



Paper to be presented at the  
35th DRUID Celebration Conference 2013, Barcelona, Spain, June 17-19

## **The Technological Origins and Novelty of Breakthrough Inventions**

**Sam Arts**

KU Leuven

Faculty of Business and Economics  
sam.arts@kuleuven.be

**Reinhilde Veugelers**

KU Leuven

Faculty of Business and Economics  
reinhilde.veugelers@kuleuven.be

### **Abstract**

We explore the relationship between the technological origins and novelty of inventions on the one hand and their technological impact on the other hand. In particular, we are interested into the technological origins and novelty of breakthrough inventions. By jointly looking at the effects of the origins and novelty on an invention's average impact, on the likelihood of a very poor invention, and on the likelihood of a breakthrough, we identify some trade-offs researchers face when exploring breakthroughs. For evidence, we consider the US patent record in biotechnology from 1976 to 2001. Our analysis shows that breakthroughs in biotechnology rely more on non-technical and technical prior art, particularly more recent technical prior art, prior art from many different technology fields, and prior art from unfamiliar technology fields. Yet, breakthroughs are less likely to have a dissimilar set of technical prior art citations, as they are more likely to use prior art previously cited by many other inventions. Besides differences in the origins, we find significant differences in the technological novelty. Breakthroughs are more novel in the sense that they are more likely to recombine technological components for the first time in history, particularly more familiar technological components.

## **1. Introduction**

The main objective of this paper is to study the technological origins and novelty of breakthrough inventions. The large majority of all inventions are incremental, further improving existing inventions. Only a very small minority of all inventions are breakthroughs. Although they are small in numbers compared to incremental ones, breakthroughs are generally considered as most important. They generate order-of-magnitude improvements rather than incremental improvements over existing technologies (Tushman and Anderson, 1986) and serve extensively as prior art for many subsequent incremental inventions who further improve on the breakthrough. They lie at the heart of changes in technological paradigms (Freeman, 1992), create new technological systems and even new industries (Nelson and Winter, 1982).

The breakthrough versus incremental characterization of technological inventions is most commonly traced to Schumpeter (Freeman, 1992). Schumpeter's work placed far greater emphasis on the discontinuous nature of technological change than on smaller more gradual improvements. Because of their high direct and indirect impact, he deemed technological breakthroughs as a key component of value creation and growth (Schumpeter, 1942). Subsequent work has shown how firms' value creation and performance often depends on the few breakthroughs in their portfolio. Scherer and Harhoff (2000) for example estimated for different samples of company and university patents that 10% of the patents generated 48% to 93% of the total returns. Other studies showed how companies pioneering breakthroughs grow at a much faster rate (Tushman and Anderson, 1986). But also from a societal perspective, breakthroughs are an important target. Public R&D funding agencies such as the National Science Foundation (NSF) increasingly aim at transformative research, i.e. research that holds the potential to radically change our understanding of current science or engineering concepts (NSB, 2007). Similarly, the European Research Council (ERC) mission statement includes that its grants should help to bring about new and unpredictable scientific and technological discoveries, the kind of inventions that form the basis of new industries, markets and broader social innovations in the future. In addition, Europe's Horizon 2020 states as ambition to help innovative firms to develop their technological breakthroughs.

A major issue in the academic literature studying breakthrough inventions is how they arise and evolve. Most empirical papers study the emergence of breakthroughs from a company

(e.g. Ahuja and Lampert, 2001) or inventor perspective (e.g. Singh and Fleming, 2010), looking for which type of firm or team of inventors is most likely to introduce breakthroughs. A particular focus in the literature is on whether large incumbents or rather small start-ups are the main source of breakthrough inventions. Older and larger firms, having to surmount an incumbency trap, are typically taken to be disadvantaged in introducing breakthroughs (e.g. Henderson, 1993; Sorenson and Stuart, 2000). But this literature by and large fails to come with conclusive evidence. Large firms are not necessarily doomed to incremental improvement of well-established technology and there is substantial heterogeneity across and within firms on who introduces breakthroughs and how (Ahuja and Lampert, 2001; Hill and Rothaermel, 2003). We believe part of this confusion is caused by the lack of a proper understanding of the technological origins and nature of breakthroughs. For instance, one of the main reasons typically raised for why incumbent firms fail is because breakthroughs are frequently based on scientific and technological knowledge unfamiliar for incumbents (Tushman and Anderson, 1986). We lack however empirical evidence on whether technological breakthroughs indeed more likely originate from unfamiliar knowledge and technology.

In this paper, rather than studying the associated companies or inventors, we look at different ex-ante characteristics of breakthroughs at the invention level. More specifically, we want to study the technology shoulders on which they build and the new configurations of technologies they make. Analyzing the technological origins and novelty of breakthroughs and how both interact can provide a more comprehensive picture on which company or team of inventors is most likely to introduce them. Furthermore, it potentially can inform researchers, companies and governments on different pathways to create breakthroughs, i.e. on how they could organize to increase their chances at a breakthrough. For instance, what kind of search and exploration strategies are most successful? Should companies stay in familiar territory or venture into new and unfamiliar technology fields? Should researchers search emerging technologies as locus for breakthroughs? Should they experiment with new recombinations of well-established technologies? Can R&D funding agencies stimulate the creation of breakthroughs by stimulating cross-disciplinary collaboration among researchers?

Despite the importance of breakthrough inventions, little is actually known about their technological origins and novelty, besides seminal historical analyses (e.g. Usher, 1954; Basalla,

1988), a number of more recent case studies (e.g. Fleming, 2002; Hargadon, 2003; Cattani, 2006) and scarce empirical evidence on particular characteristics of their origins. Using patent data, Fleming (2001) provides larger empirical evidence focusing on technological component and combination familiarity of inventions and their effect on the average and variance of patent value. He finds that experimentation with less familiar technological components and component combinations leads to less useful inventions on average but also increases the variability resulting in higher probability for both failure and breakthrough. Using a sample of tennis racket patents, Dahlin and Behrens (2005) find technologically radical patents have backward patent citations which are dissimilar from prior patents. Studying the technological origins of breakthrough patents, Schoenmakers and Duysters (2010) find breakthroughs, which they list as radical, use more, broader and younger technical prior art as compared to a control group of patents.

In this paper, we look at both technological origins and technological novelty of breakthroughs. We identify different trade-offs researchers face by taking more inventive risks in order to discover breakthroughs. First, by jointly looking at the effects of the origins and novelty on an invention's average value, on the likelihood of a very poor invention, and on the likelihood of a breakthrough. Second, by studying the interaction between technological origins and novelty. In line with prior empirical studies, we use patent data to identify and characterize breakthrough inventions. Patents are the single most important source of information on the origins, novelty and impact of technological inventions. They contain detailed and consistent information on both the prior art used to create the invention as well as on the nature and impact of the invention itself. Using forward patent citations to measure impact restricts the analysis to assessing breakthroughs in the technological dimension only (Jaffe et al., 2000), being ignorant on the commercial dimension. Nevertheless, previous research has indicated that forward citation counts are related to social value (Trajtenberg, 1990) and private value (Harhoff et al, 1999; Hall et al., 2005, Gambardella et al, 2008). Moreover, we restrict our analysis to the field of biotechnology since the vast majority of inventions in biotechnology is patented (Arundel and Kabla, 1998), and because biotechnology experienced many technological breakthroughs over the past years.

Our analysis on biotech patents suggests breakthroughs rely more on non-technical and technical prior art, particularly more recent technical prior art, prior art from many different technology fields, and prior art from unfamiliar technology fields. Yet, breakthroughs are less likely to have a dissimilar set of technical prior art citations, as they are more likely to cite technical prior art previously cited by many other inventions. Besides differences in the origins, we find significant differences in the technological novelty. Breakthroughs are more novel in the sense that they are more likely to recombine two technological components for the first time in history and have a larger proportion of such new to the world recombinations. Furthermore, they are broader in scope and serve a more general purpose as they cover more claims and technology fields. Testing a number of interaction effects between the origins and novelty of inventions, we find that most impactful technological inventions are based on new recombinations of most familiar technological components, i.e. components which have been used extensively in many technological applications in the past. Before we present the results in section 4, we first review the literature in section 2 and detail our data and empirical strategy in section 3.

