Paper to be presented at the

DRUID Society Conference 2014, CBS, Copenhagen, June 16-18

Technology Brokering and Breakthrough Invention

Sam Arts
KU Leuven
Faculty of Business and Economics
sam.arts@kuleuven.be

Reinhilde Veugelers
KULeuven
Managerial Economics, Strategy and Innovation (MSI)
Reinhilde.Veugelers@kuleuven.be

Abstract

Interpreting technological invention as an evolutionary and recombinant search process, we study the effect of technology brokering, i.e. creating new inventions by combining formerly disconnected but familiar technology components, on the likelihood of inventing more useful and particularly breakthrough inventions. For evidence, we consider 26 years of the US patent record in biotechnology. We find that recombining more familiar components and creating new component combinations stimulate average usefulness and likelihood of breakthrough. In particular, interaction effects illustrate how recombining more familiar components continues to foster the creation of more useful and breakthrough inventions but only as long as the familiar components are recombined in unprecedented ways.

Jelcodes:O31,O32

1. Introduction

Technological progress is a major driver of firm performance, economic growth and social prosperity (Schumpeter, 1934; Freeman, 1992). At the same time, technological inventions vary considerably in value. The value or usefulness of an invention is typically assessed by the extent to which it serves as prior art for subsequent technical progress (Trajtenberg, 1990; Harhoff et al, 1999). The large majority of technological inventions marginally improve existing technology and have a relatively small impact. Only a handful of inventions have a strong impact by serving extensively as prior art for many subsequent inventions and can be considered to be breakthroughs (Rosenberg, 1994). Although they are small in numbers, breakthroughs are generally considered as most important for value creation and growth (Schumpeter, 1942; Scherer and Harhoff, 2000).

Despite their importance, we know relatively little about the evolutionary origins of breakthroughs and the search process governing their discovery (e.g. Rosenberg, 1994). What characteristics of the search process are more likely to lead to big successes? On this, the literature is still scarce given that most of the literature looks at explaining average usefulness or value of inventions (e.g. Griliches, 1998; Rosenkopf and Nerkar, 2001). An exception are the detailed historical analyses on the evolutionary origins of breakthroughs (e.g. Usher, 1954; Basalla, 1988), a number of more recent qualitative case studies (e.g. Rabinow, 1996; Fleming, 2002; Hargadon, 2003a; Cattani, 2006), and a few empirical studies on particular characteristics of their origins (e.g. Dahlin and Behrens, 2005).

The literature portrays technological invention as the result of an evolutionarily recombinant search process (e.g. Schumpeter, 1939; Usher, 1954; Mokyr, 1990; Scotchmer, 1991; Basalla, 1998). This search process is typically characterized by the familiarity of the technological components which are being recombined and by the linkages made between these components, both affecting the usefulness or impact of an invention (e.g. Henderson and Clark, 1990; Dahlin and Behrens, 2005; Fleming, 2001).

The more technology components are used by prior inventions, the more inventors learn about successful and unsuccessful applications and the better their foresight in how to reuse these components in different ways and contexts (e.g. Cohen and Levinthal, 1990; Hargadon, 2003b). As such, the recombination of familiar, well-understood, components improves the screening and results in more useful inventions. Using patent data, Fleming (2001) confirms that recombining more familiar components results in more useful inventions on average. But how does familiarity of components affect the likelihood of creating exceptionally valuable inventions? Postulating that a higher variance of success is associated with a higher mass in both sides of the distribution and hence a higher chance at a breakthrough, Fleming (2001) finds that above a certain threshold of familiarity, recombination increases the variance of success and hence the likelihood of breakthroughs. However, other research suggests that experience and learning improve average performance but reduce experimentation and variability (March, 1991; Levinthal and March, 1993). Reusing familiar components might lead to local search and learning traps. As such, firms and inventors should explore unfamiliar and emerging technologies in order to create breakthroughs (e.g. Ahuja and Lampert, 2001). In light of this research, it remains unclear why the recombination of very familiar technologies continues to stimulate rather than restrict the development of more useful and breakthrough inventions. Are there no decreasing marginal and ultimately negative effects associated with the recombination of familiar components, particularly for the creation of breakthroughs? How can such decreasing effects be avoided? Moreover, is achieving higher variance of inventive success necessary to ensure an increase in the probability of breakthrough?

In this paper, we extend prior research by illustrating how particularly the interaction between the familiarity and the combination of components is key to understand the search process governing valuable and breakthrough inventions, as well as to understand the conditions under which the reuse of familiar technology can continue to drive inventive success. Relying on the idea that the combinatorial potential of familiar technology components is virtually unlimited (e.g. Romer, 1994; Weitzman, 1998), we explore technology brokering, i.e. the creation of new inventions by brokering formerly uncombined but familiar technologies, as an important search process rendering more usefulness and breakthrough inventions (Hargadon and Sutton, 1997; Hargadon, 2003b)[i]. To do so, we identify for each invention whether it was created by brokering formerly uncombined technologies for the first time in history. Subsequently, we test whether brokering formerly uncombined technology components results in more useful and breakthrough inventions. In addition, we study whether brokering allows to mitigate any potential diminishing returns from recombining more familiar components. To this end, we look at the interaction between component familiarity and new combinations. To analyze the effect of the recombinant search process on the usefulness of inventions, we look explicitly at the effect on the likelihood of exceptional success, i.e. breakthroughs, rather than on the variance of success. As an extension, we also look at average usefulness to identify if the recombinant search process governing breakthroughs results at the same time in lower average usefulness.

For evidence, we use 26 years of the US patent record in biotechnology, from the beginning of the modern biotechnology industry in 1976 until 2001 (Rothaermel, 2000). Biotechnology provides an interesting setting to study brokerage of formerly uncombined but familiar technologies. It is an interdisciplinary field of R&D bridging different fields of science and technology including chemistry, microbiology, biochemistry, chemical engineering and computer science (Smith, 2009; Phene et al., 2006). The evolution of biotechnology has been driven by recombining familiar but disparate strands of knowledge and technology (Phene et al., 2006; Smith, 2009). Moreover, biotechnology is a very innovation intensive field with many breakthrough discoveries over the studied period and with the large majority of inventions being patented (Arundel and Kabla, 1998). In line with prior empirical research, we use patent data to characterize both the recombinant search process and the usefulness of inventions. Patents provide detailed information on the components of an invention allowing us to trace prior use of the same components to assess their familiarity (e.g. Fleming, 2001; Kaplan and Vakili, 2014), and to identify inventions with formerly uncombined components (Fleming et al., 2007; Carnabuci and Operti, 2013). Moreover, forward citations allow us to identify an invention's usefulness, i.e. its impact on subsequent technological evolution (e.g. Trajtenberg, 1990; Harhoff et al, 1999). Patents receiving a disproportionately large number of citations can be considered as breakthroughs (e.g. Ahuja and Lampert, 2001; Conti et al., 2013).

The results strongly support the importance of technology brokering, i.e. combining for the first time formerly disconnected but familiar technologies. Recombining more familiar components improves inventive success but only as long as the familiar components are recombined in unprecedented ways. Reusing familiar components in conventional combinations does not result in more valuable inventions. On the contrary, the threats inherent in the reuse of more familiar components is that they are typically used in more conventional combinations while the latter reduces average usefulness and the likelihood of breakthrough. 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. Literature

2.1. The Evolutionary Origins of Breakthroughs

Technological progress and innovation is an evolutionary process whose origins can be traced (Schumpeter, 1934; Usher, 1954; Basalla, 1988; Mokyr, 1990). For example, Usher (1954, p.11) defines technological invention as "the constructive assimilation of pre-existing elements into new synthesis". Besides scholars from the history of technology (e.g. Usher, 1954; Basalla, 1998), scholars from the economics and management of ideas and innovation acknowledged how technological novelty is driven by the recombination of pre-existing knowledge and technology (e.g. Nelson and Winter, 1982; Weitzman, 1998; Arthur, 2009). Nelson and Winter (1982) state that "the creation of any sort of novelty in art, science or practical life consists to a substantial extent of a recombination of conceptual and physical materials that were previously in existence". Varian (2009) coins the term "combinational innovation" to denote the creation of innovation as a recombination of component technologies.

This holds not only for inventions in general but also for breakthroughs, i.e. those foundational inventions with a disproportionately large impact on subsequent technological progress. Although breakthroughs are frequently labeled as revolutionary or discontinuous to prior art, they typically also have a history of technical prior art (Basalla, 1988; Levinthal, 1998; Hargadon, 2003a; Adner and Levinthal, 2002). Also Utterback (1996) points out that "radical innovations often are seen to be based on the synthesis of well-known technical information or components". For instance, the transistor, often labeled as a radical or breakthrough invention replacing vacuum tubes, has its evolutionary origins in crystal radio sets which preceded vacuum tubes (Basalla, 1988).

