DISCUSSION PAPER SERIES

DP13447

HOW FAST IS THIS NOVEL TECHNOLOGY GOING TO BE A HIT?

Reinhilde Veugelers, Michele Pezzoni and Fabiana Visentin

INDUSTRIAL ORGANIZATION

ISSN 0265-8003

HOW FAST IS THIS NOVEL TECHNOLOGY GOING TO BE A HIT?

Reinhilde Veugelers, Michele Pezzoni and Fabiana Visentin

Discussion Paper DP13447
Published 13 January 2019
Submitted 12 January 2019

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre’s research programme in **INDUSTRIAL ORGANIZATION**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

HOW FAST IS THIS NOVEL TECHNOLOGY GOING TO BE A HIT?

Abstract

Despite the high interest of scholars in identifying successful inventions, little attention has been devoted to investigate how (fast) the novel ideas embodied in original inventions are re-used in follow-on inventions. We overcome this limitation by empirically mapping and characterizing the trajectory of novel technologies' re-use in follow-on inventions. Specifically, we consider the factors affecting the time needed for a novel technology to be legitimated as well as to reach its full technological impact. We analyze how these diffusion dynamics are affected by the antecedent characteristics of the novel technology. We characterize novel technologies as those that make new combinations with existing technological components and trace these new combinations in follow-on inventions. We find that novel technologies combining for the first time technological components which are similar and which are familiar to the inventors' community require a short time to be legitimated but show a low technological impact. In contrast, combining for the first time technological components with a science-based nature generates technologies with a long legitimation time but also high technological impact.

JEL Classification: O33

Keywords: technological novelty, Diffusion, combinatorial components, patent data

Reinhilde Veugelers - reinhilde.veugelers@kuleuven.be
KU Leuven and CEPR

Michele Pezzoni - michele.pezzoni@unice.fr
University Cote d'Azur

Fabiana Visentin - visentin@merit.unu.edu
MERIT-University of Maastricht

Acknowledgements
Comments received from the participants at GSE Summer Forum (Barcelona), European Association for Evolutionary Political Economy (NICE), STI conference (Leiden), Collegio Carlo Alberto seminar (Turin), Economics of Research & the Design of Science Policy workshop (Bordeaux), KUL-MSI, UMaastricht, and KOF-ETH-UZH International Economic Policy seminar (Zurich) are gratefully acknowledged. Veugelers acknowledges the financial support from FWO (G.0825.12).

How fast is this novel technology going to be a hit?
Antecedents predicting follow-on inventions

Michele Pezzoni[1], Reinhilde Veugelers[2], Fabiana Visentin[3]

(Draft - Please do not quote or cite without permission from the authors)

Abstract: Despite the high interest of scholars in identifying successful inventions, little attention has been devoted to investigate how (fast) the novel ideas embodied in original inventions are re-used in follow-on inventions. We overcome this limitation by empirically mapping and characterizing the trajectory of novel technologies' re-use in follow-on inventions. Specifically, we consider the factors affecting the time needed for a novel technology to be legitimated as well as to reach its full technological impact. We analyze how these diffusion dynamics are affected by the antecedent characteristics of the novel technology. We characterize novel technologies as those that make new combinations with existing technological components and trace these new combinations in follow-on inventions. We find that novel technologies combining for the first time technological components which are similar and which are familiar to the inventors' community require a short time to be legitimated but show a low technological impact. In contrast, combining for the first time technological components with a science-based nature generates technologies with a long legitimation time but also high technological impact.

JEL code: *O33*

Keywords: technological novelty, diffusion, combinatorial components, patent data

Acknowledgements: Comments received from the participants at GSE Summer Forum (Barcelona), European Association for Evolutionary Political Economy (NICE), STI conference (Leiden), Collegio Carlo Alberto seminar (Turin), Economics of Research & the Design of Science Policy workshop (Bordeaux), KUL-MSI, UMaastricht, and KOF-ETH-UZH International Economic Policy seminar (Zurich) are gratefully acknowledged. Veugelers acknowledges the financial support from FWO (G.0825.12).

