure 7 reports the shape of the left-hand-side (LHS) of (4) as a function of j. As the curves show higher T or π reduce the optimal j. Plots for different values of γ, δ, or α yield in most of the cases similar patterns.

Figure 7: Optimal j and Comparative Static vs T and π
($\delta = 0.05$, $\gamma = 0.15$, $\alpha = 0.5$)



(a) Optimal j vs T ($\pi = 0.035$)



(b) Optimal j vs π ($T = 1$)

Curves represent the left-hand-side (LHS) of (4) as a function of j. The optimal j is the one that corresponds to the point in which the curves cross the 0-line.

5 Implications

5.1 Search speed and discovery rate

Search speed γ and discovery rate π depend on general conditions of the economy and idiosyncratic characteristics of firms. We posit that there is a distribution of values γ or π across firms, and common factors that raise or lower the values of γ or π of all firms. For instance all firms benefit from new opportunities that raise speed (like AI) or by scientific or technological advances that open new domains. We assume for simplicity that the idiosyncratic components of γ and π are "endowments" of the firms. Firms can be faster or slower or more or less able to or interested in exploring new domains.

In principle, firms could optimize their γ and π at the outset of their dynamic program as investments that set the underlying (long-term) choice of the firms in terms of speed and ability to discover new theories. To streamline our analysis we assume that they are endowments of the firm after this optimization process. Optimization would still depend on exogenous factors that raise or lower the optimal γ or π of the firm. We focus on this reduced form and discuss the effects of changes in γ and π as the exogenous parameters that affect the choice of how fast the firm wants to be or how much it wants to explore. This also implies that, in our discussion below, firms can reduce π or γ if they are sub-optimally high, but cannot increase them. Increases in γ and π are costly, and the optimization has already reached the optimal level of γ and π given costs. However, if γ and π are higher than optimal, reducing them (with implied lower costs) is feasible. *De facto*, this means that the optimal γ or π is below the frontier that they can reach, and firms can move below their frontier, but they cannot push γ and π beyond their frontier.

Firms choose, instead, t_i explicitly. Thus, the allocation of resources between current and new theories is the strategic choice that firms make given these parameters. A lower t_i

reflects a higher $T - t_i$ – that is, firms invest to a greater extent in the growth of new theories.

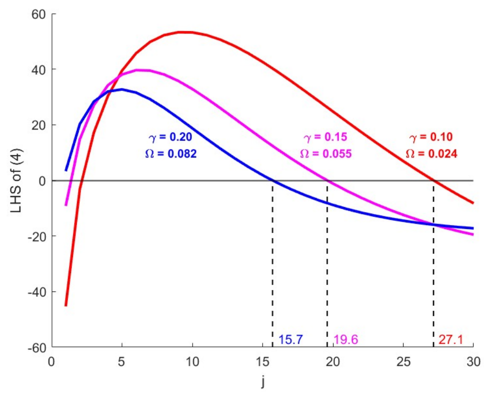
Theorem 1. *The discovery of new theories is limited by the speed of search:* $\gamma < \gamma' \implies \pi > 0$ *is higher than optimal.*

Proof. See Appendix. ∎

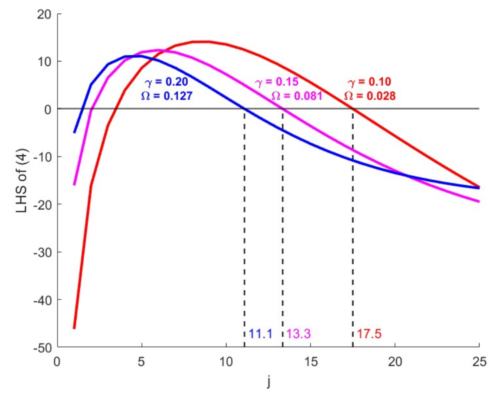
This theorem states that given the endowment π of the firm, there is a threshold γ' such that for smaller γ, the endowment of the firm is suboptimal. In other words, the firm has a higher ability to discover new theories than it is ideal. The intuition is that, for a low search speed γ, there are enough discoveries to be made in the current theory in later stages, which lowers the value of a higher ability to discover new theories. It is natural to assume that firms endowed with too high a π either lower it, or, if not, they produce suboptimal knowledge because they switch "too early" to new theories leaving ideas from the old theory "on the table". Conversely, a sufficiently high search speed makes any given ability π fully valuable. Early switch are beneficial because the old theory exhausts rapidly and a greater ability to discover a new theory pays off. Firms reallocate resources t_i away from the current theory to the new theory, and j declines.

Figure 8 helps to illustrate Theorem 1. It shows the left-hand-side of (4) as a function of j and the optimal j for different levels of π and γ. The figures also report the optimal Ω using the given π and γ and the optimal j. From (7), K increases with Ω, and the figures show that the optimal π is 0.05 when speed is high ($\gamma = 0.20$), 0.035 when speed is medium (0.15), and 0.02 when speed is slow (0.10). As implied by Theorem 1, when speed is slow, a high ability π to discover new theories is suboptimal. It is simply too high, leading too early to the discovery of new theories, before firm have exhausted, optimally, the exploitation of the current theory. Only when speed is high ($\gamma = 0.20$), the high $\pi = 0.05$ is optimal. In this case, high speed enables firms to exhaust theories more quickly, raising the value of a higher ability to discover new theories. Similarly, medium speed ($\gamma = 0.15$) makes $\pi = 0.035$ optimal, and when speed is low ($\gamma = 0.10$), firms need to be slow in discovering new theories, making $\pi = 0.02$ optimal.

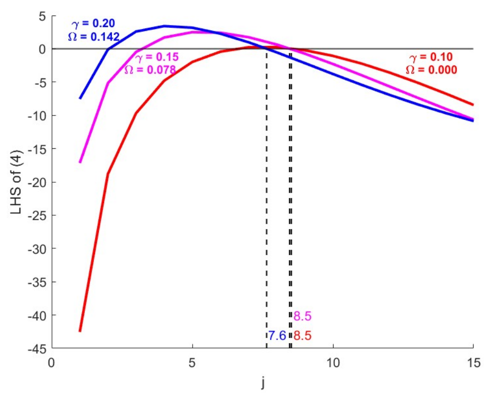
Figure 8: Optimal π at Different Speed γ
($\delta = 0.05$, $T = 1$, $\alpha = 0.5$)



(a) $\pi = 0.020$



(b) $\pi = 0.035$



(c) $\pi = 0.050$

Curves represent the LHS of (4) as a function of j for different π and γ. Ω is evaluated at the given π and γ, and the corresponding optimal j. From (7), K increases with Ω. The figures show that for $\pi = 0.05$, 0.035, and 0.020, the highest Ω corresponds, respectively, to high-, medium-, and low-speed ($\gamma = 0.20$, 0.15, 0.10). As implied by Theorem 1, $\pi = 0.05$ can only be sustained by high speed and $\pi = 0.035$ by medium speed. With slow speed, $\pi = 0.02$ is optimal.

Interestingly, the proof of this Theorem shows that a higher π always reduces Ψ_j in the expression for Ω. This is because a higher π provides fewer resources to explore the current theory, and thus makes it less likely to develop ideas in the current theory. However, a higher π reduces the optimal j, anticipating the opportunity to produce new ideas in the new theory. When speed is low, the cost of giving up the current theory can be so high that this anticipation will never offset the costs of giving up resources allocated to the old theory. When speed becomes sufficiently large, the quicker exhaustion of theories in the current theory starts making this anticipation potentially beneficial. Firms start switching, and the earlier switch provides the opportunity to improve the benefits of finding and searching within new theories vis-à-vis continuing the search in the old theory.

Indeed, Figure 8 shows that the highest $\pi = 0.05$ always corresponds to the lowest optimal j. However, this anticipation is suboptimal unless speed is very high. Lower speed makes smaller π and a higher waiting time before exploration more valuable. In this respect, Theorem 1 also provides an explanation for Christensen's (1997) paradox whereby incumbent firms overlook new opportunities. This behavior is driven by a low search speed.

Corollary 1. *A sufficiently high search speed is a necessary condition to ensure that knowledge increases with the probability of discovering new theories:* $\dfrac{\partial K}{\partial \pi} \implies \gamma > \gamma'$

Proof. See Appendix. ■

This corollary highlights an implication of Theorem 1. If γ is small ($\gamma < \gamma'$), increases in π reduce K. The intuition is again that firms are more likely to jump to new theories before they have exhausted, optimally, the current theory. Thus, they want to be less impatient about producing new theories, and lower π. Only if speed is sufficiently high, firms can fully exploit a given ability $\pi > 0$ to discover new theories, and firms with π higher than the given endowment π enjoy a higher K.

Figure 8 provides, again, an illustration of these patterns. As noted, (7) implies that K

increases with Ω, which is affected by π. However, π does not affect K_0. Therefore, as the figures show, a higher π corresponds to the optimal Ω, and therefore to the optimal K only if γ is higher. Thus, a high γ makes it more likely that K increases with π.