2.2 Component familiarity, Technology Brokering and Breakthrough Invention

Interpreting invention as a process of recombining technology components, the inventive search process can be characterized along two dimensions: (i) by the familiarity of the components which are combined and (ii) by the linkage between these components (e.g. Henderson and Clark, 1990; Fleming, 2001; Kaplan and Vakili, 2014).

2.2.1. Component familiarity

When studying the familiarity of the technology components being recombined, the literature has looked at how much these components have been used by prior art (Fleming, 2001; Kaplan and Vakili, 2014). On the one hand, the frequency of components' prior use increases knowledge and understanding of successful and failed combinations, and stimulates insight in how to reuse the components in different combinations and contexts (e.g. Cohen and

Levinthal, 1990; Fleming, 2001; Hargadon, 2003b). Prior use directs inventors to successful combinations and to avoid less successful ones. As such, recombining well understood components, i.e. those used by many and recent prior inventions, will increase an invention's average usefulness and the likelihood it will be a breakthrough. On the other hand, reusing more familiar technology might reduce experimentation and variability, making it more difficult to come up with something exceptionally valuable (March, 1991; Levinthal and March, 1993; Ahuja and Lampert, 2001). By means of patent data, Fleming (2001) studies the effect of component familiarity on the mean and variance of invention value, the latter capturing both failure and breakthrough. He finds a positive and increasing effect of reusing more familiar components on average value and a u-shaped effect on the variability of value. Initially, the reuse of familiar components decreases the variance. This would imply a lower probability of breakthroughs. However, above a certain level of familiarity, reuse has a positive and increasing effect on the variance of success. As such, recombining "very familiar" components is most likely to direct inventors to the discovery of more useful and breakthrough inventions. Fleming's results would therefore surprisingly suggest no decreasing but increasing returns from familiarity on average invention value and breakthrough success.

2.2.2 Technology Brokering and Breakthrough Invention

A second dimension to characterize the recombinant search process is the nature of the combinations being made. Using knowledge and technology in new ways correspond to a search process labelled as exploration while using knowledge and technology in well-understood ways corresponds to a search process labelled as exploitation (March, 1991). The literature on creativity and innovation particularly looks at new combinations of pre-existing knowledge or technology. Using existing components in new combinations offers the possibility to create completely novel inventions (e.g. Arthur, 2009). Romer (1994), Weitzman (1998) and Varian (2009) argue that new combinations of existing components provide a potentially huge source of important new discoveries. In a theoretical model, Weitzman (1998) illustrates how the ultimate limits to growth do not lie so much in our ability to generate new ideas and technology as in our ability to successfully recombine the existing knowledge and technology in new ways.

The concept of technology brokering, introduced by Hargadon and Sutton (1997), refers explicitly to the process of creating new innovations from original combinations of familiar technology. These new combinations are objectively new inventions because they are built from existing but previously unconnected technologies (Weitzman, 1998; Arthur, 2009; Varian, 2009). In fact, many of the breakthrough inventions in the past originated from brokering familiar but formerly uncombined components (e.g. Adner and Levinthal, 2002; Hargadon 2003a; 2003b). For instance, Kary Mullis' Nobel prize winning invention of polymerase chain reaction (PCR), a biomedical technique for identifying and multiplying DNA sequences, which revolutionized technical progress in biotechnology. Kary Mullis described the process of inventing PCR as: *"In a sense, I put together elements that were already there, but that's what inventors always do. You can't make up new elements, usually.* The new element, if any, it was the combination, the way they were used*"* as cited by Rabinow (1996, p 6-7). Each of PCR's components were developed before the invention of PCR and were available to laboratories around the globe. Du Pont even filed a law suit against the PCR patents arguing that all of its components existed since the 1960s and were invented by Gobind Khorana, another Nobel prize laureate. The novelty or inventive step involved in the discovery of PCR is described by Rabinow (1996, p. 7) as: *"What was* original, powerful, and significant was the concept that combined -and reconfigured- these existing techniques.*"* This perspective is in line with Levinthal (1998) and Adner and Levinthal (2002) who discuss how breakthroughs can originate from combining existing but formerly disparate technological lineages. For example, the CAT scanner resulted from combining X-ray technology and computer technology (Adner and Levinthal, 2002). Cattani (2006) describes how Corning pioneered fiber optics by redeploying familiar knowledge and technology in specialty glass into the new application domain of fiber optics.

Besides anecdotal and case study evidence, the link between technology brokering and inventive success has not been investigated on a larger scale. Trajtenberg et al. (1997) argue that patents originating from the synthesis of dissimilar technology fields, as measured by patent classes of the cited patents, are more basic or original and diffuse more broadly. Rosenkopf and Nerkar (2001) find that inventions created by moving beyond local search, i.e. by citing more patents from other classes than the classes to which the patent itself belongs, are more useful or valuable. For U.S. biotechnology patents, Phene et al. (2006) find that patents which cite more non-biotech patents but from the same national origin are more likely to be breakthroughs. Similarly, Shane (2001) argues that inventions which cite patents from a large number of technology fields are more radical and more likely to be commercialized through the foundation of a new firm. Fleming (2001) studies the effect of combination familiarity, i.e. the number of prior patents using exactly the same combination of components, on the average and variance of inventive success. He finds that combination familiarity improves average usefulness but reduces the variance of success and hence the probability of breakthrough. Finally, Dahlin and Behrens (2005) find that inventions which rely on a dissimilar set of prior inventions are more novel and radical. As such, all these studies indicate that sourcing knowledge and technology from distant technology fields result in more novel and valuable inventions. Bridging different fields and communities increases the likelihood that novel technological combinations will emerge (Hargadon, 2003a). Yet, none of these studies identifies inventions created by combining formerly separate components, neither whether such new combinations include more or less familiar components. As such, the impact of brokering formerly uncombined technologies on inventive success has not been looked at in the empirical literature. In this paper, we hypothesize that recombining formerly disconnected components increases an invention's average usefulness and the likelihood of being a breakthrough.

H1: Recombining formerly disconnected components will increase an invention's usefulness and likelihood of being a breakthrough

As discussed above, the familiarity of the technologies or components is an important aspect of the recombinant search process and affects the usefulness of inventions (e.g. Levinthal, 1998; Fleming, 2001; Hargadon 2003a). However, it remains unclear why there are increasing positive rather than negative effects associated with the recombination of most frequently used components (Fleming, 2001). This finding stands in contrast with the literature on local search and learning traps which argues that firms and inventors should experiment with unfamiliar and emerging technologies in order to create more useful and breakthrough inventions (e.g. March, 1991; Ahuja and Lampert, 2001). One could expect that after a certain amount of reuse, the most valuable technological applications and combinations might have been exhausted, eventually reducing average usefulness and especially reducing the likelihood of becoming seminal for future technology progress. In this paper, we posit that such diminishing returns to the reuse of familiar components can be avoided by recombining the familiar components in novel ways. Recombining existing knowledge and technology in new ways correspond to a search process labelled as exploration (March, 1991). Building on the idea that the combinatorial potential of technology components is virtually unlimited (Romer, 1994; Weitzman, 1998), we hypothesize that reusing more familiar components continues to result in the creation of more valuable and breakthrough inventions as long as the familiar components are recombined in unprecedented ways.

H2: Recombining familiar components will increase an invention's usefulness and likelihood of a breakthrough as long as the familiar components are recombined in unprecedented ways

3. Research Setting, Data, Variables and Methodology

3.1 Research Setting

Biotechnology provides an interesting setting to study technology brokering given that it is an interdisciplinary field of R&D with a very long history. Biotechnology bridges separate fields of science and technology such as chemistry, microbiology, biochemistry, chemical engineering and computer science (Smith, 2009; Phene et al., 2006). The evolution of biotechnology has been driven by recombining familiar but disparate strands of knowledge and technology (e.g. Smith, 2009; Phene et al., 2006). For instance, Smith (2009, p. 6) argues: *"It involves the marshalling of concepts and methodologies from a number of* separate disciplines and applying them to a specific problem". Moreover, it is a very innovation intensive field with the large majority of technological inventions being patented (Arundel and Kabla, 1998). Although breakthrough inventions are relatively rare, biotechnology experienced a significant number of breakthroughs over the years providing us with a sufficient number of observations to test our predictions (Smith, 2009). A milestone in biotechnology was the discovery of the double helix structure of DNA by Watson and Crick in 1953. Yet, their findings penetrated very slowly into technological inventions until Cohen and Boyer's breakthrough discovery of recombinant DNA in 1973, illustrating the possibility to manipulate the structure of microorganisms. The beginning of the modern biotechnology industry is marked by the foundation of Genentech, the first biotechnology firm, in 1976 (Rothaermel, 2000). Our study will focus on the period 1976 until 2001.