## **2. Breakthrough Inventions and their Technological Origins and Novelty**

### **2.1. Characterizing Breakthrough Inventions**

The literature struggles with a lack of a commonly used definition of what breakthrough inventions are. This is partly due to the fact that this topic has been studied by many scholarly communities including the economics and strategy of innovation, industrial organization, evolutionary economics, corporate strategy, marketing and the history of technological change. Researchers have tended to draw upon a wide range of terminologies where beyond breakthroughs also discontinuous, radical and disruptive technologies meet.

Tushman and Anderson (1986) define technological breakthroughs or discontinuities as inventions which offer order-of-magnitude improvements in the price versus performance ratio over existing technology. They classified these major technological shifts as either competence-enhancing or competence-destroying, depending on if they reinforced or destroyed established firms' existing competencies, skills and knowledge. Rosenberg (1994) identifies major innovations as those that provide a framework for many subsequent innovations, each dependent

upon or complementary to the original one. Similarly, Fleming (2001) and Ahuja and Lampert (2001) identify breakthroughs as those foundational inventions that serve as the basis for many subsequent technological inventions.

A number of other concepts closely related to breakthrough inventions are popular in the management literature. A notion often associated with breakthroughs is radicalness. Chandy and Tellis (1998) define radical inventions as those that incorporate substantially different technology compared to existing products and can fulfill consumer needs better than existing products. They separately define technological breakthroughs as those that satisfy the first, but not necessarily the second condition. Utterback (1996) defines radical innovations or discontinuous change as “change that sweeps away much of a firm’s existing investments in technical skills and knowledge, designs, production technique, plant and equipment,” and Henderson (1993) described an innovation as radical if it renders a firm’s information filters and organizational procedures partially obsolete. Henderson and Clark (1990) introduce the idea of an architectural innovation in which core concepts or components remain unchanged but are linked together differently in a new architecture. Christensen (2003) focuses on disruptive technologies and their implications for established firms in an industry. Christensen’s theory of disruptive technologies focuses on the value a technology brings to the market and whether or not it is a dimension customers have traditionally valued. A disruptive technology will have features that initially only a fringe market segment will value. It redefines the performance trajectory. For example, shrinking the size of disks in the disk drive industry. These disruptive technologies need not be technological breakthroughs. In fact, Christensen notes that generally disruptive innovations are technologically straightforward.

In this paper, we follow Rosenberg (1994), Fleming (2001) and Ahuja and Lampert (2001) and define technological breakthroughs as those inventions that disproportionately serve as prior art for many subsequent inventions. We follow the tradition in the empirical literature by identifying technological breakthroughs by means of forward patent citations (Fleming, 2001; Ahuja and Lampert, 2001; Singh and Fleming, 2010). The count of forward citations a patent receives is correlated with its technological importance (Albert et al., 1991; Carpenter et al, 1993; Jaffe et al., 2000) as well as its social (Trajtenberg, 1990) and private value (Harhoff et al, 1999; Hall et al., 2005). The distribution of forward citation counts is very skewed with a large

share of patents receiving no citations and a small minority of patents with a very large count of forward citations. This pattern resembles the distribution of the actual impact and value of inventions which also tend to be very skewed (Schumpeter, 1942, Scherer & Harhoff, 2000). Therefore, it is most likely that outliers in the top of the distributions of forward citation counts are most important technological inventions.

Schoenmakers & Duysters (2010) simply identify breakthroughs as those patents with more than 20 forward citations, ignoring technology and time specific effects driving the number of citations. Most other prior studies have identified breakthrough patents as the top 1% or 5% in terms of citations received compared to patents filed in the same year and in the same primary 3-digit technology class (e.g. Ahuja and Lampert 2001; Singh and Fleming 2010). This definition assumes each technology field to have a fixed share of high impact inventions each year and does not compare patents across years in order to identify breakthroughs. To avoid a definition that forces a fixed share of breakthroughs every year in each class while allowing similar patents to be compared across years, we consider the distributions of both un-truncated and truncated forward citations counts and identify outliers in the top of these distributions. This methodology is explained in more detail in Arts (2012) and in the variables section below.

## 2.2 The Technological Origins of Breakthroughs

Although frequently labeled as revolutionary or discontinuous to prior art, breakthroughs generally also have a pre-history of technical developments. Grounded in historical analyses of the evolution of technology, Usher (1954) and Basalla (1988) show that breakthroughs, and technological inventions in general, originate from the recombination of knowledge and technological components already in existence. As such, technological progress is an evolutionary rather than a revolutionary process whose technological origins can be traced. For instance, the transistor, often labeled as a disruptive or radical invention replacing vacuum tubes, has its origins in crystal radio sets preceding vacuum tubes (Basalla, 1988). Also management scholars such as Utterback (1996) point out that “radical innovations often are seen to be based on the synthesis of well-known technical information or components”. Fleming (2001) argues and shows breakthroughs most likely originate from new recombinations of most familiar

technological components, i.e. components which have been used extensively in the past in different applications. Levinthal (1998) and Adner and Levinthal (2002) discuss how radical or discontinuous technologies result from a shift of an existing technical lineage to a new domain of application, from leveraging knowledge and technology from unfamiliar technology fields.

To trace the technological origins of patented inventions, scholars have used backward patent citations, citations to non-patent literature, the technology fields of the patents which are being cited and the technology subclasses or components of the patented technology itself. A first way to identify “pioneering” inventions is to look at patents without any citations to prior patents (Ahuja and Lampert 2001). Granted patents without technical prior art citations have no direct technological antecedents and potentially cover very original technologies. By contrast, the count of backward patent citations indicates on how much technical prior art the patent is directly relying. A large number of backward citations is associated with a larger base of technical prior art.

Another way to identify technologically radical inventions using backward patent citations was developed by Dahlin and Behrens (2005). They argue and show, using a sample of tennis racket patents, that patents protecting technologically radical inventions have a dissimilar set of technical prior art citations compared to prior patents filed in the same field, i.e. they recombine technical prior art which is not frequently used by prior patents and particularly not used together by one and the same patent. This definition is in line with the conceptualization of invention as a recombinant search process and the view that technological breakthroughs originate from new recombinations of existing technological components (Fleming, 2001; Hargadon, 2003).

Besides looking at the number of backward patent citations and at which patents are being cited, others have analyzed the technology fields of the cited patents. Technological breakthroughs are expected to source prior art from a broader part of the technology landscape. While Schoenmakers & Duysters (2010) simply count the number of different technology fields being cited, Trajtenberg et al. (1997) and Hall et al. (2001) developed a concentration measure based on the spread of backward citations over technology fields, arguing that the lower the concentration, the more original or basic the invention is.

Comparing the technology classes of the cited patents to the technology classes of the citing patent, allows testing whether breakthrough inventions build upon different technological knowledge and paradigms than the one in which it is currently applied (Shane, 2001; Adner and Levinthal, 2002), whether unfamiliar technology fields were sourced (Rosenkopf and Nerkar, 2001). Citing proportionally more patents from technology fields different from its own technology fields reflects the degree to which knowledge and technology is sourced from different fields and is expected to increase the likelihood at a breakthrough.

Furthermore, citations to non-technical prior art, predominantly scientific literature, may also reflect the evolutionary origins of the invention. Science alters inventors search processes because it directs researchers to the most useful combinations of technological components. Fleming and Sorenson (2004) show how the value of science for technological inventions increases with the difficulty of the inventive problem. As such, we expect technological breakthroughs being big inventive steps to be more science rooted.