Put differently, when speed increases beyond the threshold γ', not only is a higher endowed π more valuable, but firms with higher π also perform better. This speaks, for instance, to the fact that, with low speed, firms may invest too much in basic research to find new theories. Conversely, with higher speed, the value of basic research to discover new theories increases: With low speed basic research is redundant, with high speed, it is a scarce resources. Also, an economy in which speed increases (e.g. because of AI or more generally because of new technologies that accelerate the production o ideas) will experience increasing differences in knowledge accumulation across firms. The mechanism of these growing differences is the higher productivity in more basic research to discover new theories.

5.2 Resources T and Growth of Ideas

The choice of T reinforces the effects discussed in the previous section. Higher γ and π raise T, and the combined effect of these three variables raises the growth of new ideas. As discussed in the previous section, low speed implies that firms with low or high π do not differ systematically because low speed dampens the potential of higher π. In this section we show that high speed raises the value of π, and they both raise T and a higher growth of ideas. This widens gaps across firms further.

To keep matters simple, let $V(K, T)$ be the ultimate value that the firm looks at, where K is our knowledge asset that depends on T, and the argument T represents an alternative allocation of resources such that a higher T increases K and V indirectly; the direct effect of T is negative because it represents an alternative allocations that generate value for the firm. We assume that V is concave in T, and thus there is an optimal allocation of T.

One interpretation of the direct effect of T is that not only does value depends on knowledge but also on the implementation of knowledge to generate products, services or other goods that agents use. To the extent that resources T are rivalrous, this generates a trade-off between production of knowledge and implementation. Another interpretation is that T can be allocated alternatively to activities other than the growth of knowledge of the firm. A classical example is "leisure" or more generally reallocation of the firms' proceeds to shareholders in the form of dividends rather than investment. In what follows, we keep these interpretations open. They all have similar implications for our main points.

Theorem 2. *A sufficiently high search speed is a necessary condition for K to be supermodular in π, γ, and T:* $\frac{\partial^2 K}{\partial \pi \partial \gamma}, \frac{\partial^2 K}{\partial \pi \partial T}, \frac{\partial^2 K}{\partial \gamma \partial T} > 0 \implies \gamma > \gamma'$.

Proof. See Appendix. ■

This theorem establishes that only if $\gamma > \gamma'$, the endowment π of a firm is not "too high". A low speed does not trigger the complementarity stated by Theorem 2. In this case, firms lower π and increase the allocation of resources t_i on the current theory. Conversely, if $\gamma > \gamma'$, they respond to the higher γ by fully exploiting their higher ability to discover new theories, lowering allocations to t_i and increasing T, so that $T - t_i$ increases, and the endogenous T raises further the richness of the newly explored domains and growth of new ideas.

Table 4 and Figure 9 provide an illustration of Theorem 2. The table reports different values of K for the eight triplets of low/high γ, π, and T. We employ the same values of the parameters used in Figures 7 and 8, including $\delta = 0.05$ and $\alpha = 0.05$, and j is set optimally, using (4), given the parameters. The table also reports the difference-in-difference that corresponds to the second partial derivatives of any two of the three parameters, for the low and high values of the third parameter. As the table shows, all these differences are positive, but the case for the pair (T, π) when γ is low. As implied by the theorem, $\gamma < \gamma' \implies \Omega_\pi = 0$ because, as shown in the proof of the theorem, $\Omega_\pi < 0$ and firms reduce π optimally from their endowment up to the point where $\Omega_\pi = 0$. As the proof of the theorem shows, this implies $\dfrac{\partial^2 K}{\partial \pi \partial T} = 0$. This is the case in which the complementarity breaks down: With low speed, increases in T do not raise the marginal value of the ability π to discover new theories.

Table 4: Values of K for Different Combinations of T, γ, and π
($\lambda = 1$, $\delta = 0.05$, $\alpha = 0.5$)

$\boldsymbol{\gamma}$	$\boldsymbol{T}$	$\boldsymbol{\pi} = 0.020$	$\boldsymbol{\pi} = 0.050$
0.10	0.9	0.4296	0.4213
0.20	0.9	0.6484	0.6792
0.10	1.1	0.4791	0.4657
0.20	1.1	0.7333	0.7809

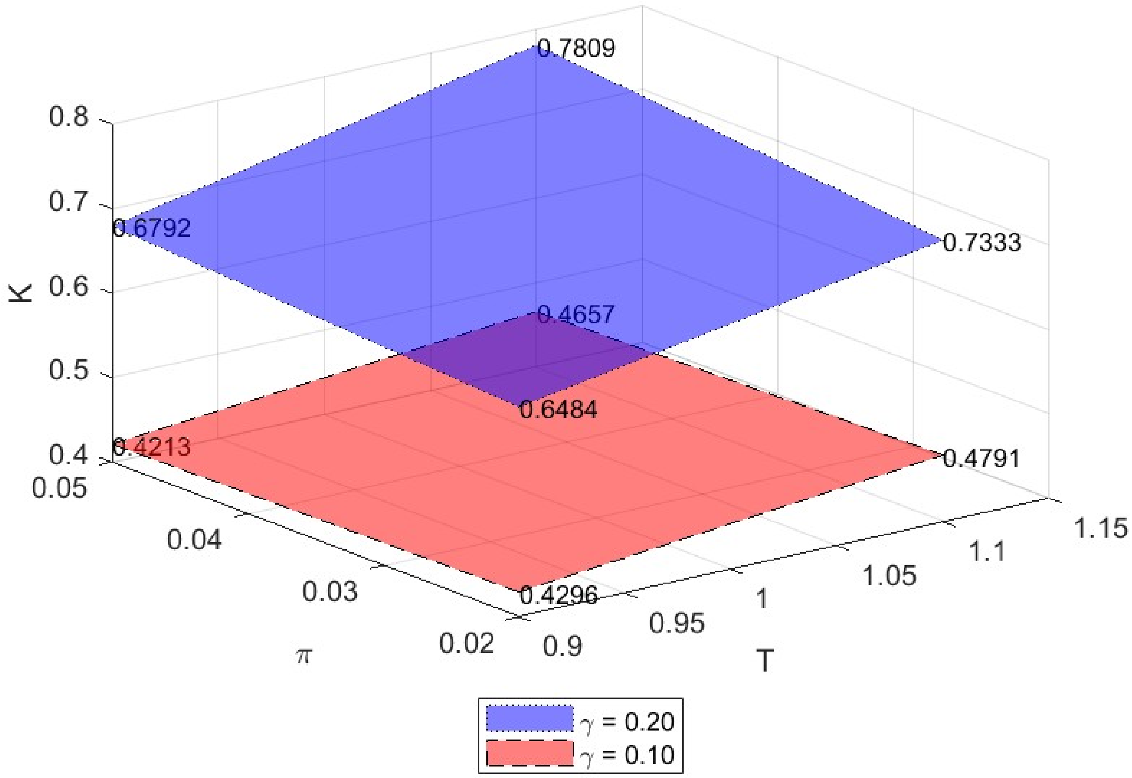
$\dfrac{\Delta^2 K}{\Delta\gamma\Delta\pi}$	$T = 0.9$	0.0391
	$T = 1.1$	0.0610
$\dfrac{\Delta^2 K}{\Delta\gamma\Delta T}$	$\pi = 0.02$	0.0354
	$\pi = 0.05$	0.0572
$\dfrac{\Delta^2 K}{\Delta\pi\Delta T}$	$\gamma = 0.10$	-0.0050
	$\gamma = 0.20$	0.0168

The first prospect reports the values of K for the corresponding combinations of T, γ, π given the optimal j in Figures 7 and 8. The second prospect reports the two differences in the differences of any two of T, γ, and π, for the two values of the third variable. The second prospect shows that the variables are complementary. However, for γ low, π and T are not complementarity, as implied by Theorem 2.

Figure 9 provides a graphical illustration. It reports the two planes corresponding to γ low and high in the (K, T, π) space. As the figure shows, the plane corresponding to the low γ is flatter than the place corresponding to the high γ. Specifically, when T is small ($T = 0.9$), the higher π reduces K, as also shown by the slightly negative cross-partial in Table 4. When T is high, a higher π increases K. These differences are more pronounced when speed is high, as shown by the upper plane in the figure.

Theorem 2 generates several insights. First, it aligns with Bloom et al. (2020), who note that new ideas are harder to find because they require an increasingly large scale

Figure 9: Values of Different Combinations of T, γ, and π
($\delta = 0.05$, $\alpha = 0.5$)



The figure reports graphically the two planes that correspond to the first prospect of Table 4 for $\gamma = 0.10$ or 0.20

of resources. In our model, a low search speed prevents resources (T) from accumulating fast enough to discover new theories, suggesting that the discovery of new theories is an antecedent to the lack of resources. New technologies that increase search speed can enable firms to discover new ideas at a higher rate.

