HARVARD | BUSINESS | SCHOOL

The Evolutionary Nature of Breakthrough Innovation: Re-Evaluating the Exploration vs. Exploitation Dichotomy

Dominika K. Sarnecka
Gary P. Pisano

Working Paper 21-071

The Evolutionary Nature of Breakthrough Innovation: Re-Evaluating the Exploration vs. Exploitation Dichotomy

Dominika K. Sarnecka
Harvard Business School

Gary P. Pisano
Harvard Business School

Working Paper 21-071

Funding for this research was provided in part by Harvard Business School.

The Evolutionary Nature of Breakthrough Innovation:

Re-Evaluating the Exploration vs. Exploitation Dichotomy

Dominika K. Sarnecka*

Gary P. Pisano**

Harvard Business School

December 1, 2020

* Ph.D. Candidate in Organizational Behavior, Harvard Business School

**. Harry E. Figgie Professor of Business Administration and Senior Associate Dean, Harvard Business School

The authors wish to thank Xiang Ao and Patrick Clapp of the Harvard Business School for their assistance with data collection and computational consultation. We would also like to gratefully acknowledge the financial support of the Harvard Business School Division of Faculty Research and Development. All errors and omissions are the responsibility of the authors.

Abstract

Over the past few decades, a consensus has emerged that breakthrough innovations emerge from exploration of novel terrain while more routine innovations are the product of exploitation. In this paper, we revisit this explore versus exploit dichotomy with an analysis of over two and half thousand firm-level innovation histories spanning three decades. Our data and a novel measure of search (Technological Focal Proximity) enable us to characterize at a detailed level the search strategies of firms and to examine breakthroughs and non-breakthroughs are associated with different search strategies. Using our novel firm-level data and method, we find (contrary to the existing literature) that breakthrough innovations evolve through a process involving *both* exploration (initially) and exploitation (subsequently). The breakthrough innovation process appears to evolve through phases. In the early phases, firms explore unfamiliar terrain. However, as the process unfolds, firms shift their search strategies to focus on exploiting cumulative knowledge. Our findings call into question the strong dichotomy between exploration *versus* exploitation that has played such a prominent role in thinking about the origins of breakthrough innovation, and have potential implications for strategy, organizational design, management practice, and corporate culture.

The Evolutionary Nature of Breakthrough Innovation: Re-Evaluating the Exploration vs. Exploitation Dichotomy

INTRODUCTION

"Successful ideas are the result of slow growth. Ideas do not reach perfection in a day, no matter how much study is put upon them. It is perseverance in the pursuit of studies that is really wanted."
Alexander Graham Bell

A rich theoretical tradition conceptualizes innovation as a search process in which organizations recombine and manipulate existing knowledge to produce new knowledge (Henderson and Clark 1990; March and Simon 1958; Nelson and Winter 1982; Schumpeter 1939). Innovating firms can be thus thought of as searching over a landscape of technological possibilities. As far back as Schumpeter (1939), scholars of innovation have been distinguishing between *exploratory* search of novel terrain versus *exploitative* search of more familiar ground (March 1991; Nelson and Winter 1982). This "exploration versus exploitation" dichotomy has played a prominent role in research on organizations in general, and on origins of breakthrough innovations in particular (Jung and Lee 2016; Rosenkopf and Nerkar 2001). A strong consensus has formed in both the theoretical and empirical literature that breakthrough innovations arise from exploration whereas less impactful innovations emerge from exploitation (Ahuja and Lampert 2001; Cohen and Levinthal 1990; Henderson and Clark 1090; Katila and Ajuja 2002; Kogut and Zander 1992; March 1991; O'Reilly and Tushman 2013).

However, an increasing number of case studies provides evidence that breakthroughs may emerge from a search process that is more nuanced and complex than previously thought. These case studies suggest that breakthroughs emerge through a process that combines *both* exploration and exploitation, and not one at the exclusion of the other. For instance, one of the most important inventions of the 20th century, the transistor, was created at the Bell Labs through a process that exhibited the classic hallmarks of both exploration (e.g. working on the frontiers of known science, experimentation with unprecedented design concepts, unpredictable paths forward, and failure) and exploitation (cumulative learning, incremental refinement and improvement of materials' purity, and iterative experimentation) (Gertner 2012; Riordan et al. 1999). Similarly, in aircraft design, Honda's unprecedented "over the wing engine" configuration started with an exploratory question: would it be feasible — contrary to the precepts of aeronautical engineering accepted by contemporary to that era scholars — to design an aircraft with engines mounted *over* the wings? Theoretical inquiry by the Honda team lead to a novel (and initially controversial) hypothesis: if the engine was mounted in exactly the right spot, the aerodynamic "interference" (predicted by existing theories) could be avoided. This initial hypothesis then triggered a wave of exploitative search (using computer simulations and physical prototype tests) to iteratively find "the sweet spot" on the wing where the critical airflows would "cancel each other out" (Pisano and Shulman 2018). The breakthrough engine-wing configuration required both exploration into unknown terrain and subsequent exploitation to iteratively find a solution. Similar patterns of breakthroughs emerging through a dual process of first exploration, followed by exploitation, can be found also in a range of other domains, for example life sciences (Luo et al 2018; Pisano 2006), liquid crystal displays (Kawato 2002), and material science (Jung and Lee 2016).

This paper tests the proposition that development of breakthrough innovations is a process sequentially involving *both* exploration and exploitation. We examine this premise using three decades of patenting histories of over two and a half thousand organizations (2559 firms), which, in aggregate, applied for over 1.3 million patents between 1975-2005. Because our dataset disambiguates names of patents' assignee firms, we are able to construct longitudinal histories of each firm's patent portfolio on a yearly basis. This enables us to characterize patterns of corporate search at a highly detailed level and over time. To conduct this analysis, we develop a new measure of innovative search, *Technological Focal Proximity*, that compares "the distance" in the theoretical knowledge space between any given patent (focal patent) and the assignee firm's patent portfolio in any prior year. Closer proximity between the technological content of a patent and a firm's past portfolio suggests a higher degree of exploitation, whereas a greater distance suggests a higher degree of exploration. In our main analysis, we take each firm's 2005 patents to be the focal sample and classify them as either breakthroughs or non-breakthroughs using methods existing in the literature (e.g., Singh and Fleming 2010). Then, we compute the distance between each of these patents and the firm's portfolio in every past year going back to 1975 (or the earliest possible year if the firm began applying for patents after 1975). Advantages of our analytical strategy over prior work on this topic are related to tracing back in time decades of firm-level patent portfolios for thousands of firms. This allows us to examine at a very detailed level the corporate search strategy behind each individual invention, while allowing for variation *within* the assignee firm and *between* many firms. In contrast, the majority of existing studies either do not examine the relationship between corporate breakthroughs and the inventor firms' prior inventive activities, or have focused on only a limited number of firms within a specific industry.

Our empirical analysis of decades of corporate innovation suggests that breakthrough innovation is an evolutionary process that combines both exploration and exploitation. We find that it begins with a period of exploration, which is then followed by a period of exploitation. Specifically, when firms innovate by first venturing into a distant area of the theoretical knowledge space and then focus their activities on that area and iterate within it, they are more likely to produce a breakthrough, than when following a different search process. Further, our results show that both the shape of the search trajectory pursued by a firm and the intensity of its search activity in the selected area influence the likelihood that the developed invention will be a breakthrough. Our findings have implications for understanding why some organizations may be more capable of achieving breakthrough innovations than others.