With respect to the age of the cited prior art, breakthroughs are expected to rely on more recently developed technical prior art (Ahuja and Lampert, 2001; Schoenmakers & Duysters, 2010). As a technology matures, it becomes better understood and more reliable. Therefore, recombining more mature technologies reduces the recombinant uncertainty but also the likelihood to hit a breakthrough. By contrast, emerging or relatively young technologies are less familiar and reliable but provide more opportunities to be used in novel ways and therefore may generate breakthroughs more often. Schoenmakers & Duysters (2010) calculate the mean average age of the cited patents and find that breakthroughs rely on younger technical prior art on average.

To capture not only the directly cited technical prior art but also the use and recent use of the technological components or subfields of the patent, Fleming (2001) develops a measure of component familiarity capturing how many prior patents have used a patent's technological components<sup>1</sup>. On the one hand, the reuse of more familiar technological components might reduce the recombinant uncertainty and increase the average value of the invention but at the same time also reduce the likelihood at a breakthrough. On the other hand, inventors might take more inventive risks with more familiar components. Fleming (2001) shows first that

---

<sup>1</sup> Fleming (2001) argues the technology subclasses of a patent reflect the technological components of the patented technology.

breakthroughs most likely emerge from the recombination of most familiar technological components, and second that re-using exactly the same combination of technological components as prior patents reduces the likelihood at a breakthrough. The more a particular combination is used by prior patents, the less likely a breakthroughs will emerge. Therefore, his findings suggest technological breakthroughs most likely originate from recombining disconnected technological components each having a distinct but very fruitful history of progress in different inventive applications. Experimenting with new combinations of technological components can help organizations create technological breakthroughs, but at the same time it also reduces the average value of the inventions and increases the number of failures.

### 2.3 The Technological Novelty of Breakthroughs

Beyond having different technological origins compared to incremental inventions, breakthroughs are also expected to be more novel in a technological sense. Usher (1954) and Basalla (1988) describe how technological novelty is mainly driven by the recombination of pre-existing technological knowledge and components. An invention can be novel either by combining particular technological components for the first time or by differently linking previously combined components together in a new architecture (Henderson and Clark, 1990; Fleming, 2001).

As postulated by Arthur (2009), existing technologies used in new combinations provide the possibility to create novel technologies. There are many examples of important inventions involving combinations and transfers of ideas across technological domains (Mowery and Rosenberg, 1998). Hargadon (2003) provides many examples of how brokering disconnected but well-understood technological components led to the discovery of breakthroughs such as the light bulb or the reebok pump shoe.

To identify particularly novel technological inventions with a potentially high impact on future technological progress, Fleming et al. (2007) proposes to look at patents which are the first in history to recombine at least 2 previously disconnected technological components. Arts (2012) studies how technological incumbent firms can create breakthroughs by brokering very familiar technological components with disconnected technological components. As such, we

expect patents with new recombinations of technological components and with a large proportion of such new recombinations are more likely to be breakthroughs.

### **3. Data, Variables and Hypotheses**

In this paper, we look at both technological origins and technological novelty of breakthrough inventions. More particularly, we want to see whether (i) inventions with more diversified, dissimilar, younger and more familiar technical prior art and (ii) inventions with new recombinations of technological components are more likely to turn into breakthroughs. We also explore any possible effect from combining novelty with familiarity.

#### **3.1 Patent Data**

In line with the literature, we use patent data to identify breakthrough inventions and to characterize the technological origins and novelty of inventions. While not all technological inventions are patented, patent data still include a large proportion of both failed and very successful technological discoveries, making the data appropriate to study breakthroughs (e.g. Fleming, 2001; Ahuja and Lampert, 2001). Patent data provide a detailed window on when and what kind of technology inventions are created. Because inventors are legally obliged to cite all relevant prior art, tracing these citations allows us to carefully trace the evolutionary origins of patented inventions. The technology classes of a patent capture the technology fields or domains covered by the patent. Furthermore, the systematic classification of patents into very detailed technological main- and subclasses (currently more than 140,000) allows the identification of the different subfields or components of the patented technology (Fleming, 2001). Finally, forward citations can be used to measure technological impact and to identify technological breakthroughs (e.g. Ahuja and Lampert, 2001; Singh and Fleming, 2010).

#### **3.2 Sample Selection**

Our study will focus on the field of biotechnology given that patenting is a very common practice and given the significant number of breakthroughs in the field over the past years. The technological origins and novelty of breakthrough inventions might differ across technology domains so that combining patents from multiple domains into the same analysis might blur the effects. To identify all USPTO biotechnology patents, we made use of the OECD classification

scheme which relies on IPC codes (OECD, 2005). Data have been extracted from the Patstat patent database (version October 2011) and include all patents filed at the USPTO between 1976 and 2001 and granted before 2004 which fall into at least one of the IPC classes associated with biotechnology. This sample consists of 83,091 patents. For the calculation of citation related indicators, we made use of the Harvard patent database (Lai et al, 2009). For each breakthrough patent, we develop a control sample of patents sharing the same 3-digit technology classes, application year and grant year. Only control patents which have exactly the same technology classes, application year and grant year as one of the breakthrough patents are retained. Breakthrough patents are also excluded in case they do not have at least one matching control patent.

### 3.3 Dependent Variable: Technological Breakthroughs

To identify technological breakthroughs, we calculate for all granted US patents the count of forward citations as the number of patents citing the patent (citations from patents filed until 2006 inclusive) and the truncated count of forward citations as the number of citations received within 5 years after application. We use the full count of forward citations to compare all patents sharing at least one 3-digit US technology class filed within the same year and the truncated citation count to compare all patents sharing at least one 3-digit technology class irrespectively of their time of filing. For each of the distributions, we calculate the mean and standard deviation of (truncated) forward citation counts. A patent is labeled as being a technological breakthrough in case both its truncated and full count of forward citations are larger than the mean plus  $n$  times the stdev in at least one of its technology classes. So for each technology class of a patent, the patent is compared with 2 distributions: the distribution of full forward citation counts and the distribution of truncated forward citation counts (see Arts (2012) for a more elaborate discussion of the methodology). Using a 1 and 3 standard deviation rule to identify outliers in the distribution of forward citations, we respectively find 6.9% and 1.5% of biotechnology patents in our sample to be labeled as breakthrough inventions. The use of a relatively less and more restrictive definition of breakthroughs allows us to not only study differences between breakthrough patents and control patents but also to look at breakthroughs with different levels of impact.

### 3.4 Independent Variables: Technological Origins of Inventions

## Dissimilar backward patent citations

To identify which patents have a dissimilar recombination of backward patent citations with respect to prior art, we closely follow the methodology of Dahlin and Behrens (2005). We calculate the overlap scores between the backward patent citations of each patent with all other previously granted patents in the same field, and averaging these overlap scores within each year relative to the grant year of the patent.

Concretely, we calculate for each patent  $P$  granted in year  $t$  the average annual overlap scores between the backward citations of  $P$  with respectively all other patents filed in the same field (3-digit US technology class) within a time window of 5 years before the grant year of  $P$ . We extent Dahlin and Behrens (2005) by comparing a patent to all other US granted patents with at least one similar 3-digit technology class. So for each of a patent's technology classes, we follow the methodology as outlined in Dahlin and Behrens (2005). First, we label  $P$  as having dissimilar backward citations compared to prior art in case the average annual overlap score is 0 or in case the average standardized annual overlap score is smaller than or equal to the 10<sup>th</sup> percentile of all patents for each year  $t-n$  with  $n>0$  and  $n\leq 5$ . Due to truncation, not all patents have a time window of 5 years before their grant year. For patents which we can only be observed 3 or 4 years before grant, we require the patent's average overlap score to be 0 or its standardized average annual overlap score to be equal to or smaller than the 10<sup>th</sup> percentile threshold for each of the observed years before grant. Patents which cannot be observed at least 3 years before the grant year are not taken into account during the analysis. Those patents which have the lowest overlapping scores compared to prior art in the field are considered to be dissimilar in technological origins. Following this methodology, 46% of the patents in our sample are labeled as having dissimilar technological origins. We expect inventions with dissimilar technological origins are more likely to be breakthroughs.