3.2 Patent Data

In line with prior empirical research, we use patent data to study both the recombinant search process behind the creation of each technological invention as well as to identify each invention's impact on subsequent technological progress. While not all inventions are patented, patent data includes a large share of both failed and very successful technological discoveries. Patent data provides a detailed window on the components or technologies used to create an invention. The technology classes of a patent capture the technology fields covered by the patent while the subclasses (currently more than 150,000) correspond to the different components or technologies used to create the invention (Fleming, 2001; Carnabuci and Operti, 2013). Unlike patent citations which might be driven by strategic considerations of inventors or assignees, technology subclasses are assigned by the USPTO and therefore provide an objective indication of the different components of an invention. In case new technology (sub)classes are added or existing (sub)classes altered, all patents are retrospectively reclassified, allowing us to consistently measure the prior use of components over time. To identify inventions which broker formerly uncombined components, we identify patents which are the first in history with particular pairwise combinations of subclasses (Fleming et al., 2007).

3.3 Sample Selection

To identify all USPTO biotechnology patents, we made use of the OECD classification 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. The sample consists of 84,119 patents. Notice that while the analysis is restricted to this sample of 84,119 biotechnology patents, the whole US patent database is used to calculate some of the variables as explained below. Information on patents comes from the most recent National Bureau of Economic Research (NBER) patent database (www.nber.org/patents) and from the Harvard patent database (http://dvn.iq.harvard.edu/dvn/dv/patent).

3.4 Dependent Variables: Breakthrough, Usefulness and Failure

We define breakthroughs as those foundational inventions that have a disproportionately large impact on subsequent technological progress, independent from their evolutionary origins or degree of novelty (e.g. Rosenberg, 1994). We follow the tradition in the literature by identifying breakthroughs by means of forward patent citations (e.g. Ahuja and Lampert, 2001; Conti et al., 2013; Kaplan and Vakili, 2014). The number of citations a patent receives is correlated with its technological importance (Albert et al., 1991) as well as its social (Trajtenberg, 1990) and private value (Harhoff et al, 1999). As such, we interpret the number of citations a patent receives as it **usefulness** or impact. In line with prior research (Sing and Fleming, 2010), patents receiving no forward citations are treated as **failures**[ii]. In line with the actual value distribution of inventions (Schumpeter, 1942, Scherer and Harhoff, 2000), the distribution of forward citation counts is very skewed with a large share of patents receiving no or very little citations and a small share of patents receiving a large number of citations. Therefore, it is most likely that patents receiving most citations pertain to most important technological inventions.

Most 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 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. We calculate for all granted US patents the count of forward citations as the number of patents citing the patent and the truncated count of forward citations as the number of citations received within 5 years after application. The untruncated citation count is used to compare all patents sharing the same 3-digit technology class filed within the same year and the truncated citation count is used to compare all patents sharing the same 3-digit technology class irrespectively of their time of filing. For each of the distributions, we calculate the mean and standard deviation. Subsequently, a patent is labeled as **breakthrough** in case both its truncated and untruncated count of forward citations are larger than the mean plus three times the standard deviation in the respective distributions. The sample of breakthrough patents includes some of the most seminal inventions identified by experts as having a profound impact on subsequent technical progress and generating very large private and social value. For instance, the sample of breakthroughs includes the Cohen and Boyer recombinant DNA patent (US4237224) and the two polymerase chain reaction patents (US4683202 and US4683195) invented by Kary Mullis and colleagues. Both recombinant DNA and PCR are labelled by experts as some of the most important biotechnology inventions to date (e.g. Rabinow, 1996; Trajtenberg et al., 1997). Arts et al. (2013) provide more general validity for our measure of breakthrough in the field of biotechnology. The authors illustrate how patents labelled as breakthrough by means of forward citations are a very strong and significant predictor of being externally identified as breakthrough by experts.

3.5 Independent Variables

New Combinations. For each patent, we calculate a measure of recombinant novelty, capturing the extent to which the invention is created by combining particular components or technologies for the first time in history[iii]. To identify inventions created by brokering formerly disparate components, we use the 2008 US technology class concordance to go through all subclass assignments of all US granted patents in order to identify all previously uncombined pairs of subclasses (e.g. Fleming et al., 2007; Carnabuci and Operti, 2013). At the patent level, we construct new combinations as the patent's number of unprecedented subclass pairs divided by the patent's total number of subclass pairs. We look at the first combination of two technology subclasses in history independent from the order in which the subclasses appear on the patent document. Alternatively, prior research has used the cumulative number of prior patents using exactly the same combination of subclasses as a measure of combination familiarity (e.g., Fleming, 2001; Kaplan and Vakili, 2014). Given our purpose to study brokerage of formerly uncombined components, we opted for a measure identifying for each subclass whether it previously was combined with each of the other subclasses of the same invention in any kind of configuration. As such, our measure is more conservative. Looking at the whole set of subclasses rather than pairs of subclasses would imply that patents with a different number of subclasses are not compared to each other in order to identify recombinant novelty while they actually may be very similar[iv]. In addition, this would create a bias given that patents with a large number of subclasses would almost per definition be classified as representing a new combination. If we identify new combination as patents having a set of subclasses which has never appeared before, we find 82% of all patents have a new combination compared to 48% having a new pairwise combination. Finally, by averaging the number of novel subclass pairs by the patent's total number of subclass pairs, we get a continuous measure ranging from zero to one, and not just a binary variable, which is comparable across patents with a different number of subclasses.

Component Familiarity. To capture the extent to which an invention was created by recombining familiar components, i.e. components frequently used by prior inventions, we calculate a patent-level measure of component familiarity by taking the average familiarity of each of a patent's subclasses (Fleming, 2001; Kaplan and Vakili, 2014). Individual subclass familiarity is measured by the use of the same subclass by prior patents. More formally, familiarity of patent i's subclass j is calculated as

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

In line with prior research (e.g. Katila and Ahuja, 2002), we take 5 years as the time constant of knowledge loss, i.e. a yearly knowledge loss of 18%. See appendix 1 for an example of how component familiarity and new combinations are calculated for a particular patent.

3.6 Control Variables

We include additional invention and team level characteristics which might affect inventive success as control variables. Teams are found more likely to create breakthroughs compared to lone inventors, an effect which is partially mediated by the diversity of technical experience of the team and the size of their collaboration network (Singh and Fleming 2010). We include **team size** as the number of inventors on the focal patent, **average experience** as the average number of prior patents by the focal patent's inventors, **experience diversity** as the number of technology classes at least one of the focal patent's inventors has patented in before, and **network size** as the number of inventors at least one of the focal patent's inventors has collaborated with before the filing of the focal patent.

Finally, we include a number of invention level characteristics in line with prior research. First, we include **patent references** as the number of citations to prior patents and **non-patent references** as the number of citations to non-patent literature. In addition, we control for **number of classes** as the number of main 3-digit US technology classes. We also include the **number of subclasses** and a **single subclass** dummy. The latter control is necessary given that the recombinant search process underlying single subclass patents remains unobservable. For patents with a single subclass, new combinations is put to zero. Furthermore, the USPTO updates from time to time the classification of technology classes and subclasses, updating existing subclasses and creating new subclasses typically for the most successful lines of technology. Such a reclassification of patents creates new subclasses and subclass combinations after the initial patent filing. Because the ex post reclassification is likely related to the success of the initial inventions using the particular subclasses, we need to control for the potential bias. To do so, we include for each patent **newest subclass** as the minimum number of previous uses among the focal patent's subclasses (Fleming, 2001). In addition, the newest subclass variable controls for the alternative explanation that brokering formerly uncombined components increases inventive success because one of the components is relatively new, per definition increasing the likelihood of new combinations. As such, the positive effect of the new combinations measure could reflect the effect of including novel components rather than of brokering pre-existing but formerly uncombined components. Moreover, we include technology class dummies (3 digit) and application year dummies as controls. Finally, we cluster standard errors at the assignee level to control for remaining unobserved heterogeneity at the assignee level.