EXPLORATION AND EXPLOITATION IN BREAKTHROUGH INNOVATION

Innovation has long been conceptualized as a recombinant search process (March and Simon 1958; Nelson and Winter 1982; Schumpeter 1939) involving technological knowledge components. Schumpeter (1939:88) was the first to raise the idea that innovation might involve either new components or new combinations of existing components. This distinction has led researchers to classify innovation search processes according to the extent to which innovators use components (or component combinations) that are familiar or foreign. Exploitation means searching over component spaces that are well-known to the innovator (March 1991). Explorations means searching over unfamiliar spaces (March 1991).

A large body of theoretical and empirical work has suggested that breakthrough innovations are more likely to emerge through exploration than exploitation (Ahuja and Lampert 2001; Cohen and Levinthal 1990; Henderson and Clark 1990; Katila and Ahuja 2002; Kogut and Zander 1992; March 1991; Tushman and O'Reilly 1996). Exploration enriches the firm's knowledge pool (March 1991), enhances combinatory possibilities (Fleming and Sorenson 2001; Nelson and Winter 1982), decreases rigidity (Leonard-Barton 1995), and minimizes the threat of competency traps (Levitt and March 1988). The idea that breakthroughs require an organization to be good at exploration has given rise to a number of practical suggestions for management. For instance, O'Reilly and Tushman have argued that organizations need to be capable of exploitation to support routine innovation in their existing business and exploration to spawn disruptive inventions that drive transformation (2013). They call this characteristic "ambidexterity". Others have suggested that because of the trade-offs between exploration and exploitation, firms should isolate efforts to develop breakthrough innovations from the "mainstream" business maintained through exploitation (Benner and Tushman 2003; Bower and Christensen 1995; O'Reilly and Tushman 2013).

Yet, despite the strong consensus that breakthroughs are spawned through exploration, a number of case studies of corporate innovations provide evidence that the process may be more nuanced than portrayed in the extant literature. Examples of breakthrough innovation in such contexts as semiconductors (Gertner 2012; Riordan et al. 1999), LCD displays (Kawato 2002), aircraft and aeronautical engineering (Pisano and Shulman 2018; Vincenti 1990), life sciences (Luo et al 2018; Pisano 2006), and materials science (Jung and Lee 2016), all provide a picture of exploration and exploitation working in concert. Consider the well-known case of perhaps the most important breakthrough of the 20th century—the transistor invented at the Bell Labs in December 1947 and January 1948.[1] The first stage of the process could certainly be characterized as exploratory in many ways. Scientists at the Bell Labs were working on the frontiers of solid-state physics and made discoveries even in the realm of basic science (e.g. the discovery of surface states). They were creating, building, and testing completely unprecedented device designs and using combinations of materials in novel ways. The research leading up to the transistor had its share of failed experiments, blind alleys, and unexpected turns, which are all hallmarks of an explorative effort. However, following this exploratory phase, much of the process resembled classic exploitive search. Bardeen described the research leading up to creation of the successful transistor in his 1956 Nobel Lecture: "It was dependent both on the sound theoretical foundation, largely built up during the thirties, and on improvement and purification of materials, particularly of germanium and silicon, in the forties" (Bardeen 1956: 319). Throughout the program, the purity of various semiconductor materials (which at that point had been already known for decades) was continuously improving, device configurations were progressively evolving, and each promising design was iteratively altered. Shockley's initially failed field-effect proposal, for instance, stimulated further research leading to comprehension of surface states, a critical discovery in the transistor program. Similarly, Brattain and Bardeen's point-transistor provided Shockley with insight proving to be key in development of his junction transistor. Corporate knowledge behind the ultimate breakthrough invention thus evolved cumulatively through an iterative process of addressing gaps between experimental evidence and understanding. The picture we get from the Bell Labs' history and other accounts of processes behind corporate inventions is that of initial exploration followed by a period of intensive exploitation, not simply one or the other.

Scholars studying innovation using quantitative methods also provide some empirical evidence that when firms innovate within the context of their expertise, they can learn from past failures and

---

[1] There were actually two transistor inventions. The point-contact transistor invented by Walter Brattain and John Bardeen in December 1947 and the junction transistor invented by William Shockley in January 1948. Because the two inventions were developed only few months apart and were a part of the same corporate effort, we refer to these as a single invention.

leverage their deep understanding of the components' functionality. For example, Fleming found that inventions that incorporate technological components that have been frequently used in the past are more likely to serve as basis for future creations than those which do not (2001). Similarly, Jung and Lee found that, conditional on the type of knowledge contained in the used components, local search can outperform distant search in generating high-impact breakthroughs (2016).

Our proposition is that breakthrough innovation involves an iterative search process combining both exploration *and* exploitation. We will use the familiar landscape metaphor to illustrate our argument (see e.g. Fleming 2001). The technological knowledge landscape is a multi-dimensional space composed of a variety of technological components that firms can draw on to build inventions. At any point in time, a firm occupies a "place" in that landscape in terms of the combinations of components that it utilizes. During the innovative search process, firms scan the knowledge space and choose a number of components as inputs to the innovations they are creating. As noted earlier, exploitative search means looking at components and their combinations in the company's existing "neighborhood", whereas exploration means looking further afield. Firm knowledge is dynamic. It is constantly evolving as a result of the firm's search strategies (whether those result in successful innovations or not). Once explored, an unprecedented space becomes a familiar ground for further exploitation. As a result, the concept of exploration versus exploitation must itself be dynamic.

Prior approaches to analysis of corporate innovation

Evidence from researchers like Fleming (2001) connivingly demonstrate that the search process that a firm pursues to develop an invention effects its chances for success, which is most commonly measured as the novelty's impact on future technologies (e.g., Jung and Less 2016; Singh and Fleming 2010). However, what we do not know from this existing work is whether the path to these novelties is continuous or discontinuous, and how changes in the firm's knowledge base influence this process. Our hypothesis is that successful breakthrough innovation involves both leaps into novel spaces (exploration) and periods of intensive iterative exploitation in the selected areas. To test this proposition, we follow the well-established convention of using patent data to study corporate innovation (e.g., Henderson and Cockburn 1994; Jaffe, Trajtenberg, and Henderson 1993; Katila and Ahuja 2002; Rosenkopf and Nerkar 2001). However, our analytical approach differentiates our work from earlier literature on this topic. Specifically, we construct a novel measure of corporate innovation, *Technological Focal Proximity*, described in detail in the next section of this paper. We develop this measure to overcome challenges of prior research, which has been limited by methodological strategies and data constraints.

Unlike some of the past work in this area, our framing of the problem focuses heavily on the nature of search occurring within firms. That is, we view the innovating firm as the critical searching entity. This perspective is different from research that has analyzed corporate innovation without reference to the firms engaging in the search process (e.g., Fleming 2001). Studies in this vein abstract from firms and focus solely on the content and antecedents of a patent (irrespective of whether those antecedent inventions were generated by the firm that created the patent in question or not). Prior literature that assumes this firm-agnostic perspective can be broadly divided into two analytical approaches. First, some researchers look at the process of how knowledge is recombined at a very global scale (e.g., Fleming 2001). They classify inventions as products of explorative or exploitative search based on how often their inputs components have been previously used in the context of all patents ever grated by the U.S. Patent and Trademark Office (USPTO). Their analysis makes no reference to whether those components were used by the firm behind the focal patent or not. As a result, those studies cannot distinguish between search that is exploitive to one firm, but may be exploratory to another, based on its own prior knowledge. Consider the example of Google developing autonomous cars. There are many automotive technologies like brakes and suspension systems that are quite familiar to traditional auto companies like Ford, but that would represent a new terrain for Google. This means that what is exploitive to Ford might be exploratory for Google and vice versa.