## Backward citation lag

The average backward citation lag captures the extent to which the patent is building on emerging or rather maturing technologies. Rather than the mean age of the cited patents, as used in Schoenmakers & Duysters (2010), which ignores the age of the citing patent, we calculate the average backward citation lag as the average number of years between the application of the citing patent and the application of the cited patents. A relatively short backward citation lag

indicates the patent builds on more recently developed technical prior art. We expect these inventions are more likely to become breakthroughs.

#### Spread of backward citations across technology fields

To measure the spread of backward citations across fields, we use the measure developed by Trajtenberg et al. (1997) and Hall et al. (2001) reflecting the concentration of backward patent citations in technology classes. Their measure, which they list as “originality” is calculated as 1-bias-corrected<sup>2</sup> Herfindahl index of the technological classes (3-digit) of all cited patents. A larger originality score indicates the patent is recombining prior art from a broad base of technology fields. We expect these inventions are more likely to become breakthroughs.

#### Proportion of backward citations from different technology fields

To capture the degree to which patents build upon prior art from technology fields different from the one(s) in which it is currently applied, we calculate the proportion of cited patents which cover 3 digit technology class(es) different from the technology class(es) of the focal patent. We expect breakthrough inventions to proportionally cite more patents from different technology classes.

#### Familiarity of technological components

In line with Fleming (2001), we calculate a patent level measure of component familiarity by taking the average familiarity of each of a patent’s subclasses. Individual component familiarity of patent i’s subclass j is calculate as

$$\sum_{\substack{\text{all patents } k \\ \text{filed before} \\ \text{patent } i}} 1\{\text{patent } k \text{ uses subclass } j\} \times e^{-\left(\frac{\text{application year patent } i - \text{application year patent } k}{\text{time constant of knowledge loss}}\right)}$$

In line with Fleming (2001), we take 5 years as the time constant of knowledge loss, i.e. a yearly knowledge loss of 18% and take the square root of the average component familiarity. We expect breakthrough inventions to recombine less familiar technological components.

#### No backward citations

---

<sup>2</sup> A bias correction is necessary because not all patents have the same number of backward patent citations so that the originality score will be biased downwards for patents with few backward citations. Therefore, the originality score is multiplied by (count backward citations/ (count backward citations-1)).

We calculate a dummy for patents without any backward citations to prior patents. These patents have been labeled as pioneering as they do not built directly on any technical prior art (Ahuja and Lampert, 2001). We expect these inventions more likely to be breakthroughs.

Count backward patent citations and citations to non-technical prior art

We calculate the number of citations to prior patents to measure the extent to which the patent relies on relatively more technical prior art. Moreover, we calculate the number of citations to non-technical prior art to measure the extent to which the patent relies on relatively more non-technical prior art. We expect breakthrough inventions to cite more non-technical prior art.

### 3.5 Independent Variables: Technological Novelty of Inventions

New recombination of technology subclasses or components

To identify patents which combine technological components for the first time in history, we use the 2008 US technology subclass concordance to go through all technology subclass assignments of all US granted patents in order to generate all first pairwise subclass combinations (Fleming et al., 2007). At the patent, level, we calculate a dummy in case the patent makes a least one new recombination as well as a the proportion of new pairwise recombinations among all possible pairwise combinations of a patent's technology subclasses. We expect breakthrough inventions to be more novel and to have a larger proportion of new component combinations.

### 3.6 Other Variables

We include the number of claims to capture the scope of protection granted to a particular patent. The larger the number of claims a patent covers, the larger the scope of protection and the more valuable a particular patent. Previous studies have found the number of claims to be associated with more forward citations received (Brandstetter and Sakakibara, 2001). Similarly, the number of patent classes measures the scope of the patent, i.e. the extent to which it covers multiple technology domains. It determines the extent of patent protection and monopoly power and hence the economic value of an invention (Scotchmer, 1991). Therefore, patents that cover more claims or technology fields are more likely to be breakthroughs. Finally, we also include the number of inventors and the number of applicants, technology main class dummies and application year dummies as controls.

### 3.7 Empirical methodology

To analyze the effects of the various dimensions of technological origins and novelty of inventions on the likelihood of being a breakthroughs as measured by being an outlier in the distribution of forward citation counts, we use a probit model.

Beyond looking at the likelihood of a breakthrough invention, we also want to compare the effects of technological origins and novelty on the average value of an invention, as measured by forward citation counts received within 5 years, and on the likelihood at a particularly poor invention, as measured by receiving no forward citations. In order to discover breakthroughs, researchers might need to take more inventive risks and for instance explore emerging technologies, source unfamiliar prior art from distant technology fields, or experiment with new recombinations of technological components. Such an explorative search might increase the chances at creating breakthroughs but at the same time reduce the average value of the created inventions and increase the chances at particularly poor outcomes (Fleming, 2001).

To this end, we also use negative binomial models to estimate the effect of our ex-ante patent characteristics on an invention's average impact as measured by forward citations counts received within five years after application. We use a probit model to predict the effects on the likelihood to have no forward citations, i.e. the likelihood of a particularly poor technological invention.

## 4. Results

### 4.1 Descriptive Statistics

Table 1 presents descriptive statistics for the samples of breakthrough and matched non-breakthrough patents according to the different measures of breakthrough (1 stdev and 3 stdev outliers in forward citation counts). 8.9% and 3.6% of the patents are breakthroughs in the 1 stdev and 3 stdev samples respectively. Table 1 includes a T-test for mean differences between the breakthrough patents and the control patents.

Insert table 1

In terms of the technological origins, we find breakthrough patents rely to a greater extent on technical and non-technical prior art. 12% and 13% of 1 stdev and 3 stdev breakthroughs have no backward citations compared to 24% and 25% for the non-breakthroughs. Technological breakthroughs on average cite 10 or 11 prior patents compared to 4 for the control group. In terms of non-technical prior art citations, the differences are even larger. 1 stdev and 3 stdev breakthroughs have on average 42 and 46 non-patent references compared to 24 or 27 for non-breakthroughs. The dissimilarity measures indicates breakthroughs have a set of backward patent citations which is more similar to prior art compared to the control group, 51% of the control patents are dissimilar compared to 43% and 44% for breakthroughs. So breakthroughs cite prior patents which have frequently been cited by other biotech patents. In terms of the age of the technical prior art cited in patents, we find most influential inventions on average rely on more recently developed technologies. The average backward citation lag for breakthroughs is 6 years and 10 months compared to 7 years and 3 or 4 months for the control groups. Furthermore, breakthrough inventions seem to use prior art from many different technology fields, particularly technology fields different from its own technology fields. They have a spread of cited patents across technology fields of 0.61 for 1 stdev breakthroughs and 0.62 for 3 stdev breakthroughs compared to 0.57 and 0.56 for the non-breakthroughs. 16% and 17% of all patents cited by breakthroughs are from technology fields different than the patent itself. For the control samples, the share is 14% and 12% respectively. Finally, breakthroughs seems to combine less familiar technological components or subfields, i.e. components which are used less and/or less recently by prior patents. On average, high impact patents have a component familiarity around 15 while the control groups have an average component familiarity around 18 and 19.

In terms of the technological novelty of the invention itself, breakthroughs are found significantly more likely to recombine two technology subfields or components for the first time in history. 47% of 1 stdev breakthroughs and 50% of 3 stdev breakthroughs have at least one new pairwise combination compared to 34% and 31% for the control samples. Breakthroughs also have a larger proportion of new pairwise component combinations among all potential pairwise combinations with an average score of 0.13 and 0.14 for 1 stdev and 3 stdev breakthroughs respectively, compared to 0.08 or 0.07 for non-breakthroughs.

Furthermore, we find evidence for breakthroughs to be broader in scope as they have a larger number of claims (22 and 23 on average compared to 15 for the control groups), more technology classes (2.1 and 2.2 compared to 1.8 for the control patents) and more technology subclasses (6.5 compared to 5.7 for the control patents).