Second, according to Theorem 2, a low speed ($\gamma < \gamma'$) levels off firms. Differences in π, γ, or T do not generate higher gaps. This is no longer the case when speed is high. Speed can trigger increasing inequality in the growth of firms. Firms with higher endowments of

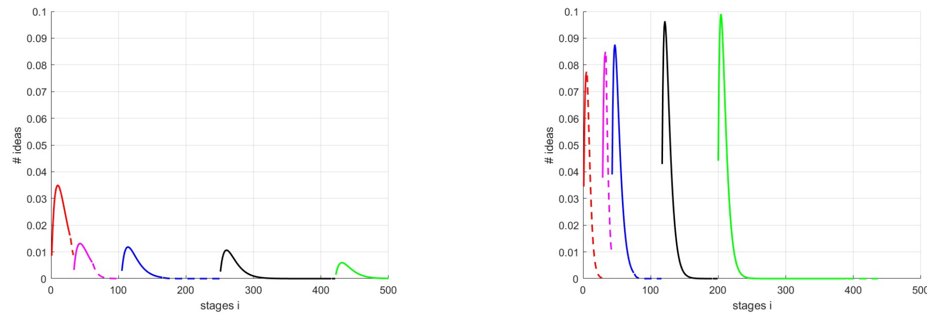
π or γ, or a greater ability to exploit economy-wide opportunities that raise π or γ, will enjoy increasing gaps also thanks to their greater incentives to raise T, which generates a virtuous circle among these three variables. Simply put, speed is likely to trigger of increasing returns via faster speed, a greater rate of discovery of new theories, and a greater incentives to invest resources to increase T in order to spin the virtuous cycle further.

Finally, it is interesting to understand how the mechanism of higher growth unfolds as speed increases. As noted in Section 4.1, each new theory has a pool of ideas that increases by $(1 + g_i)\,\lambda$, where λ is the pool of ideas of the previous theory, while $1 + g_i = T - t_i$. Higher speed increases T and π, lowers j, and lowers t_i. The complementarities triggered by high speed create the conditions for exhausting the ideas of the current theory faster, jumping on a new theory earlier, and discover new theories with a higher pool of ideas.

Figure 10 provides a representation of these processes. The figure reports two cases. The first one represents low speed ($\gamma = 0.10$), and then endogenously (by Theorem 1 and Corollary 1) low π (0.02) and (by Theorem 2) low T (0.9). The second one represents the case of high γ, π, and T (0.20, 0.05, and 1.1).

As the figure shows, low speed generates longer and less steep cycles than high speed. As discussed in Appendix A4, we draw randomly the stage at which, after the optimal j, firms discover a new theory from an exponential distribution with expected value equal to the inverse of π. This is the waiting time distribution corresponding to the Poisson process that explains the discovery of new theories. Thus, while Figure 10 represents one draw under our two cases, Figure 11 reports the averages of the number of ideas developed up to 50, 100, 150, 200, and 500 stages in the first or second case, along with the 95% confidence intervals, obtained from 100 rounds of simulations of each case. As the figure shows, not only does the high-speed case generate more ideas, but the gap across stages

Figure 10: Effects of γ, T, and π on the Growth of Ideas
($\lambda = 1$, $\alpha = 0.5$)



(a) $\gamma = 0.10$, $T = 0.9$, $\pi = 0.02$ (b) $\gamma = 0.20$, $T = 1.1$, $\pi = 0.05$

The figure shows the growth of ideas ($\lambda_i t_i^{\alpha}$, with $\alpha = 0.5$) with low speed, low resources, and the corresponding low π (prospect a), or high speed, high resources, and corresponding high π (prospect b). Appendix A4 reports how we computed the growth of ideas from simulating our model given the parameters. The figure shows that with low speed, resources, and π, cycles are longer, less frequent, and they tend to decline across stages. The opposite is true with high speed, resources, and π. Low speed also produces fewer innovations. As discussed in the text, high speed also produces more ideas, and a growing gap in the number of ideas across stages.

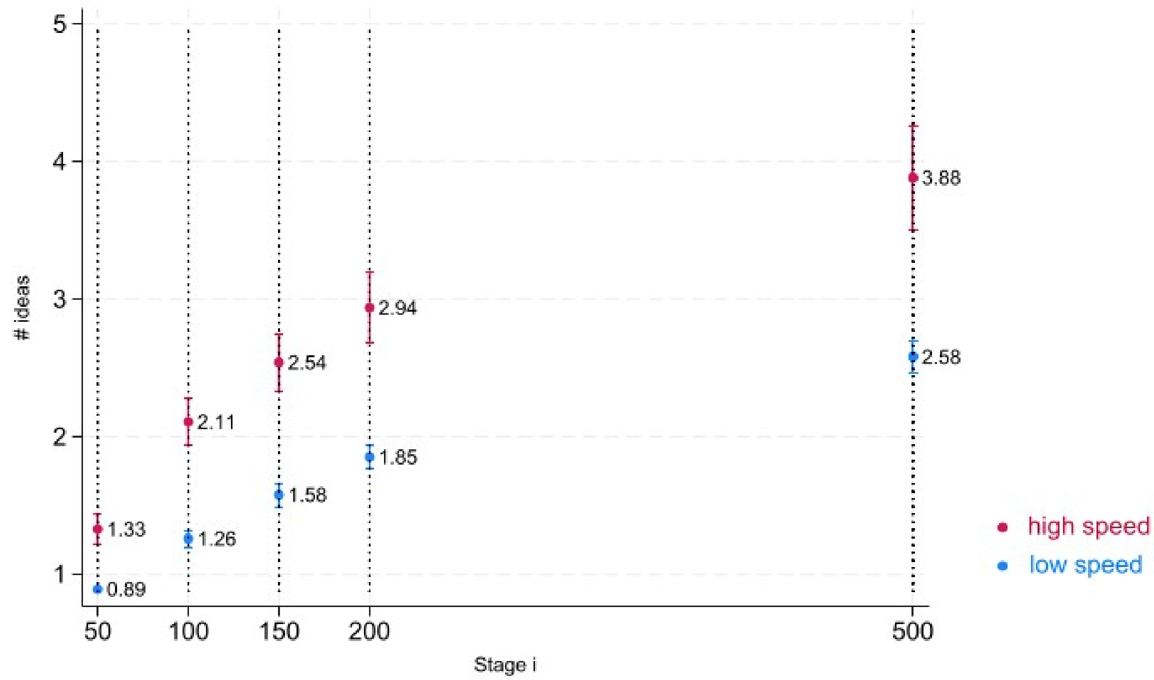
widens. Our model predicts that speed, and the associated endogenous generation of π and T, can increase gaps across firms.

6 Discussion

Our model offers several insights into key issues discussed in the literature and the current dynamics of firms, industries, and technologies.

First, we provide an explanation for Christensen's (1997) paradox, which posits that established firms often miss new opportunities. Our model shows that in a slow search environment, it is optimal for firms to focus on current theories and postpone the search for new ones. Conversely, in a fast search environment, Christensen's conditions do not apply, and limited investment in new theories is sub-optimal.

Figure 11: Speed and Growing Gaps in Ideas



The figure reports graphically the estimated average # of ideas, and their 95% confidence intervals, after 50, 100, 150, 200, 500 stages under low and high speed (respectively $\gamma = 0.10$, $\pi = 0.02$, $T = 0.9$, and $\gamma = 0.20$, $\pi = 0.05$, $T = 1.1$). We obtained averages and confidence intervals from 100 rounds of simulations of each one of the to cases.

Since at least Arrow (1962), it has been understood that firms are unlikely to invest in basic research due to its lack of immediate application and the spillovers it generates. Arrow concluded that basic research must be publicly funded and considered the presence of individuals and institutions that enjoy conducting research and combining it with teaching a "lucky accident." Nelson (1959) argued that large, diversified firms primarily conduct basic research because they can internalize its broad spillovers. Rosenberg (1990) added that firms invest in basic research with their own funds to develop the absorptive capacity needed to understand public scientific research.

Our model makes the novel prediction that in a high search speed environment, firms will find it optimal to invest in basic research to discover new theories, a possibility not found in the existing literature. Although the model does not account for spillovers, which could reduce firms' incentives to invest in basic research, the influence of high search speed is likely to counteract this effect. Firms may focus on generating intellectual property from their basic research, potentially hindering their engagement with academia, as noted by Rosenberg.

This prediction aligns with Yue (2023), the first large-scale analysis showing a positive effect of corporate involvement on scientific progress. Focusing on AI researchers affiliated with both universities and firms, Yue (2023) reveals a significant increase in citations for papers with corporate involvement, suggesting this leads to more groundbreaking research, albeit aligned with corporate interests. In contrast to Arora et al. (2021a) , who noted a decline in U.S. firms' basic research investments from 1980-2015 due to low-speed sectors, Yue (2023) highlights the growing importance of basic research in the high-speed AI sector.

Third, our discussion aligns with Ludwig & Mullainathan (2024). They show in the context of judicial decisions that hypothesis generation, traditionally reliant on human intuition, can be systematically enhanced using AI. In our model innovations in AI such as Large Language Models can accelerate the search process by enhancing the productivity of generating new theories.