3.7 Empirical Methodology

To analyze the recombinant search process most commonly underlying the creation of breakthroughs, we estimate the likelihood of a patent being a breakthrough by means of probit models. In order to discover breakthroughs, inventors might need to embrace more uncertainty for instance by using less familiar technology or by experimenting with new combinations. Such an explorative search might increase the likelihood of creating breakthroughs but at the same time reduce the average usefulness. To this end, we also use negative binomial models to study the effect of technology brokering on the average usefulness of an invention as measured by the number of forward citations received within five years. Alternatively, one could estimate the effect of component familiarity and new combinations for different quantiles of the number of forward citations, and check to what extent the estimated coefficients vary across different quantiles. Results obtained from the quantile regressions are similar to the ones presented in the paper (results from quantile regressions are not reported but available upon request from the authors). To test H1 our independent variable of interest is new combinations where we expect a positive effect on breakthrough and number of forward citations.. H2 is tested by looking at the interaction of the two independent variables, i.e. new combinations and component familiarity. We expect a positive effect of the interaction on the likelihood of breakthroughs and average invention value (H2). To interpret the marginal effects of the different variables and their interactions, we use a simulation approach recommended by King et al. (2000) and Zelner (2009). To do so, we use the clarify package in Stata (Tomz et al., 2003). After estimating the probit model, the estimated coefficients are simulated thousand times based on the predicted coefficients and their associated confidence intervals. Subsequently, we simulate the percent change in the probability of a breakthrough associated with a particular change in covariates, as well as the confidence interval around the change in probability.

4. Results

4.1 Descriptive Statistics

Table 1 presents descriptive statistics for our sample of biotechnology patents. We find 1.5% of the patents to be labeled as breakthroughs using a three standard deviation cut off in forward citations. While representing only 1.5% of all patented inventions, breakthroughs receive 17% of all forward citations. They receive on average 80 citations (31 citations within 5 years) compared to 6 citations (2 within 5 years) for the non-breakthrough patents. Using the more restrictive four standard deviation cut off, 0.8% of the patents are classified as breakthroughs while the two standard deviation cut off classifies 2.9% of the patents as exceptionally valuable.

Insert table 1

In addition, we find 48% of the patents combine at least two formerly uncombined technology components while the average share of new combinations is 0.19, suggesting the large majority of components has been combined before the invention of the focal patent. As illustrated in table 2, 2.2% of the patents covering new combinations are breakthroughs compared to 0.9% of the patents without any new combination. As such, brokering formerly uncombined technologies significantly increases the chance of inventing a breakthrough, supporting H1. Besides a larger probability of breakthrough, these inventions receive more citations on average. As such, brokering formerly disparate technologies seems to increase average and exceptional success without increasing the variance of success.

Insert table 2

Inventions originating from the recombination of more familiar components are less valuable on average and are less likely to be breakthroughs, compared to inventions reusing less familiar components (table 2). This might suggest that there are diminishing returns to the reuse of familiar technology. However, these diminishing returns only hold for the subset of patented inventions with a low share of new combinations. For the subset of patents with a below average share of new combinations, there are no significant differences in likelihood of breakthrough or average usefulness between patents recombining more familiar components and patents recombining less familiar components. By contrast, for the subset of patents with an above average share of new combinations, the reuse of more familiar components seems to have a strong positive effect on breakthrough and average usefulness, supporting H2. The hazard associated with the reuse of more familiar components is that more familiar components are typically used in more conventional combinations. Only 7% of all biotechnology patents in our sample have an above average component familiarity and an above average share of new combinations. Nonetheless, the descriptive statistics already illustrate the power of creatively recombining familiar technology to stimulate inventive successes.

Table A.1 in appendix displays correlations among the different variables. Some variables, particularly average experience, experience diversity and network size, are highly correlated. Multicollinearity does not seem to be an issue given that the average variance inflation factors (VIF) of the full models are below 3, and the VIF of each variable is below 5 (Baum, 2006). Nonetheless, excluding the highly correlated variables does not change the results of the regression analysis with respect to our key explanatory variables. As expected, component familiarity and new combinations are negatively correlated. This negative correlation shows how more familiar components are typically used in more conventional combinations.

4.2 Multivariate Analysis: Probability of Breakthrough

As illustrated in Table 3, findings from the regression analysis are generally in line with our initial hypotheses and the results from the descriptive statistics. When looking at the recombinant novelty of inventions, we find strong evidence that brokering formerly disparate components increases the likelihood of breakthrough, supporting H1. The effect is substantial: a change in the share of novel combinations from 0 to 0.5 increases the probability of breakthrough with 95% while a change in new combinations from 0 to 1 increases the probability of breakthrough with 263% (table 3 column 2).

Insert table 3

Unlike the descriptive statistics, the regression analysis shows a positive effect from recombining more familiar components on the probability of breakthrough (table 2 column 2). A standard deviation increase in component familiarity is associated with a 46% increase in the likelihood of breakthrough. In contrast with Fleming (2001), who found a U-shaped relationship between component familiarity and variance of success, we do not find a significant non-monotonic effect of component familiarity on breakthrough (results not reported). Yet, we do find a critical moderator for the positive effect of component familiarity on breakthrough success, namely the recombinant novelty of the focal invention. The model in the third column of table three includes the interaction between component familiarity and new combinations, and shows that component familiarity increases the probability of breakthrough but only insofar the components are recombined in unprecedented ways. The difference between recombining more or less familiar technological components remains insignificant as long as the recombinant novelty of the invention is low, confirming H2.

Figure 1 illustrates the positive interaction between component familiarity and new combinations. For patents with a low share of new combinations, there is no difference between recombining either more or less familiar components on the creation of exceptional successes. By contrast, for patents with a high share of new combinations, using more familiar components increases the probability of breakthrough. The figure thus illustrates how breakthroughs most likely originate from brokering formerly uncombined but very familiar technologies. This result sheds a new insight on the moderator for the increasing returns from component familiarity on inventive success. Building further on more familiar components -standing on bigger shoulders- only allows to stand out -to see further- if these components are recombined in a novel fashion.

Insert figure 1

An alternative way of interpreting the positive interaction between component familiarity and new combinations is that exploiting the learning advantages from using more familiar and better understood components allows to improve the chances of breakthrough the higher the recombinant novelty of the invention. Figure 2 illustrates this interpretation of the positive interaction.

Insert figure 2

With respect to the control variables, we find a strong positive effect of experience diversity of the team, i.e. the number of technology fields any of the focal patent's inventors has experience with. This finding is in line with the literature suggesting that valuable inventions most likely originate from the collaboration between inventors from different communities (Hargadon, 2003). Bridging different communities increases the diversity of knowledge and technology available for recombination and increases the likelihood that successful combinations will emerge. The positive effect of the number of technology classes is in line with this hypothesis. The fact that the coefficient of newest subclass, measuring the minimum number of previous uses among the focal patent's subclasses, is very small and insignificant suggest that using relatively novel components does not improve the probability of breakthrough.

4.3 Multivariate Analysis: Average Usefulness

To test whether the effect of recombining familiar technologies and formerly disparate technologies on breakthrough is accompanied by a higher average value, we estimate the same regressions but with number of citations as dependent variable (table 4). As such, we get a better insight into the uncertainty and risks associated with the recombinant search process. If technology brokering results in a lower average usefulness besides a higher likelihood of breakthrough, this would be evidence in support of a positive effect of brokering on variance of success. If technology brokering results in a higher average usefulness, this would be evidence in support of a mean shifting effect.

Insert table 4

The results in table 4 show that the effect of new combinations and component familiarity work in the same direction for average usefulness and breakthrough. Brokering improves average usefulness, even more so if accompanied by the reuse of familiar components. As for breakthroughs, reusing more familiar component increases average usefulness of an invention but only as long as the familiar components are recombined in novel ways. The biotech data therefore do not support the increasing variance effect but favor a mean shifting effect. Further evidence in support of a mean shifting effect is found by estimating the likelihood of breakthrough by excluding failures from the sample, i.e. by comparing breakthroughs with moderately valuable patents (table 3 column 9), and by estimating the average usefulness while excluding breakthroughs (table 4 column 4). Patents receiving no forward citations are treated as failures. All results indicate that brokering familiar but formerly disconnected technologies increases the probability from moving from failure to moderately valuable and from moderately valuable to breakthrough. Running regressions with failure as dependent variable leads to identical conclusions (results not reported but available upon request).