Second, the other group of researchers, who also disassociate the invention from the inventor, classifies patents on a very micro scale (e.g. Jung and Lee 2016). Their method characterizes a patent based on its distance in the theoretical knowledge space from the antecedent patents referenced as prior-art on its application materials. The argument here is that since prior-art patents that the patent in question references should, by design, represent inventions to which it is most closely related, then if it is very different (i.e. very "distant") from even those referenced inventions, (which are, again, the inventions to which it is expected to be the most proximate), then it must be very new. There is, again, no reference to the innovating firm. The distance measure derived in this fashion provides valuable information about how novel a given invention might be in the context of all previously patented inventions in its domain and beyond. However, by ignoring the assignee firm, this method has no power to tell us anything about the corporate search process that lead up to the invention. This is problematic because the exploration versus exploration distinction is inherently a classification of firm activities and not of inventive outputs. The inventions themselves can be more or less novel, however, in the context of the exploration-exploitation debate, it is only the location and pattern of corporate search that led to their construction that provides ground for their classification as explorative or exploitative outputs.

Yet, a number of other scholars recognized the importance of including firms in the analysis of corporate innovation. Their work provided important insight into the phenomenon, but was often limited by data availability. Comprehensive investigation into corporate search activities necessarily requires detailed data on historical patent portfolios of a large number of firms. However, inconsistencies in names of patent assignees prevalent in the USPTO's records presented a major challenge prior to recent advances in algorithmic text disambiguation techniques (Bena, Ferreira, Matos, and Pires 2017). Because an assignee firm's name may appear in different ways at different times in the patent database (sometimes simply due to typographically errors in the USPTO system or errors in the patents themselves), a longitudinal analysis of firms' patenting activity could only be accomplished through manual construction of datasets. As a result of this, majority of prior firm-centric studies included only limited historical data and a small number of firms from a specific industry (for example, optical disks (Rosenkopf and Nerkar 2001), robotics (Katila and Ahuja 2002), or nanotechnology (Jung and Lee 2016)). Since patenting varies significantly across industries (Cockburn and Griliches 1988; Levin et al. 1987), many of those studies face generalizability constraints.

Nevertheless, they provide an important foundation for studying corporate innovation. In those studies, scholars often focused on classifying the firm itself as an explorer- or an exploiter- type. That is, rather than focusing on the search pattern behind each individual invention, they characterized the firm's entire patent portfolio on a scale from explorative to exploitative. For example, Rosenkopf and Nerkar examined all prior-art patents cited by optical disk manufacturers in a given year (2001). From it, they derived a firm-year level classification of explorative or exploitative innovation strategy and tied it to new product introductions in the subsequent periods. This approach shined meaningful light on the innovation process, but did not incorporate the consideration of organizational ambidexterity. This is a limitation, since we know that firms can simultaneously engage in exploitative search processes to develop some inventions and explorative search processes to develop others (for review O'Reilley and Tushman 2013). In fact, the large literature on organizational ambidexterity focused specifically on providing evidence that firms can parallelly carry out the two search strategies and suggested a number of ways for how they can do so (Hayward 2002; He and Wong 2004; Jansen et al. 2006; Lavie and Rosenkopf 2006; Sidhu et al. 2007; Tushman and O'Reilly 1996). Thus, a more robust way to classify the process of innovation as exploration or exploitation should focus on the corporate knowledge used to build each individual invention, rather than the aggregation of knowledge used in all of a firm's simultaneous search activities.

Finally, we know that firm knowledge evolves over time (Argote, Beckman and Epple 1990; Darr Argote, and Epple 1995; Epple, Argote, and Devadas 1991). As a result, what might be an unfamiliar, exploratory terrain at one point in time could well become a familiar, exploitive ground at another. Yet, the majority of analytical approaches pursued in prior research implicitly ignore the possible variation in firms' search strategies over time. We believe that new light could be shed on the theorization of corporate innovation by longitudinally examining the search process in its entirety, and allowing for the possibility of a firm pursuing different search strategies at the same point time and changing those strategies over time.

METHODS

We followed the long-standing method of using patent data to analyze firms' innovative activities (e.g., Benner and Tushman 2003; Fleming 2001; Katila 2002; Katila and Ahuja 2002; Rosenkopf and Nerkar 2001), the strengths and weaknesses of which are well known. To compose our dataset, we started with all patents granted by to the USPTO before March 2013 that had application dates between 1975 and 2011. Next, we filtered the dataset to include only the patents that reference at least one prior-art patent, meaning include the information necessary to diagnose their location in the theoretical knowledge space (e.g., Fleming 2001; Fitzgerald et al. 2019; Jaffe 1989; Jung and Lee 2016; Rosenkopf and Nerkar 2001). Finally, we constrained our sample to include only the patents with assignee firms that appeared in the University of Virginia's The Global Corporate Patent Dataset (Bena, Ferreira, Matos, and Pires 2017) and The Derwent World Patents Index. Our full sample included 1,639,591 patents.

Our analysis proceeded in sequential steps. First, we selected all patents with 2005 as their application date. They became the focal patents used in our analysis, simultaneously making 2005 the focal year. Then, we classified those patents into "breakthroughs" and "non-breakthroughs." To do so, we followed the long-standing convention of using forward citations received by a patent to assess its importance (Albert et al. 1991; Hall, Jaffe, and Trajtenberg 2000). Specifically, we classified as "breakthroughs" the patents that received top 5% of forward citations among patents that belong to the same application year and technological class (e.g., Jung and Lee 2016; Singh and Fleming 2010). Our sample of focal patents included 4,743 breakthroughs and 69,499 non-breakthroughs.

We focused on the patents with 2005 application date because our dataset only captures patents granted before April 2013. Since we used forward citations from patents granted after 2005 but before April 2013 to assess the focal patents' importance, 2005 was the last year for which we could observe full seven years of forward citations data. On average, there is a two to three year lag between patent application and granting, and forward citations typically plateau three years after the grant year (Jaffe et al. 1993). Consequently, examining focal patents over seven years after their application date, allowed us to capture most of the forward citations those patents are expected to receive. This increased our confidence that we have correctly distinguished patented inventions that are breakthroughs from those that are not.

Our second step was to build a complete annual patent portfolio history, from 1975-2005, for every patent assignee firm that applied for at least one patent in 2005. If a company was founded after 1975, or did not begin applying for patents until a later date, we traced back to the first time it appeared in our dataset. Because we were interested in corporate search strategies, we built the portfolios based on patent applications, rather than patent grants. A successful patent grant is indicative of how well a given invention meets prior art and other legal standards of patenting. The application tells us much more about what the firm was actually attempting to do and which technological components it was utilizing (Singh and Fleming 2010).