Fourth, the ability to discover new theories can be a significant source of comparative advantage in a high-speed environment, as a firm's value (V) increases directly with the probability (π) of discovering new theories.[4] Firms investing more in basic research or enhancing their ability to increase π will enjoy significant competitive advantages. In

[4]Conversely, in a low-speed environment, the value of the firm does not benefit from a high probability π of discovering new theories, making differences in discovery theory irrelevant.

equilibrium, firms with higher π create higher value (V), which in turn increases the resources (T) they reinvest in research. This creates a virtuous cycle explaining the high growth potential of firms investing in new theories and the hyper-growth of many high-tech firms today, especially compared to other sectors.

External shocks that increase search speed can trigger these virtuous cycles, where faster search speeds lead to higher discovery rates, which increase reinvestment resources, thereby widening gaps among firms, industries, or even countries. As shown by Corollary 1 to Theorem 1, a higher initial stock of ideas (λ) or a lower opportunity cost of capital (T) can widen these gaps. To reduce disparities or leverage new growth cycles, external interventions may be needed to create the right conditions, triggering investments in high search speed (γ), higher discovery rates (π), and greater resource accumulation (T) for reinvestment. Conversely, in a low-speed environment, the value of the firm does not benefit from a high probability (π) of discovering new theories, making differences in discovery ability irrelevant.

7 Conclusions

In this paper, we introduce a novel framework for understanding innovation by focusing on search speed—how quickly firms explore and exhaust ideas within a technological paradigm. This concept underscores the balance firms must strike between exploiting existing knowledge and exploring new paradigms, a crucial but often underexplored aspect of innovation. Through a multi-stage search model, we shed light on how firms optimally allocate resources between searching within an existing technological framework and seeking out new theoretical paradigms. This approach provides insights into the strategic trade-offs firms face in driving innovation forward.

A key finding is the identification of a critical threshold effect. When the efficiency of searching within the existing paradigm is low (low γ), firms may become trapped in a cycle of focusing on the existing paradigm. This is because the weak diminishing returns associated with the search within the existing paradigm make investment in exploring new theoretical frameworks unattractive, even with a high probability of discovering new paradigms (high π).

However, the model also shows that the optimal search strategy shifts dramatically when the speed of search increases beyond a threshold. In such scenarios, firms can effectively explore both avenues. A higher search speed within the existing paradigm (high γ) motivates reallocating resources from later stages of the search within the existing technological paradigm to earlier stages of exploration of new theories with greater innovation potential. The greater availability of resources at later stages then opens up opportunities to explore new theories. The probability of discovering new paradigms (π) becomes a critical success factor.

A higher γ can then lead to a "bifurcated world" in which the complementarity between new theories frontier π and continuous advance in search frontier γ create exponential differential performance.

The model links future search resources to the value created by past innovations highlighting a critical interplay between search efficiency and the potential value of new theoretical paradigms. When search within the existing paradigm is slow, the ability to discover new theories has minimal impact on firm performance. However, when search is fast, differences in the ability to discover new theories become a major source of performance variation between firms. This underscores the importance of not only efficient search within the existing paradigm but also the ability to explore entirely new theoretical frameworks for sustained innovation success. Furthermore, speed of adoption and access to frontier technology in search γ or new theories π can become important geopolitical strategic variables (cognitive supply chain dependency) and access to the frontier can in itself become a key geopolitical decision (for example open sourcing γ but keeping proprietary π).

Furthermore, by emphasizing the critical roles of search speed and theory generation potential, the model provides insights into how policymakers can foster innovation across industries. Policies that enhance search speeds among diverse firms and industries can promote equitable growth and reduce performance gaps. For example, reducing capital costs allows firms to direct more resources toward exploring new knowledge domains. Additionally, policies that incentivize basic research and support knowledge-sharing institutions can significantly enhance the generation of new theories, thereby enriching the pool of ideas available to firms.

The model has several limitations. Our current measure of search speed, derived from firm-level patent data, may not fully capture the complexity of the concept. Future research should consider incorporating more flexible, time-varying parameters to better reflect the dynamic nature of innovation cycles across different industries. Additionally, the model does not account for competition and spillover effects, which are crucial influences on innovation strategies. Integrating these factors would yield a more accurate understanding of how firms respond to external pressures and benefit from shared knowledge environments. Lastly, expanding the model to include inter-firm interactions could provide deeper insights into the broader innovation ecosystem, illustrating how competition and collaboration affect the speed and direction of innovation.

References

Acemoglu, D., & Restrepo, P. (2018). The Race Between Machine and Man: Implications of Technology for Growth, Factor Shares, and Employment. *American Economic Review*, 108(6), 1488-1542.

Acemoglu, D., & Restrepo, P. (2020). Automation and New Tasks: How Technology Displaces and Reinstates Labor. *Econometrica*, 88(3), 1305-1348.

Acemoglu, D., Akcigit, U., & Celik, M. A. (2022). Radical and Incremental Innovation: The Roles of Firms, Managers, and Innovators. *American Economic Journal: Macroeconomics*, 14(3), 199-249.

Acs Z.J., Audretsch, & D.B. (1988) Innovation in Large and Small Firms: An Empirical Analysis. *American Economic Review*, 78(4), 678–690.

Acs Z.J., Audretsch, D.B. (1990) *Innovation and Small Firms*. MIT Press, Cambridge, MA.

Aghion, P., & Howitt, P. (1992). A Model of Growth Through Creative Destruction. *Econometrica*, 60(2), 323-351.

Aghion, P., Bergeaud, A., & Van Reenen, J. (2023). The Impact of Regulation on Innovation. *American Economic Review*, 113(11), 2894-2936.

Agrawal, A., Gans, J., & Goldfarb, A. (2021). *The Economics of Artificial Intelligence: An Agenda*. University of Chicago Press, Chicago.

Agrawal, A., McHale, J., & Oettl, A. (2024). Artificial Intelligence and Scientific Discovery: A Model of Prioritized Search. *Research Policy*, 53(5), 104989.

Arora, A., Belenzon, S., & Sheer, L. (2021a). Knowledge Spillovers and Corporate Investment in Scientific Research. *American Economic Review*, 111(1), 871-898.

Arora, A., Belenzon, S., & Sheer, L. (2021b). Matching Patents to Compustat Firms, 1980–2015: Dynamic Reassignment, Name Changes, and Ownership Structures. *Research Policy*, 50(5), 104271.

Arora, A., Belenzon, S., Cioaca, L., Sheer, L., Shin, H.M., & Shvadron, D. (2024a). DISCERN 2.0: Duke Innovation & SCientific Enterprises Research Network [Dataset]. In Zenodo (CERN European Organization for Nuclear Research). https://doi.org/10.5281/zenodo.3594642.

Arora, A., Belenzon, S., Cioaca, L., Sheer, L., & Shvadron, D. (2024b). Back to the Future: Are Big Firms Regaining their Scientific and Technological Dominance? Evidence from DISCERN 2.0. draft.

Arrow, K. J. (1962). Economic Welfare and the Allocation of Resources for Invention. In *The Rate and Direction of Inventive Activity: Economic and Social Factors*, a conference of the Universities-National Bureau Committee for Economic Research and the Social Science Research Council, 609-626. Princeton University Press, Princeton, NJ.

Baumol, W.J. (2002) *The Free-Market Innovation Machine. Analyzing the Growth Miracle of Capitalism*. Princeton university press, Princeton, NJ).

Bloom, N., Jones, C., Van Reenen, J. & Webb, M. (2020). Are Ideas Getting Harder to Find?. *American Economic Review*, 110(4), 1104-44.

Bresnahan, T. F., & Trajtenberg, M. (1995). General Purpose Technologies: 'Engines of Growth'? *Journal of Econometrics*, 65(1), 83-108.

Camuffo, A., Gambardella, A. & Pignataro, A. (2024). Theory-Driven Strategic Management Decisions. *Strategy Science*, forthcoming (see also CEPR DP 17664v5).

Christensen, C. (1997). *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. Harvard Business School Press, Boston MA

Cohen, W. M., Levin, R. C., & Mowery, D. C. (1987). Firm Size and R&D Intensity: A Re-examination. *Journal of Industrial Economics*, 35(4), 543-565.

Cohen W.M., & Klepper, S. (1996a) A Reprise of Size and R & D. *Economic Journal*, 106(437), 925–951.

Cohen, W. M., & Klepper, S. (1996b). Firm Size and the Nature of Innovation within Industries: The Case of Process and Product R&D. *The Review of Economics and Statistics*, 78(2), 232-243.

Cohen, W.M. (2010) Fifty Years of Empirical Studies of Innovative Activity and Performance. Hall B. & Rosenberg N. (eds.) *Handbook of the Economics of Innovation*, Elsevier, Amsterdam, 129–213.

Cockburn, I.M., Henderson, R. & Stern, S. (2019). The Impact of Artificial Intelligence on Innovation: An Exploratory Analysis. In Agrawal, A., Gans, J., & Goldfarb, A. (eds.). *The Economics of Artificial Intelligence: An Agenda* University of Chicago Press, Chicago.