4.4 Robustness Checks

One might worry that our identification of breakthrough inventions is too strict or not strict enough. As an alternative for patents being three standard deviation outliers in the distribution of forward citations, we identify two and four standard deviation outliers corresponding respectively with 3% and 0.8% of the patents being classified as breakthrough compared to the initial 1.5%. As shown in columns 4 and 5 of table 3, our findings remain consistent across these different classifications of breakthrough. The effect of new combinations on breakthrough becomes larger the more exceptional we define breakthroughs. An increase in new combinations from 0 to 1 increases the probability of success with 154% for two standard deviation breakthroughs, with 263% for three standard deviations breakthroughs (cf supra) and with 432% for four standard deviation breakthroughs. Recombinant novelty becomes increasingly critical for the most influential inventions.

In addition, one might worry that our measures reflecting component familiarity and new combinations are biased because our sample includes patents from 1976 until 2001. Because our sample starts around the beginning of the modern biotechnology industry, there are relatively fewer patents filed before 1976 compared to later years resulting in a lower component familiarity and a higher likelihood of brokering formerly uncombined components in the early years. As a robustness check, we rerun the analysis on patents filed between 1990 and 2001, dropping around 21,500 patents or 26% of the sample. Our findings remain consistent (table 3 column 6). Further reducing the sample to patents filed between 1995 and 2001 does not change the results (results not reported).

Furthermore, brokering formerly uncombined components might increase inventive success because one of the components appears for the first time or is relatively new per definition resulting in new combinations. In this case, the positive effect of new combinations on breakthroughs could reflect the effect of including relatively new components rather than brokering formerly uncombined but familiar components. We can reject this alternative explanation because the newest subclass control is found to be an insignificant predictor of breakthrough and because the new combinations measure remains positive and significant while controlling for newest subclass. In addition, we check whether the effect of technology brokering holds while including a dummy variable, **new subclass**, being one for patents having a subclass which appears for the first time. As shown in table 3 column 7, new combinations and the interaction between new combinations and component familiarity remains positive and significant. The new subclass variable is negative and insignificant so using novel components does not affect the likelihood of a breakthrough.

Finally, component familiarity, being the average familiarity for each of the subclasses of a patent, reflects both the prior and recent use of the components combined for the first time and the prior use of the formerly combined components. By consequence, the positive interaction between new combinations and component familiarity does not necessarily reflect that recombining more familiar components in new ways increases the probability of breakthrough. Potentially, the less familiar rather than the more familiar components part of the same invention are recombined in new ways. To more explicitly test hypothesis H2, we calculate **familiarity of components used in new combinations** as the average familiarity of the components part of any new pairwise subclass combination. By doing so, we additionally control for the potential bias that new subclasses are retrospectively assigned to more valuable patents. If this would be the case, familiarity of components used in new combinations should have a negative and significant impact. Restricting the analysis to patents covering at least one new combination, we find familiarity of components used in new combinations to have a positive and significant effect on breakthrough supporting hypothesis two (table 3 column 8). Moreover, new combinations remains positive and significant illustrating how the share of new combinations, i.e. the degree of recombinant novelty, rather than just the fact of having a new combination, determines inventive success.

5. Discussion and Conclusion

Characterizing technological invention as an evolutionary and recombinant search process, this paper studies the role of technology brokering, i.e. the creation of new inventions by recombining formerly disconnected but familiar technology components, as a key search process behind the discovery of more useful and breakthrough inventions. Using USPTO patent data in the field of biotechnology, we identify the recombinant search process behind the creation of each patented invention at the time of filing, and subsequently analyze how successful the inventions become in the future as measured by forward citations. We characterize the search process by the prior and recent use of the focal invention's components, i.e. component familiarity, and by the fact whether the invention was created by combining formerly disconnected components.

The results strongly support the importance of brokering formerly disparate but familiar technologies to generate breakthrough inventions. Brokering leverages the power of reusing well understood technology components. Recombining frequently used rather than relatively novel components results in the creation of breakthroughs but only when the components are recombined in unprecedented ways. Besides stimulating the creation of breakthroughs, brokering familiar technologies trims the likelihood of failure and increases average usefulness. As such, it has a mean shifting effect rather than an increasing variance effect. Most frequently however, familiar components are used in more conventional combinations. Learning and experience with technology reduces experimentation and variability, which are key for the discovery of breakthroughs. The reuse of familiar components in conventional combinations does not result in more useful inventions. On the contrary, it reduces average usefulness and trims the likelihood of breakthrough.

Our findings when proven robust for other samples and specifications have important implications for inventors, companies, research organizations or policymakers willing to stimulate the creation of impactful technological discoveries. The results illustrate the power of combinatorial innovation, especially for breakthroughs. Teams which master the process of creatively recombining familiar technology will stand out (e.g. Hargadon and Sutton, 1997). Experimenting with formerly disconnected components positively affects inventive success, particularly when creatively recombining familiar components. As such, the revolutionary nature of breakthroughs should be put into perspective. Breakthroughs, in biotech at least, rely to a greater extent on pre-existing knowledge and well understood technology used by many prior inventions (Fleming, 2001), rather than being pioneering in the sense of using emerging technology or lacking links to prior art (Ahuja and Lampert, 2001). Prior and recent use of components stimulates learning about useful and useless applications and fosters insight in how to reuse the components in new combinations (Cohen and Levinthal, 1990; Hargadon, 2003a). Yet, reusing familiar components is only supportive in case they are recombined in a new way while new combinations positively affect inventive success independent from the familiarity of the components. The drawback is that reusing familiar components might lead to learning traps (Levinthal and March, 1993), i.e. more familiar components are typically used in conventional combinations. In this case, there is a higher probability of failure, a lower average usefulness and a lower likelihood of breakthroughs.

Our paper has some 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, probably a sample of more successful inventions. Furthermore, the discovery of breakthroughs is very rare. We try to mitigate these problems by limiting our study to biotechnology, a field in which patenting is a very common practice and characterized by many breakthroughs over the studied period. Extending our analysis to other fields could test the robustness of our findings. Second, we make the assumption that the (anticipated) value of patents does not affect the indicators capturing the characteristics of the recombinant search process, i.e. component familiarity and new combinations. As discussed in the paper, this might be an issue particularly for the new combinations measure as subclasses are retrospectively reclassified typically for more successful lines of technology. To control for a potential bias, we included newest and new subclass, measured the familiarity of components used in new combinations, and ran the analysis for different time windows, without finding any evidence of a potential bias. Nevertheless, results should be interpreted with this potential bias in mind. Third, there might be important characteristics in the search process not captured by our indicators. In this paper, we limit ourselves to studying new combinations and component familiarity as main characteristics of the search process. Fourth, according to the original definition of technology brokering, it involves recombining old ideas and technologies in new ways by "spanning multiple, otherwise disconnected industries and markets and, by doing so, put themselves in a position to be the first to see how existing technologies in one market could be used to create breakthrough innovations in another" (Hargadon, 2003b). In this paper, we use unprecedented pairwise subclass combinations to identify inventions resulting from what we labeled in this paper as technology brokering. As such, our identification of brokering does not explicitly take into account whether the formerly uncombined technologies come from otherwise disconnected industries and markets, just the fact whether pre-existing and uncombined technologies are used in new ways. Nonetheless, the strong and significant effects of the patent's number of classes and of the team's experience diversity suggest bridging diverse communities spurs inventive creativity and success. Fifth, predicting rare events like breakthroughs is a difficult task and our models have only limited predictive power. This implies potentially important, potentially random, determinants of breakthroughs remain unexplained or inexplicable. The discovery of breakthroughs sometimes results from serendipitous observation rather than from an organized search process. A classic example is Alexander Fleming's accidental discovery of penicillin, or Désiré Collen's discovery of the recombinant tissue plasminogen activator, a protein involved in the breakdown of blood clots and used to treat strokes, and touted as biotechnology's first blockbuster drug. However, serendipity does not imply the absence of a recurring pattern underlying the development of significant inventions. As Pasteur's saying goes "luck favors the prepared mind". By identifying some common characteristics in the recombinant search process underlying the discovery of breakthroughs, we hope our analysis at least marginally provides insight about how to turn luck on one's side.

References

Adner, R. and D. Levinthal (2002), 'The Emergence of Emerging Technologies', California Management Review, **45**, 50-66.

Ahuja, G. and C.M. Lampert (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., D. Avery, F. Narin and P. McAllister (1991), 'Direct Validation of Citation Counts as Indicators of Industrially Important Patents', Research Policy, **20**(3), 251-259.

Arthur, W.B. (2009), The Nature of Technology: what it is and how it evolves. Free Press (Simon & Schuster): US.

Arts, S., Appio, F., Van Looy, B. (2013). Inventions shaping technological trajectories; do existing indicators provide a comprehensive picture: The case of biotechnology. Scientometrics, **97**(2), 397-419.