[1] Université Côte d'Azur, CNRS, GREDEG, France; ICRIOS, Bocconi University, Milan, Italy; BRICK, Collegio Carlo Alberto, Torino, Italy; email: michele.pezzoni@unice.fr

[2] KULeuven; email: reinhilde.veugelers@kuleuven.be

[3]UNU-MERIT, School of Business and Economics University of Maastricht, Maastricht, Netherlands; visentin@merit.unu.edu

1. INTRODUCTION

Ever since the disappointing productivity growth in the West has been related to the failure of non-frontier firms to catch up (e.g. Andrews et al. 2015), improving the diffusion of novel technologies has become a more central issue in the policy discussion on stimulating economic growth. Yet, diffusion of novel technologies is a less well studied phenomenon within the innovation literature, which concentrates mostly on understanding the generation rather than the diffusion of inventions (Arts et al. 2013). Nevertheless, several studies of the diffusion of a single industrial application of a novel technology exist (see Stoneman and Battisti (2010), Hall (2004), Geroski (2000) or Rogers (1983) for a review of this literature). Perhaps the most known and seminal one, is Griliches' 1957study of the diffusion of hybrid corn seeds in the Midwestern United States. A common finding in these studies is that the diffusion process usually starts out slowly among pioneering adopters, reaching "take-off" when the effects of peer influence kick in and a growing community of adopters is established, to then level-off as the population of potential adopters becomes exhausted, thus leading to an "S-shaped" cumulative adoption curve.

Despite the commonly found S-shape of diffusion, there is nevertheless a wide variation across different inventions in the time to take-off and to reaching full impact (Rosenberg 1976). In order to explain these differences, the literature so far has looked at the characteristics of the demand side for explaining different adoption rates of “users” over time, such as changing cost and benefits to adopting the new technology and social network factors (such as information sharing, connectivity). Being constrained by samples of individual or small sets of inventions to trace their diffusion, the empirical literature has so far not been able to look in a systematic fashion at the characteristics of the initial inventions that embed the novel technology as explanatory factors for differences in diffusion patterns.

Another neglected issue in the diffusion literature is the improvements and new inventions that may emerge along and through its diffusion. Most of the literature typically treats the focal innovation as given when following its diffusion path and the users as passive adopters. Yet, Rosenberg (1982) and Dosi (1991), among others, have emphasized that the diffusion of innovations is often accompanied by applications in different environments and learning about their use which in turn leads to improvements to the original innovation. Although many studies have described this process of innovation enhancement during its diffusion qualitatively, there has been relatively little systematic collection of data or explicit modeling of this process (Hall, 2004).

Follow-on inventions have been studied in the evolutionary economics literature on technology trajectories, initiated by Dosi (1982). Dosi (1982) makes a distinction between technological progress along a defined path or trajectory on the one hand and the process of search and selection on new technological paradigms, shifting the incumbent trajectories, on the other hand. There are however few empirical studies that try to measure and compare these trajectories for various technologies. Andersen (1999) using patent stock data between 1920 and 1990, studies the technology growth patterns across technology classes. She confirms that the S-shaped growth form is an appropriate and good approximation to describe the paths of evolution, but that there is a lot of cross-technology differences in takeoffs and periodicity of the time span of each cycle. This literature looks at the evolution of all patents within a technology class. The link between the patents belonging to the same trajectory is thus their joint belonging to the same technology class. It does not look explicitly at the relationship between the initial patent and the follow-on patents as re-users of the novel idea embedded in the initial patent. It can therefore not be used to understand differences in diffusion patterns of novel technologies.

Our contribution aims to improve our understanding of the differences in diffusion patterns of novel technologies. Using large scale patent data, we identify the follow-on inventions "using" the novel ideas embedded in original inventions. By looking at all the patented inventions that build on that novel technology, we can characterize the diffusion dynamics and analyze how these are affected by the antecedent characteristics of the novel technology.