Cohen, W. M., & Levinthal, D. A. (1989). Innovation and Learning: The Two Faces of R&D. *The Economic Journal*, 99(397), 569-596.

De Massis, A., Audretsch, D., Uhlaner, L., & Kammerlander, N. (2018). Innovation with Limited Resources: Management Lessons from the German Mittelstand. *Journal of Product Innovation Management*, 35(1), 125-146.

Dosi, G. (1982). Technological Paradigms and Technological Trajectories: A Suggested Interpretation of the Determinants and Directions of Technical Change. *Research Policy*, 11(3), 147-162.

Furman, J., & Seamans, R. (2019). AI and the Economy. *Innovation Policy and the Economy*, 19, 161-191. National Bureau of Economic Research, Cambridge.

Gimenez-Fernandez, E. M., Sandulli, F. D., & Bogers, M. (2020). Unpacking Liabilities of Newness and Smallness in Innovative Start-ups: Investigating the Differences in Innovation Performance between New and Older Small Firms. *Research Policy*, 49(10), 104049.

Hall, B.H. & Rosenberg, N. (2010). *Handbook of the Economics of Innovation.* Elsevier, Amsterdam.

Hall, B.H. & Helmers, C. (2024). *The Economics of Innovation and Intellectual Property.* Oxford University Press, Oxford.

Kuhn, T. (1962). *The Structure of Scientific Revolutions.* University of Chicago Press, Chicago IL.

Lehmann, E. E., Schenkenhofer, J., & Wirsching, K. (2019). Hidden Champions and Unicorns: A Question of the Context of Human Capital Investment. *Small Business Economics*, *52*, 359-374.

Levinthal, D. A., & March, J. G. (1993). The Myopia of Learning. *Strategic Management Journal*, 14(S2), 95–112.

Ludwig, J. & Mullainathan, S. (2024). Machine Learning as a Tool for Hypothesis Generation. *Quarterly Journal of Economics* 139(2), 751-827

Mansfield, E. (1981) Composition of R and D Expenditures: Relationship to Size of Firm, Concentration, and Innovative Output. *Review of Economic Statistics*, 63(4), 610–615.

McElheran, K., Li, J. F., Brynjolfsson, E., Kroff, Z., Dinlersoz, E., Foster, L., & Zolas, N. J. (2024). AI Adoption in America: Who, What, and Where. *Journal of Economics and Management Strategy*, 33(2), 375-415.

Nelson, R. R. (1959). The Simple Economics of Basic Scientific Research. *Journal of Political Economy*, 67(3), 297-306.

Nelson, R. R. (2006). Reflections on "The simple economics of basic scientific research": Looking back and looking forward. *Industrial and Corporate Change*, 15(6), 903-917.

Nooteboom, B. (1994). Innovation and Diffusion in Small Firms: Theory and Evidence. *Small Business Economics*, *6*, 327-347.

Romer, P. M. (1990). Endogenous Technological Change. *Journal of Political Economy*, 98(5), S71-S102.

Rosen, R. J. (1991). Research and development with Asymmetric Firm Sizes. *The RAND Journal of Economics*, 22(3), 411-429.

Rosenberg, N. (1990). Why Do Firms Do Basic Research (With Their Own Money)? *Research Policy*, 19(2), 165-174.

Rosenberg, N. (1992). Scientific Instrumentation and University Research. *Research Policy*, 21(4), 381-390.

Rosenberg, N. (1994). *Exploring the Black Box: Technology, Economics, and History.* Cambridge University Press, Cambridge.

Scherer, F. M. (1965). Firm Size, Market Structure, Opportunity, and the Output of Patented Inventions. *American Economic Review*, 55(5), 1097-1125.

Trajtenberg, M. (2019). Artificial Intelligence as the Next GPT: A Political-Economy Perspective. In Agrawal, A., Gans, J., & Goldfarb (eds.). *The Economics of Artificial Intelligence: An Agenda* National Bureau of Economic Research, Cambridge.

Weitzman, M. L. (1979). Optimal Search for the Best Alternative. *Econometrica*, 47(3), 641-654.

Weitzman, M. L. (1998). Recombinant Growth. *Quarterly Journal of Economics*, 113(2), 331-360.

Yue, D. (2023). I, Google: Estimating the Impact of Corporate Involvment in AI Research. *working paper*, Harvard Business School.

Appendix

A1. Solving the Dynamic Program

In order to solve the dynamic program (1), first solve it for K to obtain

$$K = \frac{k}{1-\Theta} \tag{A1}$$

where

$$k \equiv \sum_{i=1}^{j-1} e^{-\delta i}\lambda_i T^{\alpha} + \sum_{i=j}^{\infty} e^{-\delta i} e^{-\pi(i-j)}\lambda_i t_i^{\alpha} \tag{A2a}$$

$$\Theta \equiv \sum_{i=j}^{\infty} e^{-\delta i} e^{-\pi(i-j)}\pi\left(T - t_i\right) \tag{A2b}$$

Using (A1), (A2a), and (A2b), firms choose t_i to $\max_{t_i} K$, which yields the *foc*

$$\alpha\lambda_i t_i^{\alpha-1}\left(1-\Theta\right) - k\pi = 0 \tag{A3}$$

The ratios between each term of (A3) with the equivalent terms of the *foc* of t_j yields foc (3) in the text.

The *foc* for t_j is (A3) with t_j in lieu of t_i, which yields

$$\alpha\lambda_j t_j^{\alpha-1}\left(1-\Theta\right)-k\pi=0 \tag{A4}$$

Replace k and Θ with (A2a) and (A2b) after replacing t_i with (3). The expressions for k and Θ will only depend on t_j. As discussed in the text, replace $t_j = T$. After some tedious calculations shown in the next section of this Appendix, obtain (4) in the text.

To obtain the optimal knowledge K, use (A1) and derive $\dfrac{k}{1-\Theta}$ from (A4) after replacing $t_j = T$. This yields (6) in the text.

A2. Derive *foc* (4)

Rewrite (A4) as

$$\lambda_j\left(1-\Theta\right)-\frac{1}{\alpha}t_j^{1-\alpha}k\pi=0$$

Using (A2a) and (A2b), replacing t_i with (3) and $t_j = T$, obtain

$$\lambda_1 j e^{-\gamma(j-1)}\left[1-e^{-\delta j}\pi\sum_{i=j}^{\infty}e^{-(\delta+\pi)(i-j)}T\left(1-\Gamma_{ij}^{\frac{1}{1-\alpha}}\right)\right]$$

$$-\frac{1}{\alpha}T^{1-\alpha}\lambda_1 e^{-\delta}\left[\sum_{i=1}^{j-1}ie^{-(\delta+\gamma)(i-1)}T^{\alpha}+je^{-(\delta+\gamma)(j-1)}\sum_{i=j}^{\infty}e^{-(\delta+\pi)(i-j)}\Gamma_{ij}^{\frac{1}{1-\alpha}}T^{\alpha}\right]\pi=0 \quad ;$$

$$je^{-\gamma(j-1)}-e^{-\delta}je^{-(\delta+\gamma)(j-1)}\pi T\sum_{i=j}^{\infty}e^{-(\delta+\pi)(i-j)}\left(1-\Gamma_{ij}^{\frac{1}{1-\alpha}}\right)$$

$$-e^{-\delta}\frac{1}{\alpha}\pi T\left[\sum_{i=1}^{j-1} ie^{-(\delta+\gamma)(i-1)} + je^{-(\delta+\gamma)(j-1)}\sum_{i=j}^{\infty} e^{-(\delta+\pi)(i-j)}\Gamma_{ij}^{\frac{1}{1-\alpha}}\right] = 0 \quad ;$$

$$je^{-\gamma(j-1)} - e^{-\delta}je^{-(\delta+\gamma)(j-1)}\frac{\pi T}{\alpha}\left\{\sum_{i=j}^{\infty} e^{-(\delta+\pi)(i-j)}\left[\alpha\left(1-\Gamma_{ij}^{\frac{1}{1-\alpha}}\right)+\Gamma_{ij}^{\frac{1}{1-\alpha}}\right]\right.$$
$$\left. + \frac{e^{(\delta+\gamma)(j-1)}}{j}\sum_{i=1}^{j-1} ie^{-(\delta+\gamma)(i-1)}\right\} = 0$$

Rewrite the last term as

$$\frac{e^{(\delta+\gamma)(j-1)}}{j}\sum_{i=1}^{j-1} ie^{-(\delta+\gamma)(i-1)} - \sum_{i=j}^{\infty}\frac{i}{j}e^{-(\delta+\gamma)(i-j)}$$

Since

$$\sum_{i=1}^{\infty} ie^{-(\delta+\gamma)(i-1)} = \sum_{i=1}^{\infty} e^{-(\delta+\gamma)(i-1)} - \frac{\partial}{\partial\gamma}\sum_{i=2}^{\infty} e^{-(\delta+\gamma)(i-1)} = \left[1-e^{-(\delta+\gamma)}\right]^{-2}$$

we can rewrite the *foc* as

$$je^{-\gamma(j-1)} - e^{-\delta}je^{-(\delta+\gamma)(j-1)}\frac{\pi T}{\alpha}\left\{\sum_{i=j}^{\infty} e^{-(\delta+\pi)(i-j)}\left[\alpha+(1-\alpha)\,\Gamma_{ij}^{\frac{1}{1-\alpha}}\right]\right.$$
$$\left. - e^{-\delta(i-j)}\Gamma_{ij} + \frac{e^{(\delta+\gamma)(j-1)}}{j\left[1-e^{-(\delta+\gamma)}\right]^2}\right\} = 0$$

which, after rearranging terms, and using $\frac{i}{j}e^{-(\delta+\gamma)(i-j)} \equiv e^{-\delta(i-j)}\Gamma_{ij}$, becomes (4) in the text.