In our third step, we followed the established convention of examining components used to build each focal patent to determine its location in the figurative knowledge space. Using this approach, Jaffe was the first one to propose a method that allowed scholars to calculate the theoretical "distance" between two inventions based on the degree of overlap in the components used to construct them (1989). For the purpose of our analysis, we adapted Jaffe's original method to estimate the proximity between each focal patent and its assignee firm's full patent portfolio in each preceding year. To do this, we developed a novel measure, *Technological Focal Proximity.*

To compose this measure, we drew on the established practice of using technological classes of prior-art patents referenced by a focal patent to operationalize the components used to construct it (Jaffe 1989; Jung and Lee 2016; Rosenkopf and Nerkar 2001). Following Jaffe, we assessed the intensity of a firm's components utilization using counts of prior-art citations in different USPTO three-digit technology classes (1989). Then, we used annual values of *Technological Focal Proximity* to longitudinally characterize corporate search activities culminating in the development of each focal patent. *Technological Focal Proximity* is calculated as follows:

$$\textit{Technological Focal Proximity}_{t\text{-}i} = \frac{\sum_{c=1}^{C} f_c a_{c,t-i}}{\left(\sum_{c=1}^{C} f_c^2\right)^{1/2} \left(\sum_{c=1}^{C} a_{c,t-i}^2\right)^{1/2}}$$

where, i, indicates the distance in years from t, the focal patent's application date. f_c is the fraction of citations made by the focal patent that are in technology class c such that the vector $f = (f_1, \ldots, f_C)$ locates the patent in a C-dimensional technology space. $a_{c,\,t-i}$ is the fraction of all citations made by its assignee firm's patents, which have t-i as their application year, that are in technology class c such that the vector $a_{t-i} = (a_{1,\,t-i}, \ldots, a_{C,\,t-i})$ locates the firm's portfolio of citations in the C-dimensional technology space. The denominator corrects the measure by controlling for the number of classes cited by the focal patent and the assignee firm. The *Technological Focal Proximity* in year *t-i* will equal one when the technology class distribution of citations made by the assignee firm's patents, which have *t-i* as their application date, is identical to the technology class distribution of the citations made by the focal patent. This indicates that in the context of the theoretical knowledge space, the invention is very close to components previously used by the assignee firm. The *Technological Focal Proximity* in year *t-i* will equal zero when the technology class distribution of citations made by the assignee firm's patents, which have *t-i* as their application date, does not overlap at all with the technology class distribution of the citations made by the focal patent. This indicates that in the context of the theoretical knowledge space, the invention is very far from components previously used by the assignee firm.

We developed this measure to address limitations of methods used in the existing literature. First, in considering the assignee firm, *Technological Focal Proximity* is an improvement from methods that study outputs of corporate innovation, but abstract from the innovating firms. Second, in focusing on the search process underlying each specific patent, *Technological Focal Proximity* is consistent with the extensive body of evidence that firms can simultaneously engage in exploitative search processes to develop some inventions and explorative search processes to develop others (for review O'Reilly and Tushman 2013). Third, in characterizing firms' activities on a yearly basis, *Technological Focal Proximity* allows us to analyze the undertaken search process longitudinally.

In our main analysis (Table 4), we use *Technological Focal Proximity_5average*. It represents the average value of *Technological Focal Proximity* across the five year period immediately preceding a focal patent's application date. We choose to focus on this period, because prior research had established that a firm's component usage in the five years leading up to an invention reveals the most information about its relevant technological knowledge[2] (Fleming 2001).

Further, we used spline regression to diagnose the shape of corporate search trajectory across time. Spline modeling is a non-parametric regression technique that divides a dataset into a number of segments at selected values of the explanatory variable, called knots, and estimates a separate piecewise function for each of the constructed intervals. In our analysis, we devised the spline model by regressing annual values of each focal patent's *Technological Focal Proximity* onto a continuous variable representing the number of years until its application date. Following prior literature (Fleming 2001), we set the knot at five years prior to the application date.

Then, we used STATA functionality to calculate marginal splines. In this procedure, STATA calculated two slope coefficients. First, STATA reported the slope of a trend line fitted across an interval chosen to be "the base." As our "base" interval, we set the period between one and five years preceding the focal patent's application date. Then, for the second slope coefficient, the software reported the difference in slopes between the trend line fitted across the "base" interval and the remaining interval. This means that the second reported coefficient was the marginal return to the values of the explanatory variable beyond the "base" interval's boundary, not the full return to them. In our analysis, *Trend Change_5* is that second reported coefficient.

Thus, *Trend Change_5* is the difference in *Technological Focal Proximity*'s annual rate of change between the period of one to five years before the focal patent's application date and the period between five and thirty years[3] before it. Since the size of *Trend Change_5* represents the magnitude of the regime shift between the two intervals, it gives insight into possible changes in the corporate search path

[2] Results presented in Table 4 were equally significant and directionally suggestive when we used *Technological Focal Proximity* in the year immediately preceding the focal patent's application year (i.e. when *t-i* equaled *t-1*). The results were also equally significant and directionally suggestive when we used the average value of *Technological Focal Proximity* across three, not five, years preceding the focal patent's application year (i.e. when *t-i* equaled *t-1, t-2, and t-3*).

[3] If a patent's assignee company was founded after 1975, or did not begin applying for patents until a later date, the second interval is the period between five years before the focal patent's application date and the earliest year in which the assignee company appeared in our dataset.

followed to develop the focal invention. For example, when the *Trend Change_5* variable equals zero, this means that the annual rate of change in the theoretical distance between an assignee firm's knowledge base and the components used to construct the patent in question did not vary significantly between the five years preceding the patent's application date and earlier periods. Such trajectory would be indicative of a longitudinally uniform search strategy behind the focal invention. However, if *Trend Change_5* is significantly different from zero, this provides evidence that the inventor firm changed its search trajectory between the five years immediately preceding the invention and the earlier period.

Lastly, we included in our analysis a number of control variables suggested by prior research (e.g. Fleming 2001). *Claims* indicates the number of claims made by the focal patent and thus accounts for the scope of that patent. *Prior Art* indicates the number of prior-art patents referenced by the focal patent. *Class Focal* indicates the number of technological classes to which the focal patent belongs and thus serves as a secondary control for the scope of that patent. Further, as described above, the scaling factor included in the dominator of the annual *Technological Focal Proximity* is constructed to control for the number of classes cited by the focal patent (i.e. the number of theoretical components used to build the focal patent) as well as the number of classes cited by the assignee firm in the relevant comparison year (i.e. the number of components used to build all patents for which the firm applied in that year). Finally, in all of our analysis, we have included assignee company fixed effects and/or report robust standard errors that are clustered on the assignee firm to account for the possibility that error terms might be correlated for inventions that have been developed by the same company. We report summary statistics of

all variables in Table 1. We present the matrix of correlations among these variables in Table 2.