Arundel, A. and I. Kabla (1998), 'What percentage of innovations are patented? empirical estimates for European firms', Research Policy, **27**, 127-141.

Basalla, G. (1988), The Evolution of Technology. Cambridge University Press: Cambridge.

Baum, C. (2006). An Introduction to Modern Econometrics Using Stata. STATA Press.

Carnabuci, G., E. Operti (2013). 'Where Do Firms' Recombinant Capabilities Come From? Intraorganizational Networks, Knowledge, and Firms' Ability to Innovate Through Technological Recombination', Strategic Management Journal, **34**(13), 1591-1613.

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), 285-318.

Cohen, W.M. and D.A Levinthal. (1990), 'Absorptive Capacity: A New Perspective on Learning and Innovation', Administrative Science Quarterly, **35**(1), 128-152.

Conti, R., A. Gambardella and M. Mariani (2013), 'Learning to Be Edison: Inventors, Organizations, and Breakthrough Inventions', Organization Science, forthcoming.

Dahlin, K.B. and D.M. Behrens (2005), 'When is an invention really radical?: Defining and measuring technological radicalness', Research Policy, **34**(5), 717-737.

Fleming, L. (2001), 'Recombinant Uncertainty in Technology Search', Management Science, **47**(1), 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), 1059-1084.

Fleming, L., S. Mingo and D. Chen (2007), 'Collaborative Brokerage, Generative Creativity and Creative Success', Administrative Science Quarterly, **52**(3), 443-475.

Freeman, C. (1992), The Economics of Hope. Pinter Publishers: New York.

Griliches, Z. (1998), Patent statistics as economic indicators: a survey. R&D and productivity: the econometric evidence. University of Chicago Press, 287-343.

Hargadon, A. and R.I. Sutton (1997), ‘Technology Brokering and Innovation in a Product Development Firm’. Administrative Science Quarterly, **42**(4), 716-749.

Hargadon, A. (2003a), How Breakthroughs Happen: The Surprising Truth About How Companies Innovate. Harvard Business School Press: Boston, MA.

Hargadon, A. (2003b), ‘Retooling R&D: Technology brokering and the pursuit of innovation’, Ivy Business Journal, November/December, 1-7.

Harhoff, D., F. Narin, F. Scherer and K. Vopel (1999), ‘Citation Frequency and the Value of Patented Inventions’, The Review of Economics and Statistics, **81**(3), 511-515.

Henderson, R. and K.B. Clark (1990), ‘Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms’, Administrative Science Quarterly, **35**(1), 9-30.

Hsu D. and K. Lim (2013), ‘Knowledge Brokering and Organizational Innovation: Founder Imprinting Effects’, Organization Science, forthcoming.

Lai, R., A. D'Amour, A. Yu, Y. Sun, V. Torvik and L. Fleming (2011), ‘The careers and co-authorship networks of U.S. patent-holders, since 1975’, Harvard Business School working paper.

Kaplan, S. and K. Vakili (2014), ‘The Double-edged Sword of Recombination in Breakthrough Innovation’, Strategic Management Journal, forthcoming.

Katila R and G. Ahuja (2002). Something old, something new: a longitudinal study of search behavior and new product introduction. Academy of Management Journal, **45**(6): 1183–1194.

King, G., M. Tomz and J. Wittenberg (2000). ‘Making the Most of Statistical Analyses: Improving Interpretation and Presentation’, American Journal of Political Science, **44**(2), 347-361.

Levinthal, D. (1998), ‘The Slow Pace of Rapid Technological Change: Gradualism and Punctuation in Technological Change’, Industrial and Corporate Change, **7**(2), 217–247.

March, J. (1991), ’Exploration and Exploitation in Organizational Learning’, The Organization Science, **2**(1), 71-87.

Mokyr, J. (1990), The Lever of the Riches: Technological Creativity and Economic Progress. Oxford University Press: New York.

Nelson, R. and S. Winter (1982), An Evolutionary Theory of Economic Change. Belknap Press: Cambridge, MA.

OECD. (2005), A framework for Biotechnology Statistics.

Phene, A., K. Fladmoe-Lindquist and L. Marsh (2006), ‘Breakthrough Innovations in the U.S. Biotechnology Industry: The Effects of Technological Space and Geographic Origin’, Strategic Management Journal, **27**, 369-388.

Rabinow, P. (1996). Making PCR: A Story of Biotechnology. University of Chicago Press.

Romer, P.M. (1994), ‘Economic Growth and Investment in Children’, Daedalus, **123**(4), 141-154.

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

Rosenkopf, L. and A. Nerkar (2001), ‘Beyond Local Search: Boundary-Spanning, Exploration, and Impact in the Optical Disk Industry’, Strategic Management Journal, **22**(4): 287–306.

Rothaermel, F.T. (2000), ‘Technological discontinuities and the nature of competition’, Technology Analysis and Strategic Management, **12,** 149–160.

Scherer, F.M. and D. Harhoff (2000), ‘Policy Implications for a World with Skew-Distributed Returns to Innovation’, Research Policy, **29**, 559-566.

Schumpeter, J.A. (1934), The theory of economic development. Harvard University Press: Cambridge, MA.

Schumpeter, J.A. (1939), Business Cycles. McGraw-Hill Book Company: New York.

Schumpeter, J.A. (1942), Capitalism, Socialism and Democracy. Harper and Row: New York.

Scotchmer, S. (1991), ‘Standing on the Shoulders of Giants: Cumulative Research and the Patent Law’, The Journal of Economic Perspectives, **5**(1), 29-41.

Shane, S. (2001), ‘Technological Opportunities and New Firm Creation’, Management Science, 56(1), 41-56.

Singh, J. and L. Fleming (2010), ‘Lone Inventors as Sources of Breakthroughs: Myth or Reality?’, Management Science, **56**(1), 41-56.

Smith, J.E. (2009), Biotechnology: 5th Edition. Cambridge University Press: Cambridge, UK.

Sorenson, J.B. and T.E. Stuart (2000), ‘Aging, Obsolescence, and Organizational Innovation’, Administrative Science Quarterly, **45**, 81-113.

Tomz, M., J. Wittenberg and G. King. (2003), ‘CLARIFY: Software for Interpreting and Presenting Statistical Results’. Version 2.1. Stanford University, University of Wisconsin, and Harvard University. Available at http://gking.harvard.edu/

Trajtenberg, M. (1990), ‘A Penny for Your Quotes: Patent Citations and the Value of Innovations’, RAND Journal of Economics, **21**(1), 172-187.

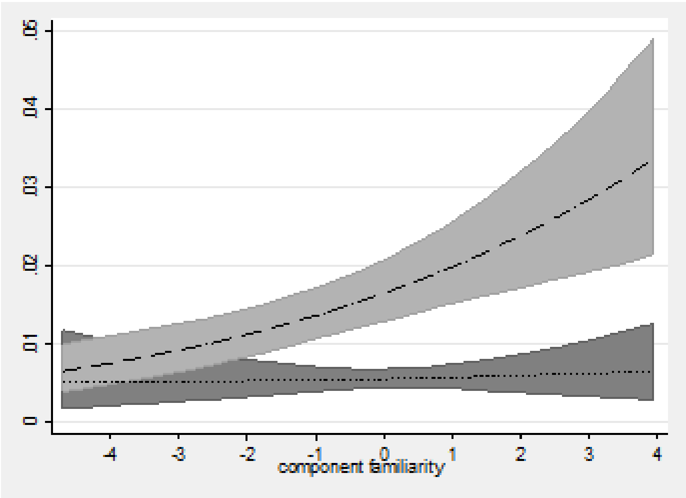
Usher, A.P. (1954), A History of Mechanical Invention. Harvard University Press: Cambridge, MA.

Utterback, J. (1996), Mastering the Dynamics of Innovation. Harvard Business School Press: Boston, MA.

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

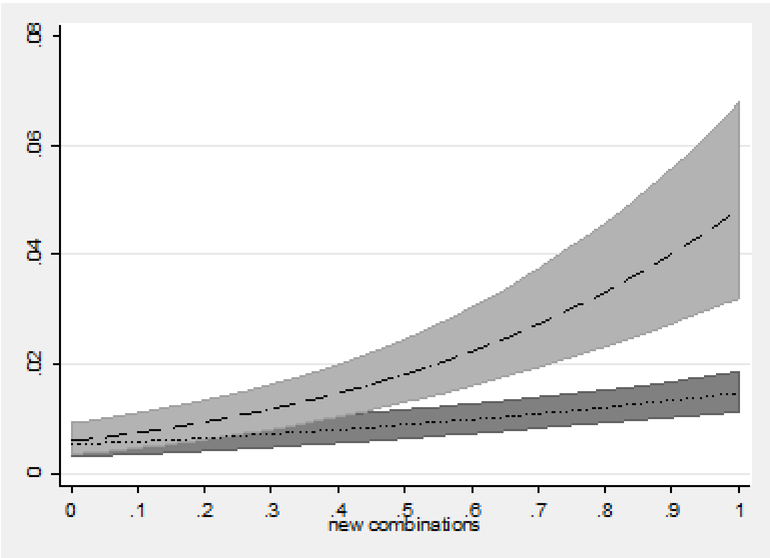
Zelner, BA. (2009), ‘Using simulation to interpret results from logit, probit and other nonlinear models’, Strategic Management Journal, **30**(12), 1335-1348.