A3. Proofs

Proof of Lemma 1. Since T is a resource constraint, the optimal t_1 is bounded. Therefore, $\mathcal{A} \implies t_1 = T$. If t_1 was the initial condition, (3) $\implies t_i = \Gamma_{i1}^{\frac{1}{1-\alpha}} t_1$, where $\Gamma_{i1} \equiv ie^{-\gamma(i-1)}$ is equal to 1 if $i = 1$ and increases till $i = \dfrac{1}{\gamma}$, after which it starts declining with i. Then, $\mathcal{A} \implies \left(t_i < T \implies i > \dfrac{1}{\gamma}\right)$ – that is, $i > \dfrac{1}{\gamma}$ is a necessary condition for $t_i < T$ – and $j \leqslant \dfrac{1}{\gamma} \implies t_j \geqslant T$. To streamline our analysis we rule out cases in which $j < \dfrac{1}{\gamma}$ and $j + 1 > \dfrac{1}{\gamma}$. Therefore, $\exists t_j$, with $j \geqslant \dfrac{1}{\gamma} : t_j \geqslant T$ & $t_{j+1} < T$, and we can approximate $t_j \approx T$. ■

Proof of Lemma 2. Since $\lim_{j\to\infty} je^{-\gamma(j-1)} = 0$, as $j \to \infty$ the first term of the LHS of (4) and Ω tend to 0. Therefore, the LHS of (4) tends to $-\frac{e^{-\delta}}{\left[1-e^{-(\delta+\gamma)}\right]^2} < 0$. As a result, $\mathcal{A}$ implies that either the optimal $j = 1$, or there is at least one $j \in (1, \infty)$ such that the LHS of (4) cuts the horizontal axis from above and the *foc* is satisfied. ■

Proof of Lemma 3. Using subscripts to denote derivatives, the sign of j_T or j_π is the same as the sign of the corresponding differential of the LHS of (4). The negative sign of j_T is straightforward. After taking the differential of the LHS of (4) with respect to π, replace the first term of (4) using the *foc*. The sign of this differential is the same as the sign of $-(1 + \Omega + \pi\Omega_\pi)$. In equilibrium, $\Omega > 0$; otherwise, from (7), firms choose $K_0 > K$, $\pi = 0$, and $j \to \infty$. It is easy to see that $\dfrac{\partial \psi_i}{\partial \pi} < 0$, which implies $\Omega_\pi < 0$. We make the fair assumption that $1 + \Omega > -\pi\Omega_\pi$, which implies $j_\pi < 0$. ■

Proof of Theorem 1. The strategy to prove this theorem is to show, first, that given $\pi > 0$, if γ is sufficiently small, $\Omega < 0$ and, using subscripts to denote derivatives, $\Omega_\pi < 0$. Thus, from (7), $K < K_0$, the given π is not feasible, and increases in π cannot raise Ω to make $K > K_0$. Second, we show that increases in γ raise Ω and eventually turn $\Omega_\pi > 0$, making $\Omega > 0$ and $K > K_0$ feasible.

We start with the first part. Using (4), the sign of Ω_π is the same as the sign of

$$j_\pi \left[1 - j\left(\delta + \gamma\right)\right] \Psi_j + j \frac{\partial \Psi_j}{\partial \pi} \tag{A5}$$

where, from Lemma 3, $j_\pi < 0$, and, from Lemma 1, $j \geqslant \dfrac{1}{\gamma} \implies 1 - j\left(\delta + \gamma\right) < 0$. Thus, the sign of the first term depends on the sign of Ψ_j.

Let $\pi > 0$. When γ is small, approximate $\psi_i \approx \psi_i^* \equiv e^{-\pi(i-j)}\left[\alpha + (1-\alpha)\left(\dfrac{i}{j}\right)^{\frac{1}{1-\alpha}}\right] - \dfrac{i}{j}$, which is ψ_i when $\gamma = 0$. Obtain

$$\frac{\partial \psi_i^*}{\partial i} = -\pi e^{-\pi(i-j)}\left[\alpha + (1-\alpha)\left(\frac{i}{j}\right)^{\frac{1}{1-\alpha}}\right] - \frac{1}{j}\left[1 - \left(\frac{i}{j}\right)^{\frac{\alpha}{1-\alpha}} e^{-\pi(i-j)}\right] < 0$$

To establish this negative sign, we only need to show that the term in the second square bracket is positive. When $i = j$, this term is equal to 0. As i increases, it increases if $i > \dfrac{\alpha}{(1-\alpha)\,\pi}$. From Lemma 1, $j \geqslant \dfrac{1}{\gamma}$, which is larger than $\dfrac{\alpha}{(1-\alpha)\,\pi}$ if $\gamma \to 0$. As a result, for $i > j$ the term in the second square bracket is positive, making $\dfrac{\partial \psi_i^*}{\partial i} < 0$. Moreover, for $i = j$, $\psi_j^* = 0$, and $\psi_j \approx 0$. Therefore, $\forall i > j$, $\gamma \to 0 \implies \psi_i < 0 \implies \Psi_j < 0 \implies \Omega < 0$.

To check the second term of (A5), the direct effect of π on ψ_i is clearly negative. Taking into account the indirect effect through j, ignore the terms $i - j$ because they are counters that start at 0. However, $\dfrac{i}{j}$ differs if j differs, and we take the differential of ψ_i with respect to the j appearing in the denominator of Γ_{ij}. To show that $\dfrac{\partial \Psi_j}{\partial j} > 0$, the sign of this differential is the same as the sign of $1 - e^{-\pi(i-j)}\Gamma_{ij}^{\frac{\alpha}{1-\alpha}}$. This expression is equal to 0 for $i = j$ and positive for $i > j$ because it increases with i if $i > \dfrac{1}{\gamma + \frac{(1-\alpha)\pi}{\alpha}}$, a threshold smaller than the threshold $\dfrac{1}{\gamma}$ of Lemma 1 such that $i \geqslant j \geqslant \dfrac{1}{\gamma}$. However, since j declines because $j_\pi < 0$, then, for $i \geqslant j$, each term ψ_i in the summation sign that generates Ψ_j, declines. As a result, both the direct and indirect effect of π reduces ψ_i and therefore $\frac{\partial \Psi_j}{\partial \pi} < 0$. Note that this result obtains $\forall \gamma \geqslant 0$.

Putting things together, when $\gamma \to 0$, the sign of (A5) is negative because $\Psi_j < 0$ and $\frac{\partial \Psi_j}{\partial \pi} < 0$. The negative sign of Ψ_j implies $\Omega < 0$. Thus, for γ small, $\Omega < 0$ and it can only increase by making π smaller. The given $\pi > 0$ is unfeasible.

The next step is to show that a higher γ can turn Ω from negative to positive, making this π feasible. The sign of Ω depends on the sign of Ψ_j. The direct effect of γ is $-\left[1 - e^{-\pi(i-j)}\Gamma_{ij}^{\frac{\alpha}{1-\alpha}}\right]\frac{\partial \Gamma_{ij}}{\partial \gamma} > 0$. This effect is positive because it is not difficult to see that $\frac{\partial \Gamma_{ij}}{\partial \gamma} < 0$, and, using Lemma 1, $j \geqslant \frac{1}{\gamma} \implies e^{-\pi(i-j)}\Gamma_{ij} < 1$. The indirect effect of γ affects Ψ_j through j. We established that $\frac{\partial \Psi_j}{\partial j} > 0$. For the sign of j_γ (the optimal change of j with respect to γ), we take the differential of the LHS of (4) and replace *foc* (4) as we did for the sign of j_T and j_π in Lemma 3. After some algebra, the sign of j_γ is the same as the sign of $-\left[j - \frac{1+e^{-(\delta+\gamma)}}{1-e^{-(\delta+\gamma)}}\right] - \Omega\frac{\partial \Psi_j}{\partial \gamma}\Psi_j^{-1}$. We established above that the direct effect $\frac{\partial \Psi_j}{\partial \gamma} > 0$. A small γ, which implies $j \geqslant \frac{1}{\gamma}$ is a necessary condition to make the term in the square bracket positive. If this positive term is sufficiently large to dominate the sign of the indirect effect, both the direct and indirect effects of a higher γ can turn Ψ_j, and therefore Ω, from negative to positive. Otherwise, the effect is ambiguous and it will be positive depending on the strength of the direct effect vis-à-vis the indirect effect.