We consider a novel technology as the result of a new combination of already existing components (Schumpeter 1939; Nelson and Winter 1982; Arthur 2009). Following Fleming (2001), Strumsky and Lobo (2015), Verhoeven et al. (2016), we take the technological classes used in the European Patent Office classification as representing technological components and novel technologies as patents with a combination of technological classes which is unprecedented in the history of patented inventions. For these novel technologies, we trace their diffusion into follow-on inventions. Follow-on inventions are identified as those patented inventions that re-use the new combination introduced by an initial invention which embedded for the first time the novel combination. We validate our set of follow-on inventions as reflecting a relevant technology link, by applying a text analysis technique.

We identify 10,782 novel technologies in the patent data that have a sufficient number of follow-on inventions to be traced over time. Looking at the diffusion curve of each of these novel technologies, we test an S-shaped curve, as commonly found in the diffusion literature. We are particularly interested in identifying the parameters of the curve that measure the time that a novel technology needs to be legitimated within the inventors' community and its maximum technological impact defined as the total number of follow-on inventions that the novel technology generates. We next look at what can explain any differences in diffusion patterns across the novel technologies. For explanatory factors, we focus on the ex-ante characteristics of the novel technology, particularly on the characteristics of the components combined for the first time in the initial invention. By investigating how the characteristics of the combined components affect the legitimation time and the technological impact, we identify the antecedent characteristics of the initial invention predicting a “fast” diffusion and a “hit” technology.

Controlling for other technology, applicant, inventor’s characteristics and time effects and testing for selection biases into successful diffusion, we find that combining for the first time technological components which are similar and which are familiar to the inventors’ community generates a novel technology that requires a short time to be legitimated but with a low technological impact. Combining for the first time technological components with a science-based nature generates technologies with a higher technological impact, but with a longer legitimation time. Our results, suggesting a trade-off between technological impact and legitimation time, thus provide new insights for the economics of innovation literature that is interested in identifying the key drivers of technological diffusion.

The remaining of the paper is organized as follow. Section 2 develops the hypothesis on how the characteristics of the newly recombined technological components affect the diffusion of the resulting novel technology. Section 3 describes the data and methods used. Section 4 presents the results. Section 5 concludes.

2. NOVEL TECHNOLOGY DIFFUSION AND ITS ANTECEDENTS

While a large number of studies has investigated the diffusion of an innovation looking at its use by the relevant population of potential adopters (Hall 2004), a limited number of studies has investigated the diffusion path of a novel technology in the follow-on pool of inventions. Previous works rely on case studies developed for specific technologies. For instance, Achilladelis

(1993) follows the diffusion of the novel technology related to Sulpholamide antibacterial drugs by tracing all the patented drugs based on Sulpholamide which appeared over 70 years after its introduction.

Differently from this literature, our study adopts a systematic large scale quantitative approach that allows us (i) to identify the introduction of a novel technology, (ii) to trace its diffusion pattern in follow-on inventions, and (iii) to assess the antecedent characteristics of the initial novel technology impacting on its diffusion pattern.

2.1. Identifying a novel technology

The first step in our analysis is to define a novel technology. In his seminal work on the origins of innovation, Schumpeter claimed that "innovation combines components in a new way, or that it consists in carrying out a new combination" (1939, pp. 88, Schumpeter 1939). Fleming (2001) elaborated on the concept of invention as a process of "recombinant search" (pp. 118) by arguing that inventors search among existing technological components and recombine them to realize something new. In the same vein, Arthur (2009) states that novel technologies are the result of the combination of existing components and that these existing components are themselves technologies. According to this "recombinant search" approach, we consider a novel technology as the result of pre-existing technological components that are combined for the first time (Verhoeven et al. 2016).