FIGURE 1: Probability of a Breakthrough by Component Familiarity for High versus Low New Combinations



The dark grey plot area represents the 95% confidence interval around the estimated probability of a breakthrough for patents with new combinations equal to the 10th percentile, i.e. no new combinations, based on the probit model in table 3 column 3. The light grey plot area represents the 95% confidence interval around the estimated probability of a breakthrough for patents with new combinations equal to the 90th percentile, i.e. 2/3 of all pairwise subclass combinations of the focal patent are new.

FIGURE 2: Probability of a Breakthrough by New Combinations for High versus Low Component familiarity



The dark grey plot area represents the 95% confidence interval around the estimated probability of a breakthrough for patents with component familiarity equal to the 10th percentile based on the probit model in table 3 column 3. The light grey plot area represents the 95% confidence interval around the estimated probability of a breakthrough for patents with component familiarity equal to the 90th percentile.

TABLE 1: Descriptive Statistics (n=84,119)

Variable	Description	Mean	Stdev.	Min.	Max.
Breakthrough	Binary: 3 Stdev outlier in distribution of forward citations	0.015	0.122	0.00	1.00
Breakthrough 2 stdev	Binary: 2 Stdev outlier in distribution of forward citations	0.029	0.169	0.00	1.00
Breakthrough 4 stdev	Binary: 4 Stdev outlier in distribution of forward citations	0.008	0.092	0.00	1.00
Forward citations	Number of forward citations received within 5 years	2.86	5.86	0.00	201.00
New combinations	The focal patent's number of pairwise subclass combinations which appear for the first time in history divided by the focal patent's total number of pairwise subclass combinations	0.19	0.30	0.00	1.00
Component familiarity	Recent and frequent usage of the focal patent's subclasses by all prior US patents (see Fleming, 2001)	4.71	1.96	0.00	8.65
Team size	Number of inventors	1.27	0.42	0.69	3.50
Average experience	The average number of prior patents by the focal patent's inventors	1.33	1.02	0.00	6.00
Experience diversity	The number of technology classes at least one of the focal patent's inventors has patented in before	1.57	0.95	0.00	5.04
Network size	The number of inventors at least one of the focal patent's inventors has collaborated with before	1.71	1.34	0.00	6.22
Patent references	The number of backward patent citations	1.37	0.99	0.00	6.60
Non-patent references	The number of citations to non-patent literature	2.26	1.41	0.00	6.98
Number of classes	The number of technology classes	0.68	0.48	0.00	2.77
Number of subclasses	The number of technology subclasses	1.62	0.67	0.00	4.50
Single subclass	Binary: single technology subclass	0.04	0.20	0.00	1.00
Newest subclass	The minimum number of previous uses among the focal patent's subclasses	3.11	1.72	0.00	8.65
New subclass	Binary: at least one subclass appears for the first time in history	0.07	0.26	0.00	1.00

In the regression analysis, all explanatory variables are logged after adding 1 for those variables with zero values with the exception of binary and fractional variables

Table 2: Descriptive Statistics on Component familiarity, New combinations and Inventive Success

SAMPLE	% of patents	% breakthroughs	Forward citations mean
Full	100	1.5	2.9
New combinations=0	52	0.9	2.5
New combinations>0	48	2.2	3.3
New combinations<mean	69	1.1	2.6
New combinations>mean	31	2.5	3.5
Component familiarity<mean	48	1.7	3.0
Component familiarity>mean	52	1.3	2.7
New combinations=0 and component familiarity<mean	17	1.0	2.5
New combinations=0 and component familiarity>mean	35	0.9	2.4
New combinations>0 and component familiarity<mean	31	2.2	3.2
New combinations>0 and component familiarity>mean	17	2.2	3.2
New combinations<mean and component familiarity<mean	24	1.1	2.7
New combinations<mean and component familiarity>mean	45	1.0	2.5
New combinations>mean and component familiarity<mean	25	2.3	3.3
New combinations>mean and component familiarity>mean	7	3.1	4.2

TABLE 3: Probit Model Technology Breakthrough

VARIABLES	(1) Breakthrough	(2) Breakthrough	(3) Breakthrough	(4) Breakthrough 2stdev	(5) Breakthrough 4stdev	(6) Breakthrough	(7) Breakthrough	(8) Breakthrough	(9) Breakthrough
SAMPLE	1976-2001	1976-2001	1976-2001	1976-2001	1976-2001	1990-2001	1976-2001	New combinations>0	Excl. failures
New combinations		0.4819***	0.6121***	0.5413***	0.6948***	0.5673***	0.6096***	0.3768***	0.5898***
		[0.079]	[0.064]	[0.052]	[0.081]	[0.078]	[0.068]	[0.098]	[0.068]
Component familiarity		0.0697***	0.0288	0.0176	0.0299	0.0202	0.0280		0.0299
		[0.018]	[0.027]	[0.018]	[0.032]	[0.031]	[0.028]		[0.029]
New combinations *Component familiarity			0.0968***	0.1014***	0.1012**	0.0969***	0.0927**		0.0880***
			[0.033]	[0.023]	[0.041]	[0.035]	[0.041]		[0.034]
Team size	-0.0369	-0.0290	-0.0280	-0.0504	-0.0230	-0.0527	-0.0281	-0.0039	-0.0194
	[0.045]	[0.045]	[0.045]	[0.036]	[0.054]	[0.053]	[0.045]	[0.056]	[0.047]
Average experience	-0.1094***	-0.1022**	-0.0952**	-0.0928***	-0.1218**	-0.0839*	-0.0952**	-0.1295***	-0.0923**
	[0.041]	[0.040]	[0.040]	[0.030]	[0.054]	[0.049]	[0.040]	[0.039]	[0.042]
Experience diversity	0.2013***	0.1982***	0.1910***	0.2072***	0.2039***	0.1619***	0.1910***	0.2156***	0.1787***
	[0.038]	[0.037]	[0.037]	[0.028]	[0.046]	[0.046]	[0.037]	[0.044]	[0.039]
Network size	-0.0269	-0.0268	-0.0294	-0.0348	-0.0196	-0.0020	-0.0294	-0.0097	-0.0228
	[0.029]	[0.029]	[0.029]	[0.022]	[0.034]	[0.035]	[0.029]	[0.029]	[0.031]
Patent references	0.1835***	0.1910***	0.1842***	0.2051***	0.1652***	0.1901***	0.1841***	0.1861***	0.1627***
	[0.020]	[0.020]	[0.020]	[0.016]	[0.022]	[0.023]	[0.020]	[0.022]	[0.021]
Non-patent references	0.0848***	0.0817***	0.0851***	0.0903***	0.0936***	0.0776***	0.0851***	0.0881***	0.0816***
	[0.015]	[0.016]	[0.016]	[0.012]	[0.020]	[0.020]	[0.016]	[0.015]	[0.016]
Number of classes	0.3136***	0.2427***	0.2459***	0.2714***	0.2790***	0.1818***	0.2461***	0.3534***	0.2505***
	[0.043]	[0.046]	[0.046]	[0.033]	[0.057]	[0.055]	[0.046]	[0.055]	[0.047]
Number of subclasses	0.1146***	0.1288***	0.1316***	0.0959***	0.1254***	0.1078***	0.1316***	0.1854***	0.1178***
	[0.033]	[0.032]	[0.032]	[0.024]	[0.039]	[0.037]	[0.032]	[0.040]	[0.034]
Single subclass	0.2626***	0.3725***	0.3266***	0.2850***	0.4064***	0.2711**	0.3282***		0.3120***
	[0.097]	[0.102]	[0.096]	[0.072]	[0.119]	[0.108]	[0.098]		[0.101]
Newest subclass	-0.0106	-0.0018	0.0175	0.0155	0.0386*	0.0268	0.0173	0.0486*	0.0196
	[0.014]	[0.017]	[0.018]	[0.013]	[0.021]	[0.020]	[0.018]	[0.026]	[0.019]
New subclass							-0.0190		
							[0.122]		
Familiarity components used in new combinations								0.0289*	
								[0.017]	
Year fixed effects	Incl.	Incl.	Incl.	Incl.	Incl.	Incl.	Incl.	Incl.	Incl.
Technology fixed effects	Incl.	Incl.	Incl.	Incl.	Incl.	Incl.	Incl.	Incl.	Incl.
Log likelihood	-5643.1052	-5612.6824	-5603.4974	-9606.3937	-3446.8549	-4222.2669	-5603.49	-3564.5516	-5209.7608
Observations	84,119	84,119	84,119	84,119	84,119	62,628	84,119	40,521	53,586