Table 2 presents the bivariate correlation matrix. The correlations among the different indicators for breakthroughs and the indicators capturing the origins and nature of inventions are in line with the findings from the mean comparison T-tests in table 1. Patents with dissimilar backward patent citations rely on younger technical prior art (correlation -0.38) and have a larger proportion of new technology subclass combinations (correlation 0.13). Patents covering more technology classes also have a larger proportion of new combinations of technological components (correlation 0.24) as are patent which mostly cite technical prior art from technology fields different from its own (correlation 0.14). The strong negative correlation (-0.35) between component familiarity and proportion of new pairwise subclass combinations illustrates how difficult new combinations of most familiar components are.

Insert table 2

#### 4.2 Multivariate Analysis

Table 3 presents the results from our multivariate analysis. Besides testing the statistical and economic significance of the different determinants of breakthroughs, we also look at the overall predictive power of our model. To do so, we calculate the recall as the proportion of technological breakthroughs which are correctly predicted to be breakthrough, i.e. patents with a predicted probability of 0.5 or higher. The precision is calculated as the proportion of predicted breakthroughs, i.e. patents with a predicted probability of 0.5 or higher, which actually are breakthroughs. Finally, we also calculate the proportion of correctly predicted patents, i.e. both breakthrough and non-breakthroughs.

Insert table 3

In general, findings from the regression analysis are in line with the results from the descriptive statistics. Technological breakthroughs cite more technical and non-technical prior art, cite

younger technical prior art, and recombine technical prior art from many different technology fields, particularly technology fields different from its own. Breakthroughs are less likely to have dissimilar technical prior art citations compared to other biotech patents and as such are more likely to re-use technical prior art previously cited by many other biotech patents. Besides being well rooted in both technical and non-technical prior art and recombining technical prior art from many technology fields, breakthroughs themselves are more novel since they have a larger proportion of new recombinations of technological components. Breakthroughs are also broader in scope since they have a larger number of claims and cover more technology fields. They are however less likely to come from cooperation among multiple assignees. In what follows, we discuss in more detail the effects of technological origins, novelty and their interactions.

More backward patent citations increase the likelihood of being a breakthrough. A stdev increase in the number of backward patent citations increase the likelihood at a 1 stdev breakthrough with 1.8% and the likelihood at a 3 stdev breakthrough with 0.5% on average. The estimated difference in the probability of being a breakthrough between a patent with the largest number of backward citations in our sample and a patent without backward citation is 65% for 1 stdev breakthroughs and 17% for 3 stdev breakthroughs. Citing more prior patents also has a positive effect on the average number of forward citations and reduces the likelihood at 0 forward cites. Also citations to non-technical prior art have a positive effect. A stdev increase in the number of non-technical prior art citations increase the chances with 1.3% for 1 stdev breakthroughs and with 0.5% for 3 stdev breakthroughs while the difference between the patent with most non-technical prior art citations and patents without any citations is 32% and 11% for 1 and 3 stdev breakthroughs respectively. As for backward patent citations, citing more non-technical literature has a positive effect on the average value and reduces the likelihood at a particularly poor invention.

Contrary to expectations, patents with dissimilar backward patent citations compared to similar prior art are 2.3% and 1.2% less likely to have a big impact. Nevertheless, having a dissimilar combination of backward patent citations reduces the likelihood at particularly poor outcomes with 1.8% while having no significant effect on the average number of forward cites.

Furthermore, breakthroughs are found to build on more recent technical prior art. A stdev increase in the average age of cited patents decreases the likelihood with 2.2% for 1 stdev

breakthroughs and 1.1% for 3 stdev breakthroughs while the difference in likelihood between the patent with the youngest and the patent with the oldest patents cited is 16% and 8%. The age of technical prior art reduces the average value of the patent and increase the likelihood at 0 forward cites.

Besides relying on younger technical prior art, high impact patents seem to cite prior art from many different technology fields, particularly technology fields different from its own. A stdev increase in the spread of cited patents across technology fields increases the likelihood with 1.9% for 1 stdev breakthroughs and 0.9% for 3 stdev breakthroughs. The difference between a patent in our sample with the lowest spread and a patent with the highest spread is 5.9% and 2.8% respectively. Breakthroughs seem to cite proportionally more patents from technology fields different from its own fields. An increase of 1 stdev in the proportion of cited patents from different fields is associated with an increase of 0.4% for both 1 and 3 stdev breakthroughs while the difference in likelihood between a patent not citing any different technology fields and patent with only cites to different fields is around 1.7%. Citing patents from many different technology fields and technology fields different from its own also increases the average number of forward citations and reduces the probability of receiving no citations

We find mixed results for the familiarity of a patent's technological components. The frequent and recent use of technological components reduces the likelihood at a 1 stdev breakthroughs. A stdev increase in component familiarity reduce the likelihood with 0.4% for 1 stdev breakthroughs while for 3 stdev breakthroughs the coefficient is insignificant. Re-using more familiar components also reduces the average impact of a patent and increases the chances at poor inventions. A stdev increase in component familiarity increases the likelihood at 0 forward citations with 1.6%. These findings are somewhat different from Fleming (2001). He finds component familiarity has a positive impact on the average number of forward citations, and a U-shape relationship between component familiarity and variance of forward citations. This difference is potentially caused by the fact that we limit our study to biotechnology whereas Fleming (2001) takes all US patents granted during May or June of 1990.

Besides different technological origins, breakthroughs are also more novel as they have a larger proportion of new pairwise technological component combinations. A stdev increase in the proportion of new recombinations increase the likelihood with 1% for 1 stdev breakthroughs

and with 0.5% for 3 stdev breakthroughs. The difference between patents having no new combinations and patents having only new combinations is 5.2% for 1 stdev breakthroughs and 3.1% for 3 stdev breakthroughs. More novel patents receive more citations on average and are less likely to receive no citations.

In line with predictions, breakthroughs are also found to cover a larger part of the technology landscape as they contain more claims and technology classes. A stdev increase in the number of technology classes makes a patent 2.2% and 1.4% more likely to become a 1 stdev and 3 stdev breakthrough respectively, while a stdev increase in the number of claims makes a patent 2.2% and 0.9% more likely. The difference in the likelihood of being a breakthrough between a patent in our sample with the largest number of technology classes and a patent with just one class is 14.4% and 16.6% for 1 stdev and 3 stdev breakthroughs while for the number of claims the difference between the minimum and maximum number of claims is 102.3% and 17.2%. More claims increase the average impact of the patent and reduce the likelihood at 0 citations while the number of technology classes has a positive impact on the average value but no impact on poor outcomes.

Finally, we look at the interactions between technological origins and technological novelty (columns with even numbers in table 3). Having different technological origins compared to similar prior art has a negative impact on 1stdev breakthroughs. But when different technological origins lead to patents with a large share of new component combinations, this negative effect is mitigated. We find a positive interaction between dissimilar backward patent citations and the proportion of new subclass combinations on the breakthrough likelihood, which is significant at the 10% level for 1 stdev breakthroughs. The interaction effect increases the chance at a 1 stdev breakthrough with maximum 3%.

Using more familiar components has no significant impact on 3 stdev breakthrough in general. Yet, new recombinations of more familiar technological components increase the likelihood at a breakthrough, as evidenced by a positive interaction between the proportion of new pairwise technological component combinations and the average familiarity of the components. For 3 stdev breakthroughs, we find a strong positive and significant interaction between component familiarity and the proportion of new subclass combinations, increasing the likelihood with maximum 6%.

Technological breakthroughs are rare events with only 8.9% and 3.6% of the patents in our sample to be labeled as breakthrough according to the different definitions. This hampers the precision and recall of the different models. Despite the many significant determinants of breakthroughs being identified, the different probit models perform rather poorly with respect to recall and precision. The minimum recall is 3.6% while the maximum is 4.4%. So the full model can only correctly identify a small minority of all breakthroughs. The minimum precision is 43% while the maximum precision is 57%. Between 91.2% and 96.4% of all patent are correctly predicted. All this is a reminder of the high unpredictability inherently associated with the creation of breakthroughs.