Table 1. Summary Statistics

(obs=74,242)	Mean	Standard Deviation	Minimum	Maximum
Trend Change_5	0.00	0.06	-0.98	0.78
Technological Focal Proximity_5average	0.39	0.29	0	1
Prior Art	12.59	27.52	1	770
Claims	17.44	10.91	1	219
Class Focal	1.67	0.88	1	10

Table 2. Correlation Matrix

(obs=74,242)	*Trend Change_5*	*Technological Focal Proximity_5average*	*Prior Art*	*Claims*	*Class Focal*
Trend Change_5	1.00				
Technological Focal Proximity_5average	-0.14	1.00			
Prior Art	-0.01	0.26	1.00		
Claims	-0.01	0.12	0.13	1.00	
Class Focal	-0.01	0.02	0.01	0.01	1.00

FINDINGS

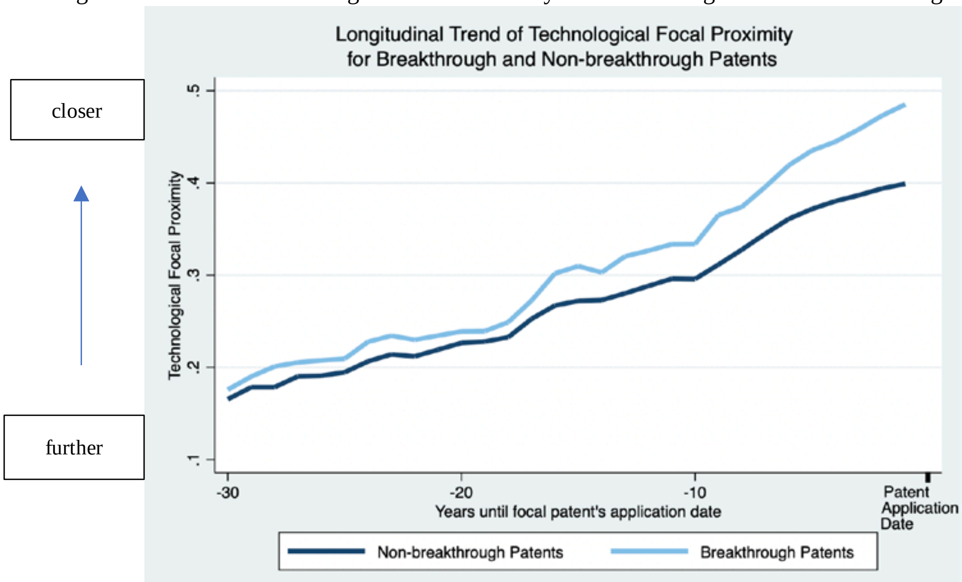
We begin by describing results of our longitudinal comparison between corporate search processes that produced breakthrough innovations and those that did not. Then, we move on to testing our proposition that both the shape of the pursued search trajectory and the intensity of the search activities influence the likelihood that the constructed invention will be a breakthrough.

Longitudinal analysis of corporate search

Figure 1 displays the distinct paths of corporate search processes culminating in development of breakthrough and non-breakthrough inventions. The graph traces mean annual values of *Technological Focal Proximity* for breakthrough and non-breakthrough patents during the thirty years before their application date. The non-overlapping trend lines provide initial evidence that the corporate search process that culminates in a breakthrough invention is different from the search process that does not produce a breakthrough.

First, annual values of *Technological Focal Proximity* are greater for breakthrough patents. This indicates that compared to an average non-breakthrough invention, a breakthrough invention is built with input components closer in the figurative knowledge space to components previously used by the assignee firm. Statistical tests presented in the Appendix provide further evidence that at five, four, three, two, and one years before the application date, the average annual value of *Technological Focal Proximity* is significantly greater for breakthrough patents than non-breakthrough patents at 95% confidence level.

Figure 1. Longitudinal Trend of Technological Focal Proximity for Breakthrough and Non-Breakthrough Patents



Second, examination of the trend lines' shape provides initial support for the proposition that the corporate search process is evolutionary and proceeds in phases. Figure 1 suggests that the search process starts with a period of exploration, throughout which annual values of *Technological Focal Proximity* are below 0.2, the commonly used threshold for differentiating explorative and exploitative activities (Benner and Tushman 2003; Fitzgerald et al. 2019). However, the period of exploration is followed by years of increasingly focused search around components used to create the focal inventions. Statistical analysis, presented in Table 3, reveals that the annual proximity of search increases at a significantly greater rate when leading up to breakthroughs than non-breakthroughs.

Specifically, to test for a difference in slopes of *Technological Focal Proximity* trendlines between the breakthrough and non-breakthrough patents, we used a spline regression model with knots at 25, 20, 15, 10, and 5 years before the patents' application date. We chose those specific periods following prior literature on organizational knowledge loss, which suggests that a firm's component knowledge changes at five-year intervals (Fleming 2001).

Results in Table 3[4], provide evidence that at ten and five years before the application date, firms increase their search proximity to components used to create breakthrough patents at a significantly greater rate than components used in non-breakthrough patents. This trend provides support for our proposition that corporate search trajectory marked with distinct stages, at which firms increasingly focus their search efforts, is positively associated with development of breakthrough inventions.

Table 3. Comparison of Technological Focal Proximity Trendlines between Breakthrough and Non-Breakthrough Patents

Dependent variable	*Technological Focal Proximity*
Regression model	*Spline*
Breakthrough	0.047** (0.008)
Breakthrough x Interval between 1 and 5 years until patent application date	-0.005** (0.001)
Breakthrough x Interval between 5 and 10 years until patent application date	-0.003** (0.001)
Breakthrough x Interval between 10 and 15 years until patent application date	0.001 (0.001)
Breakthrough x Interval between 15 and 20 years until patent application date	-0.004^ (0.003)
Breakthrough x Interval between 20 and 25 years until patent application date	0.002^ (0.001)
Breakthrough x Interval between 25 and 30 years until patent application date	-0.000 (0.002)
Interval between 1 and 5 years until patent application date	-0.006** (0.001)
Interval between 5 and 10 years until patent application date	-0.016** (0.002)
Interval between 10 and 15 years until patent application date	-0.004** (0.001)
Interval between 15 and 20 years until patent application date	-0.009** (0.002)

[4] Values of our variable representing the time interval before each focal patent's application date range from 1, indicating the time period immediately preceding the application date, to 30, indicating the time period thirty years before the application date. An increase in the values of this time variable means moving further away in time from the focal patent's application date. Due to this, the significant, negative coefficients in Table 3, indicate that over time search activities of assignee firms *increase*, rather than decrease, in proximity to the components used to construct the focal inventions.

Interval between 20 and 25 years until patent application date	-0.002 (0.002)
Interval between 25 and 30 years until patent application date	-0.006** (0.002)
Assignee firm fixed effects	*Included*
Standard Errors	*Robust; adjusted for 2,559 clusters per Firm ID*
Observations	*1,804,341*
R^2	*0.305*

^ p < 0.1 * p < 0.05 ** p < 0.01

Regression analysis of corporate search activity in the context of breakthrough development

The analyses reported in models (2), (3) and (4) in Table 4 provide evidence that corporate search activity that is concentrated in an area of the figurative knowledge space that is near the focal invention's input components has a higher likelihood of resulting in a breakthrough than a less focused search. Specifically, the significant, positive coefficients of *Technological Focal Proximity_5average* indicates that high search proximity in the five years preceding the patent's application date increase its likelihood of becoming a breakthrough.

The analyses reported in models (2), (3), and (5) in Table 4 provide support for our proposition that the shape of the corporate search path also effects the likelihood that the constructed invention will be a breakthrough. Specifically, the significant, positive coefficient of *Trend_Change_5* indicates that a corporate search trajectory that includes a significant increase in the growth rate of the *Technological Focal Proximity* measure in the five years leading up to a patent's application date is more likely to result in a breakthrough than a search trajectory not marked by such difference in growth rates. This finding supports our theory of evolutionary nature of breakthrough development. It provides evidence that corporate search trajectory characterized by a sharp increase in concentration of activity in the neighborhood of components used to construct the focal invention, at five years before the invention's application date, significantly increases its likelihood of becoming a breakthrough.