Variables used in interaction terms are centered, Robust standard errors in brackets, Clustered at assignee level, *** p<0.01, ** p<0.05, * p<0.1

TABLE 4: Negative Binomial Model Forward Citations

VARIABLES	(1) Forward cit.	(2) Forward cit.	(3) Forward cit.	(4) Forward cit.
SAMPLE				Excl. breakthroughs
New combinations		0.3080***	0.4705***	0.3192***
		[0.055]	[0.047]	[0.038]
Component familiarity		0.0423***	0.0178	0.0059
		[0.011]	[0.011]	[0.008]
New combinations*Component familiarity			0.0806***	0.0647***
			[0.017]	[0.011]
Team size	-0.1222***	-0.1171***	-0.1171***	-0.0886***
	[0.029]	[0.028]	[0.028]	[0.024]
Average experience	-0.0922***	-0.0893***	-0.0825***	-0.0732***
	[0.027]	[0.027]	[0.027]	[0.020]
Experience diversity	0.1955***	0.1950***	0.1890***	0.1648***
	[0.022]	[0.022]	[0.022]	[0.018]
Network size	-0.0148	-0.0148	-0.0166	-0.0183
	[0.017]	[0.017]	[0.017]	[0.013]
Patent references	0.2367***	0.2410***	0.2356***	0.2126***
	[0.012]	[0.012]	[0.011]	[0.010]
Non-patent references	0.0656***	0.0627***	0.0650***	0.0563***
	[0.009]	[0.009]	[0.009]	[0.007]
Number of classes	0.1246***	0.0816***	0.0840***	0.0533***
	[0.019]	[0.021]	[0.021]	[0.017]
Number of subclasses	0.0742***	0.0769***	0.0796***	0.0637***
	[0.019]	[0.018]	[0.018]	[0.014]
Single subclass	0.1114***	0.1728***	0.1392***	0.0639*
	[0.041]	[0.041]	[0.041]	[0.037]
Newest subclass	-0.0022	0.0000	0.0172*	0.0134*
	[0.008]	[0.009]	[0.009]	[0.008]
Year fixed effects	Incl.	Incl.	Incl.	Incl.
Technology fixed effects	Incl.	Incl.	Incl.	Incl.
lnalpha	0.4715***	0.4683***	0.4668***	0.3184***
	[0.018]	[0.018]	[0.018]	[0.014]
Log likelihood	-173742.84	-173664.82	-173625.55	-163923.31
Observations	84,119	84,119	84,119	82,843

Variables used in interaction terms are centered, Robust standard errors in brackets, Clustered at assignee level,
*** p<0.01, ** p<0.05, * p<0.1

Appendix

TABLE A.1: Example calculation component familiarity and new combinations for US4683195

While working for Cetus Corporation, Kary Mullis and colleagues developed polymerase chain reaction (PCR), a biomedical technique for identifying and multiplying DNA. This resulted in two patents and Kary Mullis receiving the 1993 Nobel prize in chemistry. One of the two patents, US4683195, titled “process for amplifying, detecting, and/or-cloning nucleic acid sequence” was filed in 1986 and granted in 1987. By 2006, the patent received 1,460 forward citations illustrating its impact on subsequent technological progress. The USPTO assigned 7 technology subclasses to the patent. As illustrated in the table below, each of these subclasses was previously assigned to granted patents filed before the PCR patent, ranging from 1 prior patent for subclass 435/91.2 to 324 prior patents for subclass 435/6. As argued by Kary Mullis himself (Rabinow, 1996, p 6-7), and as illustrated in the table, each of PCR’s components were used by prior inventions. In a next step, this prior use is corrected for a yearly knowledge loss of 18%, and hence for the application year of prior patents covering the same subclass, resulting in an average component familiarity of 76.6.

Subclass	# prior patents	# prior patents corrected for knowledge loss
435/6	324	189.5
435/91.2	1	0.8
435/91.41	147	84.5
436/501	218	136.6
436/508	42	15.1
436/63	159	81.6
436/94	54	28.3
average	135	76.6

For the new combinations measure, all pairwise subclass combinations of patent US4683195 are retrieved. Given that the patent covers 7 subclasses, there are 21 unique subclass pairs as illustrated in the table below. Of the 21 subclass pairs, 11 appear for the first time in the patent database resulting in the new combinations measure being 0.52. As such, about half of the subclass pairs were used by prior patents and about half of the subclass pairs appear for the first time.

Subclass pair		First time
436/94	436/501	0
436/94	435/91.41	1
436/94	435/91.2	1
436/94	435/6	0
436/94	436/63	0
436/94	436/508	1
436/63	435/91.41	1
436/63	436/501	0
436/63	435/91.2	1
436/63	435/6	0
436/63	436/508	1
436/508	435/91.2	1
436/508	436/501	0
436/508	435/91.41	1
436/508	435/6	0
436/501	435/91.2	1
436/501	435/6	0
436/501	435/91.41	0
435/91.41	435/91.2	1
435/91.41	435/6	0
435/91.2	435/6	1

TABLE A.2: Correlation Matrix (n=84,119)

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1)	Breakthrough	1.00												
(2)	Forward citations	0.60	1.00											
(3)	New combinations	0.04	0.05	1.00										
(4)	Component familiarity	-0.01	-0.03	-0.67	1.00									
(5)	Team size	0.01	0.00	-0.13	0.16	1.00								
(6)	Average experience	0.01	0.02	-0.12	0.17	0.13	1.00							
(7)	Experience diversity	0.04	0.06	-0.10	0.13	0.36	0.81	1.00						
(8)	Network size	0.01	0.02	-0.17	0.22	0.49	0.78	0.82	1.00					
(9)	Patent references	0.08	0.16	0.05	-0.08	-0.03	0.14	0.16	0.07	1.00				
(10)	Non-patent references	0.04	0.07	-0.15	0.26	0.04	0.01	-0.01	0.01	0.11	1.00			
(11)	Number of classes	0.07	0.08	0.15	0.11	0.04	0.05	0.09	0.04	0.06	0.05	1.00		
(12)	Number of subclasses	0.05	0.07	0.12	0.09	0.02	0.02	0.05	0.02	0.10	0.05	0.59	1.00	
(13)	Single subclass	-0.01	-0.02	-0.13	-0.08	-0.01	0.01	0.00	0.00	-0.03	-0.03	-0.29	-0.50	1.00
(14)	Newest subclass	-0.04	-0.05	-0.64	0.81	0.12	0.17	0.11	0.20	-0.09	0.19	-0.11	-0.23	0.10

[i] According to the original definition of technology brokering (Hargadon, 2003b), it involves recombining old ideas and technologies in new ways by "spanning multiple, otherwise disconnected industries and markets and, by doing so, put themselves in a position to be the first to see how existing technologies in one market could be used to create breakthrough innovations in another". In this paper, we do not explicitly take into account whether the formerly uncombined technologies come from otherwise disconnected industries and markets, just the fact whether formerly uncombined but existing technologies are combined in new ways.

[ii] Given the increasing use of patents for strategic reasons, even patents without (many) forward citations can still be very useful for instance for extracting licensing fees or for patent blocking.

[iii] In this paper, we explicitly focus on new combinations of technologies or components at the level of a single invention, i.e. a single invention is developed by combining formerly disparate components. However, at a different level of analysis, e.g. the entire patent portfolio of a single organization, the components might have been combined with the components of other inventions of the same portfolio. As such, our notion of new combinations varies with the level of analysis. We thank an anonymous reviewer for this comment.

[iv] This potentially results in a bias given the large difference in the number of subclasses in our sample, ranging from 1 to 90. For instance, a patent with 50 subclasses would be treated as highly novel even if there would exist many prior patents covering 49 of the 50 subclasses. The level of recombinant novelty would be exactly the same for a patent with 10 subclasses where none of the subclasses has been used with each of the others in any kind of configuration.