Finally, we need to establish whether, even if $\Psi_j \geqslant 0$, the given $\pi > 0$ is sustainable. The variation of Ω with respect to π depends on the variation of Ψ_j with respect to π and the variation of the first term of Ω with respect to j, which is affected by π. We showed that $\frac{\partial \Psi_j}{\partial \pi} < 0$, $\forall \gamma \geqslant 0$. The first term, $je^{-(\delta+\gamma)(j-1)}$ declines with $j > \frac{1}{\delta+\gamma}$, which is smaller than the threshold $j \geqslant \frac{1}{\gamma}$ established by Lemma 1. Since $j_\pi < 0$, the first term increases with π. Thus, for γ such that $\Psi_j = 0$, the overall variation of Ω with respect to π is still negative, and a higher γ is necessary to produce a sufficiently large $\Psi_j > 0$ that increases Ω as π increases. At this level of γ, which we label γ', $\gamma > \gamma'$ makes the given π sustainable. ■

Proof of Corollary 1. The proof of this corollary is straightforward from the proof of Theorem 1. A sufficiently high speed ($\gamma > \gamma'$) is necessary for Ω to increase with π. Moreover, Theorem 1 established that if Ω increases with π, then $\Omega > 0$. From (7), this implies $K > K_0$, and that K increases with π. ■

Proof of Theorem 2. Using subscripts to denote derivatives, the *foc* of $V(K, T)$ with respect to T is $V_K K_T - V_T = 0$. Using (7), and the expression for K_0, obtain $K_T = K_{0T}(1 + \Omega) + K_0\Omega_T$. It is easy to see that $K_{0T} > 0$, and from Lemma 3, $j_T < 0 \implies \Omega_T > 0$, We can fairly assume that the cross-partials of Ω_T with respect to π and γ are negligible. We can then write $K_{T\gamma} = K_{0T\gamma}(1 + \Omega) + K_{0T}\Omega_\gamma + K_{0\gamma}\Omega_T$, and $K_{T\pi} = K_{0T}\Omega_\pi$ (since $K_{0\pi} = 0$). It is easy to see that $K_{0T\gamma}, K_{0\gamma} > 0$. Moreover, in the proof of Theorem 1 we set a necessary condition for $\Omega_\gamma > 0$. If so, when $\Omega_\gamma < 0$, firms can reduce γ to make $\Omega_\gamma = 0$, and optimize with respect to the excessively high γ. As a result, $\Omega_\gamma \geqslant 0$ and $K_{T\gamma} > 0$. As far as $K_{T\pi}$ is concerned, Corollary 1 established that if $\gamma > \gamma'$, $\Omega_\pi > 0$ (and therefore $K_\pi > 0$); otherwise $\Omega_\pi < 0$. As noted above, firms can always reduce π if suboptimal, and thus $\gamma \leqslant \gamma' \implies \Omega_\pi = 0$. Therefore, $\gamma > \gamma' \implies K_{T\pi} > 0$, and $\gamma < \gamma' \implies K_{T\gamma} = 0$. Finally, $K_{\pi\gamma} = K_0\Omega_\pi + K_{0\gamma}\Omega_{\pi\gamma} \geqslant 0$ because we established that $\Omega_\pi, \Omega_{\pi\gamma} \geqslant 0$, with $K_{\pi\gamma} > 0$ if $\gamma > \gamma'$. ■

A4. Construction of Figure 10

We constructed Figure 10 as follows. Let $i \in [1, \infty)$ and $c \in [1, \infty)$ be integers that denote, respectively, stages and cycles. A cycle is the set of all stages in which firms search within a given theory. It comprises all stages in which firms focus only on searching within the given theory, which produces an expected number of ideas $\lambda_i T^\alpha$ in each stage i in which they do not search for new theories, and an expected number of ideas $\lambda_i t_i^\alpha$ in each stage in which they also search for a new theory but have not found it yet, with t_i defined by (3) and $t_j = T$, as discussed in the text and in Appendix A1.

We now show the derivation of the number of ideas by cycles, which we use to draw the curves in Figure 10. We define $x_i \equiv \lambda_i T^\alpha$ to be the expected number of ideas in each stage of a given cycle produced when the firm is not searching for a new theory, and $y_i \equiv \lambda_i t_i^\alpha$ the equivalent expected number when they also search for new theories. For simplicity, we set $\alpha = 0.5$, as in Figure 10.

Also, let q_c, $c \in [1, \infty)$, be a random draw from an exponential distribution with expected value equal to the inverse of π. Since we assumed that firms discover new theories according to a Poisson process, this is a drawn from the corresponding waiting time distribution. It represents the stochastic timing of the discovery of new theories after firms have started to search for them.

Finally, in what follows, j is the optimal number of stages after which firms start searching for a new theory since the start of the search in the current theory.

Cycle 1

Let $\mu_1 \equiv \lambda \left(1 - e^{-\gamma}\right)^2$. Then:

$$
\begin{aligned}
x_i &= \mu_1 i e^{-\gamma(i-1)} T^{\frac{1}{2}} && 1 \leqslant i < j_1 \equiv j \\
y_i &= x_{i+1} \frac{i}{j_1} e^{-\gamma(i-j_1)} && j_1 \leqslant i \leqslant j_1 + q_1 \equiv i_1
\end{aligned}
$$

Cycle 2

We now have $\mu_2 = \mu_1 T \left[1 - \left(1 + \frac{q_1}{j_1}\right)^2 e^{-2\gamma q_1}\right]$. This expression stems from our dynamic program (1). The pool of potential ideas of each theory grows according to $T - t_i$, as stated in (2), with t_i being the optimal t_i in the stage in which the firm discovers the new theory. The theory in cycle 2 is discovered in stage i_1, which, using (3) and $t_j = T$, yields the expression for μ_2. We then have:

$$x_i = \mu_2 (i - i_1) e^{-\gamma(i-i_1-1)} T^{\frac{1}{2}} \qquad i_1 + 1 \leqslant i < j_1 + i_1 \equiv j_2$$
$$y_i = x_{i+1} \frac{i}{j_2} e^{-\gamma(i-j_2)} \qquad j_2 \leqslant i \leqslant j_2 + q_2 \equiv i_2$$

<u>Cycle 3</u>

Following the same logic, $\mu_3 = \mu_2 T \left[1 - \left(1 + \frac{q_2}{j_2}\right)^2 e^{-2\gamma q_2}\right]$, and

$$x_i = \mu_3 (i - i_2) e^{-\gamma(i-i_2-1)} T^{\frac{1}{2}} \qquad i_2 + 1 \leqslant i < j_2 + i_2 \equiv j_3$$
$$y_i = x_{i+1} \frac{i}{j_3} e^{-\gamma(i-j_3)} \qquad j_3 \leqslant i \leqslant j_3 + q_3 \equiv i_3$$

A5. Construction of the Patent Clusters

Selection of Machine Learning Algorithm for Patent Clustering

The goal of the algorithm is to cluster patents issued in the US between 1980 and 2021 based on the CPC patent classification. Patents can generally be classified in one or multiple CPC categories, which leads to specific combinations of CPC classifications. Patents issued with the same or similar combinations are assumed to belong to the same field of research. When a company issues a patent with an identical or similar combination of CPC classifications, it can be assumed that the company is continuing the research in the same field. When a company issues a patent with a new combination of CPC classifications, it can be assumed that the company is exploring a new domain.

We approached this classification problem by comparing and analyzing different clustering machine learning algorithms on a dataset that only contains the patents' application IDs and the CPC classification up to the subclass level. The CPC classifications are codes for defined fields of patents and represent categorical data. The data is pre-processed by encoding the categorical CPC classifications for the use of machine learning algorithms. A CPC classification will be assigned to the respective CPC section, CPC class, and CPC subclass. Thus, patents having multiple classifications can have positive values in multiple columns. For instance, a patent having two CPC classifications A08A and A08B, would lead to a count of 2 in the A CPC section, 2 in the A08 CPC class, and 1 in each CPC subclass. This assignment was chosen based on preliminary tests with subsets of the data and assigning a CPC classification not only to one subclass but to different levels. The results led to more stringent clustering with more patents in the miscellaneous cluster and more clusters in comparison to the clustering with the CPC subclass only. The binary encoding of CPC section, class, and subclass resulted in even more miscellaneous patents and approximately the same number of clusters. Verifying the clusters manually, the non-binary encoded data including CPC section, CPC class, and CPC subclass resulted in the most consistent clusters.

Aiming to cluster patents based on the combination of CPC classifications, the number of clusters should be determined by the patent combinations themselves and should not be manually imposed. Moreover, the number of clusters should be treated as a result rather than as input. There are different ways to determine the number of clusters, such as the elbow method or the silhouette method. For subsamples of the patent data, the elbow method resulted in a deficient number of clusters, which is too coarse for the analysis.