Further, model (3), which includes both *Trend_Change_5* and *Technological Focal Proximity_5average* variables, increases the explanatory power of model (1), which includes only control variables, as well as of models (4) and (5), which each include only one of the explanatory variables. The significant positive effects of *Technological Focal Proximity_5average* and *Trend_Change_5* on a patent's likelihood of being a breakthrough in model (3) suggests that both the magnitude of the firm's search proximity and the shape of its search trajectory are important predictors of the likelihood that the developed invention will be a breakthrough. To test robustness, model (2) estimates substantive variables only. Coefficients, magnitudes, and significance levels of *Technological Focal Proximity_5average* and *Trend_Change_5* do not vary across the models.

Table 4 Logistic Models of Corporate Breakthrough Development

Model	(1)	(2)	(3)	(4)	(5)
Dependent variable	*Breakthrough*	*Breakthrough*	*Breakthrough*	*Breakthrough*	*Breakthrough*
Regression model	*Logistic*	*Logistic*	*Logistic*	*Logistic*	*Logistic*
Technological Focal Proximity_5average		0.947** (0.113)	0.615** (0.108)	0.566** (0.105)	
Trend Change_5		1.571** (0.462)	1.410** (0.430)		1.037** (0.461)
Prior Art	0.008** (0.001)		0.007** (0.001)	0.007** (0.001)	0.008** (0.001)
Claims	0.017** (0.002)		0.015** (0.001)	0.015** (0.001)	0.017** (0.002)
Class Focal	0.115** (0.020)		0.011** (0.019)	0.113** (0.019)	0.115** (0.019)
Constant	-3.312** (0.069)	-3.080** (0.060)	-3.523** (0.076)	-3.508** (0.075)	-3.311** (0.069)
Standard Errors	*Robust; adjusted for 2,559 clusters per Firm ID*	*Robust; adjusted for 2,559 clusters per Firm ID*	*Robust; adjusted for 2,559 clusters per Firm ID*	*Robust; adjusted for 2,559 clusters per Firm ID*	*Robust; adjusted for 2,559 clusters per Firm ID*
Observations	*74,242*	*74,242*	*74,242*	*74,242*	*74,242*
Wald Chi2	*250.47*	*81.6*	*311.44*	*298.62*	*264.65*
Log likelihood	*-17245.05*	*-17459.716*	*-17172.444*	*-17189.386*	*-17236.297*
R^2	*0.0221*	*0.0099*	*0.0262*	*0.0252*	*0.0226*

^ $p < 0.1$ * $p < 0.05$ ** $p < 0.01$

CONCLUSION

Our longitudinal analysis suggests that while corporate innovation is an evolutionary process, it unfolds in distinct ways for breakthrough innovations, and this process is differently than previously described in the literature. In the initial phases of the breakthrough process, firms explore unfamiliar territory. This initially unfamiliar territory becomes a focal point for subsequent search and exploitation. Over time, through a process of cumulative search, a once "unfamiliar" territory (discovered through exploration) becomes familiar. Breakthroughs ultimately emerge from the exploitation of this now-familiar body of knowledge.

Our findings have a number of potential implications for both future research and practice. They suggest that breakthrough innovation requires an organizational capability for both exploration and exploitation. Note, our findings differ from those O'Reilly and Tushman's concept of "ambidexterity". In O'Reilly and Tushman's framework, firms need to have exploration capabilities for breakthrough innovation and exploitation capabilities to pursue routine innovations. Our findings suggest that such ambidexterity is also critical just for breakthrough innovations alone. A major practical challenge—as highlighted in literature as far back as March's seminal work —is that exploration and exploitation require fundamentally different organizational structures, processes, and cultures (1991). Our results should make one suspicious of the usual advice to put exploration and exploitation related innovative efforts in different organizational units. To the extent breakthroughs require both, then understanding how these seemingly contradictory capabilities can be integrated looks to be an organizational challenge well worth understanding in future research. Our work also contributes methodologically to strategy and innovation research by developing a novel measure of corporate search process, *Technological Focal Proximity*. This construct expands the scope of theoretical questions that could be addressed by future empirical research as it will allow scholars to analyze the dynamics of knowledge accumulation and search strategies at the firm level and over time.

Nevertheless, our research is not without limitations. Patents, the basic inputs of our analysis, have well known limits as marker of firms' innovative activity. For example, important know-how inside firms is often not captured in patents and industries vary in their propensity to patent. But despite their limits, patents enabled us to explore in a very comprehensive and detailed way longitudinal innovation histories of a broad cross-section of firms. Second, our analysis was limited to characterizing the process behind breakthrough innovations. We did not probe the important question why some firms might be more capable than others at doing the type of exploration-exploitation search that successfully underpins breakthroughs. While our modeling methodology took into account firm fixed effects, we did not explicitly examine the multitude of potentially important firm-specific factors (including strategy, structure, management practices, and culture) that might contribute to one firm having greater likelihood of achieving breakthrough innovation than another. Yet, to conclude, given the importance of breakthrough innovation for economic growth and individual firm performance, we hope our analysis and methods will pave the way for future work on this topic.

REFERENCES

Ahuja, G., and Lampert, C. M. 2001. Entrepreneurship in the large corporation: A longitudinal study of how established firms create breakthrough inventions. *Strategic Management Journal*, 22: 521–543.

Albert, M. B., Narin, F., Avery, D., and McAllister, P. 1991. Direct validation of citation counts as indicators of industrially important patents. *Research Policy,* 20: 251–259.

Argote, L., Beckman, S., and Epple, D. 1990. The persistence and transfer of learning in industrial settings. *Management Science*, 36: 140–154.

Baradeen, J. 1956. Semiconductor Research Leading to the Point Contact Transistor. *Nobel Lecture*

Bena, J., Ferreira, M., Matos, P., and Pires, P. 2017. Are foreign investors locusts? The long-term effects of foreign institutional ownership. *Journal of Financial Economics* 126: 122-146.

Benner, M., and Tushman, M. 2003. Exploitation, Exploration, and Process Management: The Productivity Dilemma Revisited. *The Academy of Management Review,* 28: 238-256.

Bower, J., and Christensen., C. 1995. Disruptive Technologies: Catching the Wave. *Harvard Business Review* 73: 43–53.

Burgelman, R. 1983. A process model of internal corporate venturing in a major diversified firm. *Administrative Science Quarterly,* 28: 223–244.

Cockburn, I., and Griliches, Z. 1988. Industry Effects and Appropriability Measures in the Stock Market's Valuation of R&D and Patents. *American Economic Review,* 78: 419-423.

Cohen, W. M., and Levinthal, D. 1990. Absorptive capacity: A new perspective on learning and invention. *Administrative Science Quarterly,* 35: 128–152.

Collins, M. 2018. A Telephone for the World: Iridium, Motorola, and the Making of a Global Age. Johns Hopkins University Press, Baltimore, MD.

Darr, E., Argote, L., Epple, D. 1995. The acquisition, transfer, and depreciation of knowledge in service organizations: Productivity in franchises. *Management Science*. 41: 1750–1762.

Derwent World Patents Index (accessed in May 2020)

Epple, D., Argote, L., and Devadas, R. 1991. Organizational learning curves: A method for investigating intra-plant transfer of knowledge acquired through learning by doing. *Organization Science,* 2: 58–70.

Fitzgerald, T., Balsmeier, B., Fleming, L., and Manso, G. 2019. Innovation Search Strategy and Predictable Returns. *Management Science,* Forthcoming.