2.2. Tracing novel technology diffusion

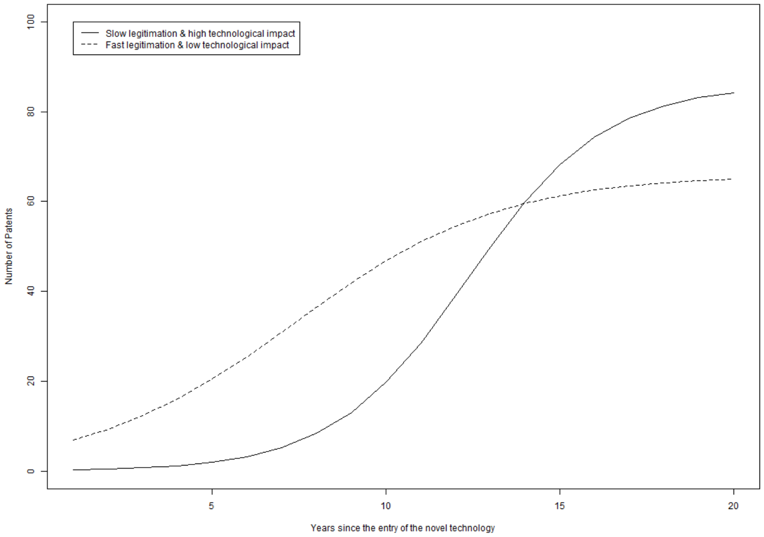
Having identified the novel technologies, the next step in the analysis is to trace their diffusion in follow-on inventions, i.e. identifying how many other inventions make "use" of the novel technology. However, not all novel technologies diffuse. As the aim of our study is to analyze the diffusion of successful technologies, we consider those novel technologies that recombine existing technological components *and* reach a minimum level of diffusion. Following Amabile (1996) and Fleming and co-authors (2007), we require the result of creative efforts to be both novel and useful. Novel technologies that do not show a minimum number of follow-ons are excluded from our analysis. We check for any possible bias from selecting only successful novel technologies (cf. Section 4.4 and Appendix E).

When characterizing the diffusion pattern of novel technologies, we follow the literature and look for S-shaped diffusion patterns (Griliches 1957; Dosi 1991). Initially, a technology diffuses slowly since it needs time to gain legitimation within the inventors' community. Inventors using the novel technology for their follow-on inventions need time to learn and familiarize with the novel technology and to abandon the established competing technologies. Once legitimated, the diffusion of the novel technology accelerates. The asymptotical convergence to a ceiling level reflects the full impact of a technology. Two different forces can lead a technology to reach its ceiling. First, all the possible applications of the technology have been implemented exhausting the inventive opportunities. Second, the technology loses its appeal in favor of emerging alternative technologies. A set of simple assumptions on the cost and benefits of adopting generates such an S-shaped curve. If the distribution of the benefits of adopting over its users is distributed approximately normally and the cost of adopting is constant or declines monotonically over time, the diffusion curve will have the familiar S-shape (Hall 2004).

An insightful example of technology experiencing a first phase of legitimation followed by a second phase of convergence to its full technological impact is represented by the Sulpholamide drugs. The Sulpholamide is a class of synthetic antibacterial drugs introduced in 1935. The first Sulpholamide-based drug, Prontosil, was developed and launched by Bayer. In the next 10 years following the Prontosil introduction, other companies and public research laboratories gradually developed and marketed more than 5,000 innovative Sulpholamide-based drugs. The success of Sulpholamide-based drugs was exhausted by the mid- '50s for two reasons. First, additional improvements of the therapeutic benefits of this class of drugs become more difficult reducing the opportunities to generate other inventions. Second, new antibiotics substituting the Sulpholamide-based drugs appeared (for details on the case, see Achilladelis 1993).

The values of the length of the legitimation period combined with the level of full technological impact define the shape of the S-curve of the novel technology. The shape of the S-curve can, however, look considerably different for different inventions. Differences in time to legitimation and in the maximum level of technological impact will generate different diffusion patterns. Figure 1 compares the diffusion curves of a technology having a long legitimation phase and a high technological impact with another technology having a short legitimation phase but a low technological impact.

Figure 1: Comparison between different S-shaped diffusion curves



2.3. Characteristics of the initial invention as predictors for its diffusion pattern

What explains the speed and extent of diffusion of a novel technology? Why does one novel technology take a long time before take-off while others take-off very quickly? Why do some novel technologies impact a large total number of follow-on inventions, while others only a few? Unlike most of the diffusion literature, we will not look at the characteristics of the "users" to explain differences in diffusion patterns but, taking advantage of observing a large set of technologies at the time of their appearance, we look at the antecedent characteristics of the novel technologies as predictors for their diffusion pattern.