## **5. Discussion and Conclusion**

Besides organizational and inventor-level determinants of breakthroughs, little empirical work has explored at the invention level the technological origins and novelty of most influential inventions. Using US patent data in the field of biotechnology, we explore different hypotheses put forward by scholars from the history of technology, and from the economics and strategy of technology and innovation. We combine various dimensions of technological origins and novelty and explore potential interactions among them. By jointly looking at the effects of the technological origins and novelty of inventions on their average value as well as on the likelihood of a very poor invention and on the likelihood of a technological breakthrough, we identify different trade-offs researchers face by taking more inventive risks while exploring breakthroughs.

A first important distinct feature of breakthrough inventions in biotechnology shown in the analysis is their technological novelty. Breakthroughs are significantly more novel as they are more likely to combine technological components or subfields for the first time in history and have a larger proportion of such new combinations.

With respect to their technological origins, our results show the evolutionary character of technological breakthroughs in biotechnology. Although breakthroughs are frequently labeled as revolutionary or discontinuous to prior art, we find biotech breakthroughs to build substantially on prior art as evidenced by having more citations to both technological and non- technological

prior art. Their technological roots are however distinct as they cite particularly more recent patents and patents from many different technology fields. Furthermore, breakthroughs rely to a greater extent on prior art from technology fields different from their own. Yet, breakthroughs are more likely to cite patents similar to the ones previously cited by other biotech patents.

Having dissimilar technological origins compared to similar prior art has a negative impact on breakthroughs. But when different technological origins lead to patents with a large share of new component combinations, this negative effect is mitigated by the positive interaction effect. Moreover, we find a positive interaction between component familiarity and the proportion of new component combinations. Re-using more familiar components leaves no significant impact on breakthroughs in general, it even increases the likelihood at zero cited inventions. Yet, re-using more familiar technological components in new recombinations leads to a higher likelihood of breakthroughs, increases the average number of forward citations and reduces the likelihood at zero citations.

Our findings when proven to be robust for other samples and specifications, have important implications for researchers, companies and other research organizations or governments willing to stimulate the creation of breakthroughs. First, the revolutionary nature of breakthroughs should be put into perspective. Breakthroughs, in biotech at least, rely to a greater extent on both technical and non-technical prior art rather than being pioneering in the sense of lacking any direct links to prior art. Therefore, more experienced researchers very familiar with a broad base of technical prior art might play an important role during the invention process. The strong effect of non-technical prior art suggests researchers should be aware of important, probably more recent, developments in other communities including scientific communities. Institutional boundary spanning individuals who are familiar with technical and scientific prior art and capable of translating this knowledge into technological applications might play an important role. Second, technological breakthroughs build on more recently developed technical prior art. As such, researchers should keep close track of emerging technological developments. Attending conferences and joining R&D consortia might help companies to keep track of potentially valuable new developments in the community. Third, the fact that most influential inventions both rely on prior art from many different technology fields as well as cover more technology fields themselves suggest breakthrough most likely come from the collaboration

between scientists and engineers with different backgrounds or from researchers with cross-disciplinary experience. Gatekeeping researchers maintaining a social network of researchers with a broad array of complementary skills and expertise might foster the diffusion of knowledge and ideas across technology fields. Companies with a broad R&D portfolio crossing many technology fields might benefit from economics of scope in the search for breakthroughs. They could stimulate the internal rotation and collaboration among inventors with different fields of expertise. Fourth, as breakthroughs rely to a larger extent on technical prior art from technology fields different from the ones in which they are currently applied, companies and inventors should be particularly aware of the importance of knowledge and technology outside their own technology communities and should engage in distant search strategies in order to tap into these potentially valuable sources of prior art. Researchers should be aware that brokering technological components or subfields for the first time in history significantly increases the likelihood of creating a technological breakthroughs, particularly when recombining more familiar components. Arts (2012) studies more carefully how established firms can create breakthroughs by recombining in a novel way the technological components with which the company has extensive experience from prior R&D.

Our paper has some obvious limitations mostly related to the use of patent data. First, not all technological inventions are patented so that our sample represents only a sample of all inventions. We try to partially overcome this problem by limiting our study to biotechnology, a field in which patenting is a very common practice and characterized by many breakthroughs. Extending our analysis to other fields could test the robustness of our findings. Second, the backward citations found in patent documents may only partially capture the technological origins of inventions. Moreover, patent citations might be added by patent attorneys or patent examiners and as such not trace the prior art used by the inventors themselves. Nevertheless, inventors have the legal obligation to disclose all relevant prior art. Third, there might be important characteristics in the technological origins and nature of inventions not captured by our indicators. Although patent documents contain a wealth of information on the origins and content of the invention, our indicators explore only most notable and quantifiable characteristics. Finally, our models have only very limited explanatory power in predicting breakthroughs. This implies important, potentially random, determinants of breakthroughs remain unexplained. Predicting rare events like breakthroughs is a difficult task, particularly

because the discovery of breakthroughs sometimes results from serendipitous observations rather than from an organized search process. A classic example is Alexander Fleming's accidental discovery of penicillin. However, as Pasteur's saying goes "luck favors the prepared mind". By identifying some common characteristics in the technological origins and novelty of breakthroughs, we hope our analysis at least marginally provides insight about how to turn luck on one's side.

## 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, pp. 521-543.
- Albert, M.B., Avery, D., Narin, F. and McAllister, P. (1991). Direct Validation of Citation Counts as Indicators of Industrially Important Patents. *Research Policy*, 20(3), pp. 251-259.
- Adner, R. and Levinthal, D. (2002). The Emergence of Emerging Technologies. *California Management Review*, 45, pp. 50-66.
- Arts, S. (2012). Path Dependency and Recombinant Exploration: How Established Firms Can Outperform in the Creation of Technological breakthroughs. MSI working paper, 1228.
- Arthur, W.B. (2009) *The Nature of Technology: what it is and how it evolves*, Free Press (Simon & Schuster), US.
- Arundel, A. and Kabla, I. (1998). What percentage of innovations are patented? empirical estimates for European firms. *Research Policy*, 27, pp. 127-141.
- Basalla, G. (1988). *The Evolution of Technology*. Cambridge: Cambridge University Press.
- Brandstetter, L. and Sakakibara, M. (2001). Do stronger patents induce more innovation? Evidence from the 1988 Japanese patent law reforms. *Rand Journal of Economics*, 32(1), pp. 77-100.
- Carpenter, M. and Narin, F. (1993). Validation study: patent citations as indicators of science and foreign dependence. *World Patent Information*, 5(3), pp. 180–185.
- Cattani, G. (2006). Technological Pre-adaptation, Speciation, and the Emergence of New Technologies: How Corning invented and Developed Fiber Optics. *Industrial and Corporate Change*, 15(2), pp. 285-318.
- Chandy, R.K. and Tellis, G.J. (1998). Organizing for Radical Product Innovation: The Overlooked Role of Willingness to Cannibalize. *Journal of Marketing*, 35(4), pp. 474-487.
- Christensen, C. (2003). *The Innovator's Solution*. Boston, MA: Harvard Business School Press.
- Dahlin, K.B. and Behrens, D.M. (2005). When is an invention really radical?: Defining and measuring technological radicalness. *Research Policy*, 34(5), pp. 717-737.
- Fleming, L. (2001). Recombinant Uncertainty in Technology Search. *Management Science*, 47(1), pp. 117-132.
- Fleming, L. (2002). Finding the Organizational Sources of Technological Breakthroughs: The Story of Hewlett-Packard's Thermal Ink-Jet. *Industrial and Corporate Change*, 11(5), pp. 1059-1084.
- Fleming, L. and Sorenson, O. (2004). Science as a map in technological search. *Strategic Management Journal*, 25, pp. 909-928.
- Fleming, L., Mingo, S. and Chen, D. (2007). Collaborative Brokerage, Generative Creativity and Creative Success. *Administrative Science Quarterly*, 52(3), pp. 443-475.