Fleming, L. and Sorenson, O. 2001. Technology as a complex adaptive system: evidence from patent data. *Research Policy,* 30: 1019-1039.

Fleming, L. 2001. Recombinant uncertainty in technological search. *Management Science*, 47: 117–132.

Gertner, J. 2012. The Idea Factory: Bell Labs and the Great Age of American Innovation. Penguin Random House, New York, NY.

Hall, B., A. Jaffe, M. Trajtenberg. 2000. Market value and patent citations: A first look. NBER Working paper 7741

Hayward M. 2002. When do firms learn from their acquisition experience? Evidence from 1990–1995. *Strategic Management Journal,* 23: 21–39.

He Z. L, and Wong P. K. 2004. Exploration vs. exploitation: an empirical test of the ambidexterity hypothesis. *Organization Science,* 15: 481–494.

Henderson, R. M., and Clark, K. 1990. Architectural invention: The reconfiguration of existing product technologies and the failure of established firms. *Administrative Science Quarterly*, 35: 9–30.

Henderson, R. M., and Cockburn, I. 1994. Measuring competence? Exploring firm effects in pharmaceutical research. *Strategic Management Journal*, 15: 63–84.

Jackson, T. 1998. Inside Intel: Andy Grove and the Rise of the World's Most Powerful Chip Company. Diane Publishing Company, Darby, PA.

Jaffe, A. 1989. Characterizing the "technological position" of firms, with application to quantifying technological opportunity and research spillovers. *Research Policy* 18: 87–97.

Jaffe, A., M. Trajtenberg, R. Henderson. 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics*. 108 577–598.

Jansen J, Van den Bosch F, and Volberda H. 2006. Exploratory innovation, exploitative innovation, and performance: effects of organizational antecedents and environmental moderators. *Management Science* 52: 1661–1674.

Jung, H., and Lee, J., 2016.The Quest for Originality: A New Typology of Knowledge Search and Breakthrough Inventions. *Academy of Management Journal,* 59: 1725–1753.

Katila, R. 2002. Newproduct search over time: Past ideas in their prime? *Academy of Management Journal*, 45: 995–1010.

Katila, R., and Ahuja, G. 2002. Something old, something new: A longitudinal study of search behavior and new product introduction. *Academy of Management Journal*, 45: 1183–1194.

Kawato, H, 2002. The History of Liquid-Crystal Displays. *Proceedings of the IEEE,* 90: 460-500.

Kogut, B., and Zander, U. 1992. Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science*, 3: 383–397.

Lavie D, and Rosenkopf L. 2006. Balancing exploration and exploitation in alliance formation. *Academy of Management Journal*. 49: 797–818.

Leonard-Barton D. 1995. Wellsprings of Knowledge: Building and Sustaining the Sources of Innovation. Harvard Business School Press: Boston, MA

Levin, R., A. Klevorick, R. Nelson, S. Winter. 1987. Appropriating the returns from industrial research and development: Comments and discussion. Brookings Paper Economic Activity, 3: 783–831.

Levitt, B, and March, J. 1988. Organizational learning. *Annual Review of Sociology*, 14: 319–340.

Luo, H., Pisano, G., and Yu, H. Institutionalized Entrepreneurship: Flagship Pioneering. *Harvard Business School Case* 718-484, January 2018. (Revised April 2018.)

March, J. 1991. Exploration and exploitation in organizational learning. *Organization Science*, 2: 71–87.

March, J., and Simon, H. 1958. Organizations. Blackwell. Cambridge, MA.

Nelson, R., and Winter, S. G. 1982.Anevolutionary theory of economic change. Cambridge, MA: Harvard University Press.

O'Reilly, C., and Tushman, M. 2013. Organizational Ambidexterity: Past, Present, And Future. *Academy of Management Perspectives,* 27: 324-338.

Pisano, G. 2006. *Science Business: The Promise, The Reality, and The Future of Biotech*. Boston: Harvard Business Review Press.

Pisano, G., and Shulman, J. 2018. Flying into the Future: HondaJet. *Harvard Business School Case* 618-012, January 2018.

Riordan, M., Lillian H., and Conyers, H. 1999. The Invention of the Transistor. *Reviews of Modern Physics*, 71:336-345.

Rosenkopf, L., and Nerkar, A. 2001. Beyond local search: Boundary-spanning, exploration, and impact in the optical disc industry. *Strategic Management Journal*, 22: 287–306.

Schumpeter, J.1939. Business cycles. NewYork:McGraw-Hill.

Sidhu J., Commandeur H. R, and Volberda H. 2007. The multifaceted nature of exploration and exploitation: value of supply, demand, and spatial search for innovation. *Organization Science* 18: 20–38.

Singh, J., and Fleming, L. 2010. Lone inventors as sources of breakthroughs: Myth or reality? Management Science, 56: 41–56.

Tushman, M., and O'Reilly C. 1996. Ambidextrous organizations: managing evolutionary and revolutionary change. *California Management Review* 38: 8–30.

Vincenti, W. 1990. What Engineers Know and How Do They Know It: Analytical Studies from Aeronautical History. The John Hopkins University Press. Baltimore, MD.

APPENDIX

CONT. Comparison of Technological Focal Proximity between Breakthrough and Non-Breakthrough patents for each of 30 years preceding focal patents' application date

Model	(6)	(7)	(8)	(9)	(10)
Dependent variable	*Technological Focal Proximity$_{t-6}$*	*Technological Focal Proximity$_{t-7}$*	*Technological Focal Proximity$_{t-8}$*	*Technological Focal Proximity$_{t-9}$*	*Technological Focal Proximity$_{t-10}$*
Regression model	*OLS*	*OLS*	*OLS*	*OLS*	*OLS*
Breakthrough	0.012^	0.008	0.010^	0.013*	0.007
	(0.007)	(0.007)	(0.006)	(0.006)	(0.005)
Constant	0.077**	0.077**	0.213**	0.464**	0.464**
	(1.82e-11)	(7.73e-12)	(2.93e-11)	(1.00e-11)	(2.92e-12)
Assignee Company fixed effects	*Included*	*Included*	*Included*	*Included*	*Included*

Standard Errors	*Robust; adjusted for clusters per 2,142 Firm ID*	*Robust; adjusted for clusters per 2,045 Firm ID*	*Robust; adjusted for clusters per 1,918 Firm ID*	*Robust; adjusted for clusters per 1,813 Firm ID*	*Robust; adjusted for clusters per 1,713 Firm ID*
Observations	*72,711*	*71,842*	*70,899*	*69,793*	*68,677*
R^2	*0.354*	*0.354*	*0.362*	*0.364*	*0.392*

^ $p < 0.1$ * $p < 0.05$ ** $p < 0.01$

Comparison of Technological Focal Proximity between Breakthrough and Non-Breakthrough patents for each of 30 years preceding focal patents' application date

Comparison of Technological Focal Proximity between Breakthrough and Non-Breakthrough patents for each of 30 years preceding focal patents' application date