Rogers in his review of diffusion studies (1983) provides a useful characterization of the attributes of the initial invention embedding the novel technology that influences its potential adoption: the complexity of the invention, the uncertainty surrounding the evaluation of the novel invention assessment, the compatibility with the potential adopter's current way of doing things and the relative advantage of the novel invention over currently available alternatives.

To focus our search for antecedent characteristics, we assume that the characteristics of the components combined for the first time in the initial invention are critical features affecting the diffusion of the embedded novel technology (Fleming 2001). Recent studies look at these antecedents for explaining the probability of observing those new combinations in the first place (Curran 2013; Caviggioli 2016). Nevertheless, none of the previous studies consider these antecedents as predictors of diffusion of the novel technology.

We focus on the technological and cognitive characteristics of the newly combined components as relevant for the diffusion of the novel technology. Technological characteristics are those related exclusively to the technological aspects of the newly combined components. As technological characteristics, we consider both the technological similarity and science-based nature of the newly combined components. Cognitive characteristics are those related to the ease of understanding of the newly combined components by the follow-on inventors' community. As cognitive characteristic, we consider the familiarity of the inventors' community with the newly combined components.

Newly combining similar technologies

Novel technologies can result from the combination of components that differ substantially or that are similar to each other. When the initial invention is for the first time combining *similar components*, it will have lower levels of uncertainty on the costs and benefits from making this new match, compared to those initial inventions that combine for the first time components that are dissimilar. Lower uncertainty decreases the cost of re-using, leading to shorter legitimation times. In addition, the adoption of a novel technology which combines dissimilar components requires cross-field competence. Inventors may need to set up new teams to be able to re-use the novel technology. Since teaming-up and acquiring new competences are time consuming activities, we expect that a novel technology resulting from the new combination of similar components is legitimated earlier than a novel technology resulting from the new combination of dissimilar components.   Although the higher risk associated with combining dissimilar components may also reduce full technological impact, we expect that the combination of components that are based on dissimilar technological principles entail a higher potential for substantial new added value compared to alternatives, boosting their technological impact once the novel technology has passed the legitimation hurdle (Verhoeven et al. 2016).

Hypothesis 1 (newly combining similar components): A novel technology resulting from the new combination of two similar components (a) requires a shorter time to be legitimated and (b) has a lower technological impact.

Newly combining science-based components

Novel technologies can result from the new combination of components, which have a close science base. *Science-based components* are those components using basic research knowledge which has abstract content, in contrast with applied components where specialized knowledge is used for specific applications (Breschi et al. 2000). Novel technology newly combining more science-based components are more likely to be complex. The population of potential follow-on inventors needs to have the scientific knowledge to absorb the advancements embedded in the novel technology. This requires education and training of cohorts of inventors employed in industrial R&D (Klevorick et al. 1995). On the contrary, when newly combining applied components, the inventors rely on a stable set of knowledge that is immediately available, needs less knowledge updating and carries less uncertainty. In view of the higher complexity and uncertainty, we expected that a novel technology resulting from the combination of science-based components takes more time to be legitimated in the users-inventors' community. This higher uncertainty surrounding more science-based inventions may reduce the likelihood with which the maximum potential of inventor-users is reached. At the same time, the generality of the scientific knowledge used in science-based components is expected to open up a broader set of technological opportunities (Klevorick et al. 1995). Novel technologies relying on applied components are expected to have a more certain, but limited set of possible technological applications. For these reasons, we expect that a novel technology resulting from the combination of science-based components has a higher expected technological impact while the combination of applied components has lower technological impact.

Hypothesis 2 (newly combining science-based components): A novel technology resulting from the combination of science-based components (a) requires a longer time to be legitimated and (b) has a higher technological impact.