- Freeman, C. (1992). *The Economics of Hope*. New York: Pinter Publishers.
- Gambardella, A., Harhoff, D. and Verspagen, B. (2008). The Value of European Patents. *European Management Review*, 5(2), pp. 69-84.
- Hall, B.H., Jaffe, A.B. and Trajtenberg, M. (2001). The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools. NBER Working Paper, 8498.
- Hall, B.H., Jaffe, A.B. and Trajtenberg, M. (2005). Market Value and Patent Citations. *RAND Journal of Economics*, 36(1), pp. 16-38.
- Hargadon, A. (2003). *How Breakthroughs Happen: The Surprising Truth About How Companies Innovate*. Boston, MA: Harvard Business School Press.
- Harhoff, D., Narin, F., Scherer, F. and Vopel, K. (1999). Citation Frequency and the Value of Patented Inventions. *The Review of Economics and Statistics*, 81(3), pp. 511-515.
- Henderson, R. (1993). Underinvestment and incompetence as responses to radical innovation - evidence from the photolithographic alignment equipment industry. *RAND Journal of Economics*, 24(2), pp. 248-270.
- Henderson, R. and Clark, K.B. (1990). Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. *Administrative Science Quarterly*, 35(1), pp. 9-30.
- Hill, C.W.L. and Rothaermel, F.T. (2003). The Performance of Incumbent Firms in the Face of Radical Technological Innovation. *Academy of Management Review*, 28(2), pp. 257-274.
- Jaffe, A., Trajtenberg, M. and Fogarty, M.S. (2000). The Meaning of Patent Citations: Report on the NBER/Case-Western Reserve Survey of Patentees. NBER Working Paper, 7631.
- Lai, R., D'Amour, A., Yu, A., Sun, Y., Torvik, V. and Fleming, L. (2009). Disambiguation and co-author networks of the U.S. Patent Inventor Database. Harvard Business School working paper.
- Levinthal, D. (1998). The Slow Pace of Rapid Technological Change: Gradualism and Punctuation in Technological Change. *Industrial and Corporate Change*, 7(2), pp. 217-247.
- Mowery, D. and Rosenberg, N. (1998). *Paths of Innovation: Technological Change in 20th-Century America*. Cambridge: Cambridge University Press
- Nelson, R. and Winter, S. (1982). *An Evolutionary Theory of Economic Change*. Cambridge, MA: Belknap Press.
- National Science Board (NSB) (2007). *Enhancing Support of Transformative Research at the National Science Foundation*. Washington, DC: NSB-07-32.
- OECD. (2005). *A framework for Biotechnology Statistics*.
- Rosenberg, N. (1994). *Exploring the Black Box: Technology, Economics, and History*. Cambridge: Cambridge University Press.
- Rosenkopf, L. and Nerkar, A. (2001). Beyond local search: boundary-spanning, exploration, and impact in the optical disk industry. *Strategic Management Journal*, 22 (4), pp. 287-306.

- Scherer, F.M. and Harhoff, D. (2000). Policy Implications for a World with Skew-Distributed Returns to Innovation. *Research Policy*, 29, pp. 559-566.
- Schumpeter, J.A. (1934) *The theory of economic development*. Cambridge, MA: Harvard University Press.
- Schumpeter, J.A. (1942). *Capitalism, Socialism and Democracy*. New York: Harper and Row.
- Scotchmer, S. (1991). Standing on the shoulders of giants: Cumulative research and the patent law. *The Journal of Economic Perspectives*, 5(1), pp. 29-41.
- Schoenmakers, W. and Duysters, G. (2010). The technological origins of radical inventions. *Research Policy*, 39, pp. 1051-1059.
- Shane, S. (2001). Technological Opportunities and New Firm Creation. *Management Science*, 47 (2), pp. 205-220.
- Singh, J. and Fleming, L. (2010). Lone Inventors as Sources of Breakthroughs: Myth or Reality? *Management Science*, 56(1), pp. 41-56.
- Sorenson, J.B. and Stuart, T.E. (2000). Aging, Obsolescence, and Organizational Innovation. *Administrative Science Quarterly*, 45, pp. 81-113.
- Trajtenberg, M., Jaffe, A. and Henderson, R. (1997). University versus Corporate Patents: A Window on the Basicness of Invention. *Economics of Innovation and New Technology*, 5(1), pp. 19-50.
- Trajtenberg, M. (1990). A Penny for Your Quotes: Patent Citations and the Value of Innovations. *RAND Journal of Economics*, 21(1), pp. 172-187.
- Tushman, M.L. and Anderson, P. (1986). Technological discontinuities and organizational environments. *Administrative Science Quarterly*, 31(3), pp. 439-465.
- Usher, A.P. (1954). *A History of Mechanical Invention*. Cambridge, MA: Harvard University Press.
- Utterback, J. (1996). *Mastering the Dynamics of Innovation*. Boston, MA: Harvard Business School Press.

**TABLE 1: Descriptive Statistics**

	FC outlier 1SD	Mean	SD	Pr( T  >  t )	FC outlier 3SD	Mean	SD	Pr( T  >  t )
No back. cit.	0	0.24	0.43	0.0000	0	0.25	0.43	0.0000
	1	0.13	0.33		1	0.12	0.33	
# Back. cit.	0	4.44	8.27	0.0000	0	4.43	8.94	0.0000
	1	10.16	22.84		1	10.81	22.95	
Dissimilar back. cit.	0	0.51	0.50	0.0000	0	0.51	0.50	0.0003
	1	0.43	0.50		1	0.44	0.50	
Back. cit. lag	0	7.26	4.01	0.0000	0	7.34	4.10	0.0017
	1	6.88	3.49		1	6.82	3.27	
Spread back. cit.	0	0.57	0.21	0.0000	0	0.56	0.21	0.0000
	1	0.61	0.18		1	0.62	0.19	
Prop. back. cit. diff. field	0	0.14	0.26	0.0000	0	0.12	0.24	0.0000
	1	0.16	0.25		1	0.17	0.25	
Component fam.	0	17.94	13.31	0.0000	0	18.96	12.91	0.0000
	1	14.65	10.52		1	14.82	10.08	
New combi dummy	0	0.34	0.48	0.0000	0	0.31	0.46	0.0000
	1	0.47	0.50		1	0.50	0.50	
Prop. new combi.	0	0.08	0.17	0.0000	0	0.07	0.16	0.0000
	1	0.13	0.20		1	0.14	0.21	
NPRS	0	24.32	40.56	0.0000	0	26.83	45.70	0.0000
	1	41.57	65.69		1	46.37	65.19	
Count claims	0	14.69	14.14	0.0000	0	15.19	14.29	0.0000
	1	21.56	19.78		1	22.64	19.57	
Count main classes	0	1.87	0.75	0.0000	0	1.77	0.67	0.0000
	1	2.15	0.86		1	2.24	0.93	
Count subclasses	0	5.65	3.58	0.0000	0	5.71	3.67	0.0000
	1	6.54	4.48		1	6.51	4.09	
Count assignees	0	1.11	0.53	0.0368	0	1.10	0.48	0.5442
	1	1.09	0.45		1	1.11	0.52	
Count inventors	0	2.91	1.78	0.0451	0	2.94	1.74	0.5205
	1	2.98	1.88		1	2.99	1.80	

The descriptive statistics are based on the matched samples of patents sharing the same 3-digit technology classes, application year and grant year as at least one of the technological breakthrough patents