Model	(1)	(2)	(3)	(4)	(5)
Dependent variable	*Technological Focal Proximity$_{t-1}$*	*Technological Focal Proximity$_{t-2}$*	*Technological Focal Proximity$_{t-3}$*	*Technological Focal Proximity$_{t-4}$*	*Technological Focal Proximity$_{t-5}$*
Regression model	*OLS*	*OLS*	*OLS*	*OLS*	*OLS*
Breakthrough	0.030** (0.007)	0.024** (0.007)	0.018** (0.007)	0.015* (0.007)	0.014* (0.007)
Constant	0.067** (1.33e-11)	0.088** (3.49e-11)	0.088** (2.25e-11)	0.077** (1.79e-11)	0.077** (4.82e-12)
Assignee Company fixed effects	*Included*	*Included*	*Included*	*Included*	*Included*
Standard Errors	*Robust; adjusted for clusters per 2,559 Firm ID*	*Robust; adjusted for clusters per 2,503 Firm ID*	*Robust; adjusted for clusters per 2,436 Firm ID*	*Robust; adjusted for clusters per 2,352 Firm ID*	*Robust; adjusted for clusters per 2,256 Firm ID*
Observations	*74,242*	*74,101*	*73,906*	*73,627*	*73,324*
R^2	*0.349*	*0.348*	*0.351*	*0.357*	*0.360*

^ $p < 0.1$ * $p < 0.05$ ** $p < 0.01$

CONT. Comparison of Technological Focal Proximity between Breakthrough and Non-Breakthrough patents for each of 30 years preceding focal patents' application date

Model	(11)	(12)	(13)	(14)	(15)
Dependent variable	*Technological Focal Proximity$_{t-11}$*	*Technological Focal Proximity$_{t-12}$*	*Technological Focal Proximity$_{t-13}$*	*Technological Focal Proximity$_{t-14}$*	*Technological Focal Proximity$_{t-15}$*
Regression model	*OLS*	*OLS*	*OLS*	*OLS*	*OLS*
Breakthrough	0.008 (0.006)	0.011 (0.006)	0.010 (0.006)	0.007 (0.006)	0.012^ (0.006)
Constant	0.019** (8.39e-12)	0.011** (1.67e-11)	0.011 (7.02e-12)	2.91e-11 (2.94e-11)	1.05e-11 (1.06e-11)
Assignee Company fixed effects	*Included*	*Included*	*Included*	*Included*	*Included*
Standard Errors	*Robust; adjusted for clusters per 1,597 Firm ID*	*Robust; adjusted for clusters per 1,521Firm ID*	*Robust; adjusted for clusters per 1,450 Firm ID*	*Robust; adjusted for clusters per 1,388 Firm ID*	*Robust; adjusted for clusters per 1,328 Firm ID*
Observations	*66,015*	*65,593*	*64,519*	*63,727*	*62,679*
R^2	*0.365*	*0.372*	*0.369*	*0.352*	*0.342*

*^ $p < 0.1$ * $p < 0.05$ ** $p < 0.01$*

CONT. Comparison of Technological Focal Proximity between Breakthrough and Non-Breakthrough patents for each of 30 years preceding focal patents' application date

Model	(16)	(17)	(18)	(19)	(20)
Dependent variable	*Technological Focal Proximity$_{t-16}$*	*Technological Focal Proximity$_{t-17}$*	*Technological Focal Proximity$_{t-18}$*	*Technological Focal Proximity$_{t-19}$*	*Technological Focal Proximity$_{t-20}$*
Regression model	*OLS*	*OLS*	*OLS*	*OLS*	*OLS*
Breakthrough	0.007 (0.006)	0.003 (0.005)	-0.003 (0.005)	-0.005 (0.005)	-0.001 (0.005)
Constant	-6.04e-12 (6.10e-12)	1.51e-11 (1.52e-11)	-1.42e-11 (1.44e-11)	-9.19e-13 (9.28e-13)	1.28e-11 (1.29e-11)
Assignee Company fixed effects	*Included*	*Included*	*Included*	*Included*	*Included*
Standard Errors	*Robust; adjusted for clusters per 1,247 Firm ID*	*Robust; adjusted for clusters per 1,184 Firm ID*	*Robust; adjusted for clusters per 1,113 Firm ID*	*Robust; adjusted for clusters per 1,039 Firm ID*	*Robust; adjusted for clusters per 978 Firm ID*
Observations	*61,417*	*60,579*	*58,672*	*57,776*	*56,473*
R^2	*0.322*	*0.325*	*0.302*	*0.295*	*0.312*

*^ $p < 0.1$ * $p < 0.05$ ** $p < 0.01$*

CONT. Comparison of Technological Focal Proximity between Breakthrough and Non-Breakthrough patents for each of 30 years preceding focal patents' application date

Model	(21)	(22)	(23)	(24)	(25)
Dependent variable	*Technological Focal Proximity$_{t-21}$*	*Technological Focal Proximity$_{t-22}$*	*Technological Focal Proximity$_{t-23}$*	*Technological Focal Proximity$_{t-24}$*	*Technological Focal Proximity$_{t-25}$*
Regression model	*OLS*	*OLS*	*OLS*	*OLS*	*OLS*
Breakthrough	-0.003 (0.006)	-0.002 (0.007)	-0.001 (0.007)	-0.000 (0.007)	-0.001 (0.007)
Constant	-1.21e-12 (1.22e-12)	-4.29e-12 (4.33e-12)	-1.06e-11 (1.07e-11)	-1.30e-11 (1.31e-11)	1.28e-11 (1.29e-11)
Assignee Company fixed effects	*Included*	*Included*	*Included*	*Included*	*Included*
Standard Errors	*Robust; adjusted for clusters per 915 Firm ID*	*Robust; adjusted for clusters per 853 Firm ID*	*Robust; adjusted for clusters per 809 Firm ID*	*Robust; adjusted for clusters per 763 Firm ID*	*Robust; adjusted for clusters per 719 Firm ID*
Observations	*52,132*	*50,252*	*50,007*	*48,516*	*47,665*
R^2	*0.266*	*0.267*	*0.246*	*0.237*	*0.223*

*^ $p < 0.1$ * $p < 0.05$ ** $p < 0.01$*

CONT. Comparison of Technological Focal Proximity between Breakthrough and Non-Breakthrough patents for each of 30 years preceding focal patents' application date

Model	(26)	(27)	(28)	(29)	(30)
Dependent variable	*Technological Focal Proximity$_{t-26}$*	*Technological Focal Proximity$_{t-27}$*	*Technological Focal Proximity$_{t-28}$*	*Technological Focal Proximity$_{t-29}$*	*Technological Focal Proximity$_{t-30}$*
Regression model	*OLS*	*OLS*	*OLS*	*OLS*	*OLS*
Breakthrough	-0.002 (0.007)	-0.001 (0.007)	0.005 (0.007)	-0.000 (.006)	0.001 (0.007)
Constant	6.12e-12 (6.17e-12)	-9.84e-12 (9.91e-12)	-6.99e-12 (7.04e-12)	8.22e-12 (8.28e-12)	0.091** (0.001)
Assignee Company fixed effects	*Included*	*Included*	*Included*	*Included*	*Included*
Standard Errors	*Robust; adjusted for clusters per 665 Firm ID*	*Robust; adjusted for clusters per 614 Firm ID*	*Robust; adjusted for 545 clusters per Firm ID*	*Robust; adjusted for 457 clusters per Firm ID*	*Robust; adjusted for 349 clusters per Firm ID*
Observations	*46,572*	*45,368*	*42,496*	*41,236*	*39,525*
R^2	*0.239*	*0.229*	*0.223*	*0.237*	*0.227*

*^ $p < 0.1$ * $p < 0.05$ ** $p < 0.01$*