The silhouette method resulted in suggested parameters that yield too many and too small clusters. For the purpose of our analysis, the clusters should not be represented by too few patents since it would not capture domains but the intricacies of the patent combinations. Consequently, clustering algorithms requiring the number of clusters as input, such as K-Means, are subsequently not further considered.

Additionally, the clustering algorithm has to be suitable for categorical data which is in our case encoded in a binary way. Various algorithms were tested on a data subset but the clusters were in most cases either combined in less than five clusters or more than 1000 clusters with the majority of clusters containing only one patent. For the purpose of our analysis, we would like to have meaningful clusters containing multiple similar patents, with no cluster containing a considerable fraction of the overall data set. We decided to test the following three clustering algorithms in more detail: DBSCAN, HDBSCAN, and Agglomerative Clustering.

First, DBSCAN evaluates the proximity of points and classifies points with a sufficient number of other points within its perimeter as core points and points with an insufficient number of points within its perimeter as noise. A cluster is created by iteratively grouping core points that are density-connected and treats points out of the proximity of core points as noise (Ester, Kriegel, Sander & Xu, 1996). The advantages of DBSCAN are that clusters can have arbitrary shapes and sizes, the number of clusters is automatically determined, and the algorithm is robust to outliers and noise. The relevant disadvantage for this dataset is that the algorithm may experience issues with clusters of similar density, and its scalability is limited for high dimensional datasets (Lytvynenko, Lurie, Krejci, Voronenko, Savina & Taif, 2019), which is the case with the categorical patent data.

Secondly, HDBSCAN is an extension of DBSCAN that includes single-linkage hierarchical clustering in addition to the DBSCAN mechanism. Thus, HDBSCAN has the same advantages as DBSCAN and is more scalable in large dimensional data. The disadvantage is that the algorithm is more computationally intense than DBSCAN.

Thirdly, Agglomerative Clustering iteratively merges the two closest clusters until a stopping criterion is met. The algorithm applies a linkage method to define cluster proximity and a distance metric. The advantage of Agglomerative Clustering is the flexible clustering of complex shapes, which determines the number of clusters itself and provides a hierarchy of clusters that can be useful for visualization and interpretation. The disadvantages of Agglomerative Clustering are the sensitivity to noise and outliers, the computational intensity, and the scalability is challenging for high dimensional datasets (Zepeda-Mendoza & Resendis-Antonio, 2013). For the purpose of our research, we prefer to have more precise clustering with noise classified in a miscellaneous category rather than having the noise included in the clusters, which can lead to deficient analysis. The clustering algorithm of DBSCAN and HDBSCAN both yield good results in separating the noise. Moreover, HDBSCAN tends to perform well for patent data in a high dimensional domain in contrast to DBSCAN, and we accept computational intensity as a tradeoff for accuracy. Thus, HDBSCAN is the suitable clustering algorithm to use for our analysis.

Regarding the inputs of HDBSCAN, we set the minimum cluster size to 10 to ensure sufficient size for clusters and not too numerous clusters. Additionally, the default metric Euclidean is applied since it delivered better results for our research considering the number of clusters, the type of data, and the computation time. The other parameters are left on default.

Selection of Algorithm for Cluster Matching

Due to the computational requirements, the clustering was executed on a yearly basis. To match the clusters over time, the cosine-similarity algorithm was used to match the cluster ids from one year to the cluster ids of the next year while all patents in the miscellaneous cluster were excluded and were not assigned a consistent cluster id. This enables us to connect the clusters over multiple years and give a more precise measure for analysis purposes. The cosine-similarity defines the similarity of two clusters through the normalized dot product of the CPC classification features of clusters. The most similar clusters are matched with one another. Attention should be paid to the varying number of clusters per year, which leads to a small mismatch for some clusters and patents. Thus, about 2500 out of 1.32 million patents are not assigned to a consistent cluster because of this mismatch in the number of clusters per year. This problem only occurs in the first years but is nonexistent in later years since the number of consistent clusters grows over time and covers more patents.

Subsequently, the yearly clusters are connected through the matched cluster ids, based on the first year 1980. The first set of consistent clusters is given by the clusters in 1980 and clusters in later years with the same or similar patterns of CPC classifications are assigned to these consistent clusters. Since the CPC patterns evolve over time, new clusters not seen in prior years create a new consistent cluster id and are assigned to this id for subsequent years. For yearly clustering, this leads to a total of 1042 clusters.

Due to the yearly clusters changing and few observations per cluster, the consistent clusters work well for consistent clusters existing only for few years, but some clusters lasting for a long time, especially clusters lasting over 35 years, tend to have changes in the pattern, leading to a different CPC pattern for the patents in the last year in comparison to the patents in the first year assigned to the same constant cluster id. Thus, consistent clusters lasting many years have “breaks” in CPC patterns that capture different CPC patterns over time.

We also conducted a clustering process with a 5-year moving window, which allows for a matching of clusters based on the overlap of the years in the clustering window. Thus, a patent generally was allocated to five clusters, except for patents at the beginning or the end of the sample. Thus, connecting the cluster ids of the individual clustering leads to sound consistent cluster ids. The consistent cluster ids lead to precise results and no observed breaks but leads to some large clusters in which many firms operate and conduct research. Examples for extensive clusters are G06F (Electrical Digital Data Processing) and H04L (Transmission of Digital Information). Using these 5-year moving window clusters does not change in meaingful ways the results of our analysis.

References

Ester, M., Kriegel, H.P., Sander, J., & Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96).*, https://www.dbs.ifi.lmu.de/Publikationen/Papers/KDD-96.final.frame.pdf.

Lytvynenko, V., Lurie, I., Krejci, J., Voronenko, M., Savina, N., & Taif, M. A. (2019). Two Step Density-Based Object-Inductive Clustering Algorithm. *Workshop Proceedings of the 8th International Conference on "Mathematics. Information Technologies. Education," MoMLeT&DS-2019*, https://ceur-ws.org/Vol-2386/paper10.pdf.

Zepeda-Mendoza, M.L., & Resendis-Antonio, O. (2013). Hierarchical Agglomerative Clustering. *Encyclopedia of Systems Biology.*, Springer, New York, NY. https://doi.org/10.1007/978-1-4419-9863-7_1371.

Comparison of Algorithms

	DBSCAN	HDBSCAN	Agglomerative Clustering
Inputs	eps=0.5, min_samples=10	min_cluster_size=10, metric='euclidean'	n_clusters=None, distance_threshold=10, compute_full_tree=True
Miscellaneous category for noise	Yes, cluster -1	Yes, cluster -1	No
Subsample (n = 16,182)			
Number of clusters	252	297	220
10 largest clusters	-1 7559 10 417 9 345 7 174 52 168 45 162 106 157 20 146 110 137 65 129	-1 5180 259 417 263 346 257 186 139 174 124 168 260 162 172 154 122 146 289 137	159 417 110 345 6 290 33 259 57 223 56 220 37 211 29 181 5 174 7 173
10 smallest clusters	232 10 100 10 230 10 198 10 90 10 200 10 202 10 42 10 220 10 78 10	100 11 77 11 17 11 73 11 147 11 115 11 151 11 168 11 35 11 185 10	163 18 124 17 101 17 67 17 103 17 170 16 213 15 197 9 143 6 207 4

Comparison of Encoding

	Only Subclass	All classifications – Binary Encoding	All classifications – Non-Binary Encoding
CPC Classifications included	CPC subclass	CPC section, CPC class, CPC subclass	CPC section, CPC class, CPC subclass
Encoding	Binary	Binary	Non-Binary
Subsample (n = 16,182) with HDBSCAN(min_cluster_size=10, metric='euclidean')			
Number of rows of dataset	578 columns	710 columns	710 columns
Number of clusters	265	298	297
10 largest clusters	-1 4837 227 452 251 369 231 215 178 186 98 181 214 177 194 176 154 175 260 174	-1 5848 295 417 293 345 149 192 238 170 199 162 188 158 178 147 292 138 159 130	-1 5180 259 417 263 346 257 186 139 174 124 168 260 162 172 154 122 146 289 137
10 smallest clusters	196 11 254 11 60 11 2 11 0 11 72 10 151 10 31 10 28 10 38 10	14 11 261 10 231 10 272 10 116 10 42 10 53 10 96 10 74 10 165 10	100 11 77 11 17 11 73 11 147 11 115 11 151 11 168 11 35 11 185 10

Consistent Cluster ID 5-year Window

	Consistent cluster ID	Number of patents
Largest 10 clusters	14490.0	108935
	14482.0	39509
	14523.0	38702
	14466.0	18844
	14572.0	15337
	14375.0	14295
	14376.0	13998
	14245.0	12604
	14576.0	12049
	14223.0	11906
Smallest 10 clusters	14497.0	31
	14123.0	31
	13792.0	30
	3807.0	30
	14251.0	30
	13595.0	29
	8170.0	27
	14193.0	26
	14398.0	20
	14111.0	19

Number of consistent clusters: 1023