**TABLE 2: Correlation Matrix**

	FC 5y	FC 0	FC outlier 1SD	FC outlier 3SD	No back. cit.	# Back. cit..	Dissimilar back. cit.	Back. cit. lag	Spread back. cit.	Prop. back. cit. diff. field	Component fam.	New combi dummy	Prop. new combi.	NPRS	Count claims	Count main classes
FC 5y	1.0000															
FC 0	-0.2880*	1.0000														
FC outlier 1SD	0.6312*	-0.1606*	1.0000													
FC outlier 3SD	0.5971*	-0.0732*	0.4556*	1.0000												
No back. cit.	-0.0732*	0.1008*	-0.0648*	-0.0313*	1.0000											
# Back. cit.	0.1390*	-0.0739*	0.1221*	0.0830*		1.0000										
Dissimilar back. cit.	0.0007	-0.0877*	0.0006	0.0010		-0.2540*	1.0000									
Back. cit. lag	0.0114*	-0.0359*	0.0246*	0.0131*		0.2656*	-0.3778*	1.0000								
Spread back. cit.	0.0921*	-0.0650*	0.0839*	0.0478*		0.2690*	-0.3604*	0.5199*	1.0000							
Prop. back. cit. diff. field	0.0223*	-0.0846*	0.0191*	0.0122*		0.1289*	-0.1184*	0.3984*	0.2323*	1.0000						
Component fam.	-0.0648*	0.2752*	-0.0640*	-0.0313*		-0.0190*	-0.0604*	-0.1842*	-0.0372*	-0.2229*	1.0000					
New combi dummy	0.0784*	-0.1329*	0.0861*	0.0553*	-0.0805*	0.0380*	0.1222*	0.0765*	0.0652*	0.0463*	-0.3170*	1.0000				
Prop. new combi.	0.0731*	-0.1356*	0.0842*	0.0576*	-0.0681*	0.0183*	0.1298*	0.0728*	0.0390*	0.1401*	-0.3471*	0.6907*	1.0000			
NPRS	0.1126*	-0.0368*	0.0965*	0.0610*	-0.0608*	0.2603*	-0.0972*	0.0822*	0.1320*	-0.0172*	0.1042*	-0.0437*	-0.0694*	1.0000		
Count claims	0.1202*	-0.0581*	0.1084*	0.0690*	-0.0848*	0.1670*	-0.0911*	0.0575*	0.1098*	-0.0342*	0.0930*	0.0314*	-0.0198*	0.1218*	1.0000	
Count main classes	0.0892*	-0.0353*	0.1239*	0.0850*	-0.0373*	0.0657*	0.2327*	-0.0035	0.1255*	-0.1271*	0.0501*	0.3417*	0.2380*	0.0258*	0.0928*	1.0000
Count subclasses	0.0803*	-0.0343*	0.0920*	0.0669*	-0.0620*	0.0870*	0.0739*	0.0334*	0.0985*	-0.1045*	-0.0129*	0.4133*	0.1351*	0.0194*	0.1574*	0.5497*

\*indicates significant correlation at the 1% level

**TABLE 3: Regression Analysis**

Variables Model	FC 5y nbreg		FC 0 probit		FC outlier 1SD probit		FC outlier 3SD Probit	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
# Back. cit.	0.0159*** [0.001]	0.0159*** [0.001]	-0.0133*** [0.001]	-0.0132*** [0.001]	0.0117*** [0.001]	0.0117*** [0.001]	0.0080*** [0.001]	0.0080*** [0.001]
Dissimilar back. cit.	0.0028 [0.016]	0.0048 [0.016]	-0.0685*** [0.014]	-0.0679*** [0.014]	-0.1625*** [0.025]	-0.1587*** [0.025]	-0.1812*** [0.048]	-0.1779*** [0.048]
Back. cit. lag	-0.0426*** [0.002]	-0.0424*** [0.002]	0.0216*** [0.001]	0.0214*** [0.001]	-0.0340*** [0.003]	-0.0339*** [0.003]	-0.0355*** [0.006]	-0.0358*** [0.006]
Spread back. cit.	0.4859*** [0.029]	0.4810*** [0.029]	-0.3444*** [0.022]	-0.3464*** [0.022]	0.4523*** [0.042]	0.4584*** [0.043]	0.4474*** [0.088]	0.4327*** [0.088]
Prop. back. cit. diff. field	0.1723*** [0.027]	0.1818*** [0.028]	-0.0787*** [0.023]	-0.0782*** [0.023]	0.1219*** [0.042]	0.1196*** [0.042]	0.2469*** [0.076]	0.2577*** [0.076]
Component fam.	-0.0038*** [0.001]	-0.0013 [0.001]	0.0045*** [0.001]	0.0033*** [0.001]	-0.0023* [0.001]	-0.0011 [0.002]	-0.0003 [0.003]	0.0040 [0.003]
Prop. new combi.	0.3208*** [0.033]	0.4955*** [0.056]	-0.2065*** [0.029]	-0.2922*** [0.047]	0.3709*** [0.050]	0.3467*** [0.094]	0.4609*** [0.097]	0.5912*** [0.170]
NPRS	0.0023*** [0.000]	0.0023*** [0.000]	-0.0013*** [0.000]	-0.0013*** [0.000]	0.0021*** [0.000]	0.0021*** [0.000]	0.0018*** [0.000]	0.0018*** [0.000]
Count claims	0.0129*** [0.000]	0.0128*** [0.000]	-0.0103*** [0.001]	-0.0103*** [0.001]	0.0104*** [0.001]	0.0103*** [0.001]	0.0093*** [0.001]	0.0093*** [0.001]
Count main classes	0.0413*** [0.008]	0.0398*** [0.008]	-0.0026 [0.007]	-0.0012 [0.007]	0.2036*** [0.016]	0.2016*** [0.016]	0.3141*** [0.034]	0.3097*** [0.034]
Count subclasses	0.0076*** [0.002]	0.0075*** [0.002]	-0.0063*** [0.002]	-0.0061*** [0.002]	-0.0005 [0.003]	-0.0007 [0.003]	-0.0032 [0.005]	-0.0032 [0.005]
Count assignees	-0.0382*** [0.012]	-0.0324*** [0.012]	0.0335*** [0.010]	0.0312*** [0.010]	-0.0432** [0.020]	-0.0409** [0.020]	0.0160 [0.037]	0.0225 [0.037]
Count inventors	-0.0059* [0.003]	-0.0065* [0.003]	0.0054* [0.003]	0.0056* [0.003]	0.0062 [0.005]	0.0060 [0.005]	-0.0011 [0.011]	-0.0018 [0.011]
Prop. new combi. * Dissimilar back. cit.		0.0033 [0.056]		0.0294 [0.056]		0.1855* [0.105]		0.2061 [0.199]
Prop. new combi. * Prop. back. cit. diff. field		-0.1410 [0.107]		0.1030 [0.088]		-0.1375 [0.176]		0.2564 [0.310]
Prop. new combi. * Spread back. cit.		-0.2619** [0.116]		-0.1102 [0.093]		0.2934 [0.184]		-0.3713 [0.359]
Prop. new combi. * Component fam.		0.0204*** [0.004]		-0.0088** [0.003]		0.0090 [0.007]		0.0363*** [0.013]
Technology Class dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Application Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log pseudolikelihood	-170647.7	-170618.74	-38972.572	-38966.837	-10726.89	-10723.803	-2510.7047	-2506.3741
Pseudo R-Sq	0.0517	0.0518	0.1798	0.1800	0.1230	0.1233	0.1691	0.1705
Recall Pr( +   D )			37.07%	37.06%	4.42%	4.36%	3.99%	3.57%
Precision Pr( D   + )			63.03%	63.01%	56.94%	56.43%	46.67%	43.10%
Correctly classified			77.90%	77.90%	91.21%	91.21%	96.39%	96.37%
Observations	83,091	83,091	83,091	83,091	40,797	40,797	19,556	19,556

Robust standard errors in brackets

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Recall represent the proportion of technological breakthroughs which are correctly predicted to be breakthroughs  
Precision represent the proportion of actual technological breakthroughs out of the predicted technological breakthroughs