Newly combining familiar components

Cognitive characteristics are relevant since novel technologies are adopted by follow-on inventors who need to receive and interpret information before being able to use it in a creative way (Cohen and Levinthal 1990). Re-use is more likely when the novel technology is compatible with the potential adopter's current way of doing things (Rogers 1983). For these reasons, we look at how familiar the inventors' community is with the components of the novel technology. A novel technology can result from the new combination of components that have been frequently or rarely used within the user-inventors' community. A component frequently used is well-known and can be readily exploited, while a component rarely used is unknown and still has to be explored (March 1991). Following Fleming's argument (2001), we assume that there is a greater chance that inventors build their inventions using components they are familiar with and can do this faster. Exploiting familiar components reduces the risks for unexpected results and facilitates the legitimation of the resulting novel technology. However, familiar components have been already largely exploited in the past so the novel technology resulting from their new combination is expected to have a lower technological impact.

Hypothesis 3 (newly combining familiar components): *A novel technology resulting from the combination of familiar components (a) requires a shorter time to be legitimated and (b) has a lower technological impact.*

All three hypotheses predict a trade-off between technological impact and legitimation: novel technologies which have characteristics that can ensure a high potential are more likely to have a longer legitimation time.

3. DATA AND METHODS

3.1. Identification of novel technologies

Our analysis relies on a sample including all the patents filed to the European Patent Office (EPO) in the period 1985-2015. Following Fleming and his coauthors (2007), we consider the technology classes in which a patent is classified (International Patent Classification -IPC- codes) as a proxy for the technological components which the patent uses (Schmoch 2008). We mark the appearance of a novel technology as the first time ever[4] appearance of a combination of IPC codes in a patent. Limiting our definition of novel technology to the first time ever appearance of a combination, we ignore novel inventions arising within a given component, not from combinations. We also exclude as novel technologies, pairs of IPC codes that appear for the first time but that are not re-used after their appearance[5]. Following Amabile (1996) and Fleming and co-authors (2007), we require the novel technology to be useful, as we need a sufficient number

[4] To flag a technology as novel, we need the complete history of its component to be able to evaluate when a pair of components appear together for the first time. We use the first available data period at EPO, 1978-1984, as a buffer period to capture the history of our components and we track novel combinations starting from 1985.

[5] The generation of pairs of IPC codes that are not re-used might be the result of some random error, for instance from the mechanical merge of all IPC codes included in all the EPO patent applications.

of followers to trace robust diffusion paths. Therefore, we restrict our novel technology definition only to those novel combinations re-used in at least 20 patents in the 20 years following the novel technology appearance. As noted by Jaffe (2002), “such selection does not create any selectivity bias” (pp. 25), as we are estimating the effects of our main variables of interest for successful novel technologies.

3.2. Identification of follow-on inventions

We trace the diffusion of a novel technology through its re-use by follow-on patents. To this end, we identify the pool of follow-on inventions as those later patents which also use the new combination of technology classes, introduced by the initial patent. Interestingly, many of the re-users are not necessarily citing the initial invention, reminiscent of the patent citation link to be an imperfect proxy of technology relatedness (cf. Appendix A).

A possible concern with our identification of a “technological trajectory” of a novel technology by tracing a new combination of two different components (represented by the combination of two IPC codes) is that there is no guarantee that the resulting combination’s re-use is meaningful. It could be that the novel combination is a mere artifact of our way of identifying novel technologies without any relevant content and that linking patents through a common use of a new combination is meaningless. To investigate the existence of a meaningful content, we implement a topic modelling analysis that verifies the coherence of the content of the patents embedding the same novel technology. If the patents associated to the same novel technology are coherent in terms of content, we can assume that they embed a meaningful common content, validating our approach for identifying the technology diffusion curves of a novel technology. In Appendix B we discuss how we validate our novel technology measure by applying a topic modelling algorithm. The results confirm that our method applied to identify the group of patents (re-)using a novel technology seems to link patents with coherent contents.

3.3. Characterizing the diffusion pattern of novel technologies

To represent the diffusion curves of each novel technology we proceed in two steps. First, we count the yearly number of patents that re-use the novel technology (y_t). We construct the actual cumulated distribution over a window-period of 20 years ($Y_t = \sum_{p=1}^{t} y_p$) following the appearance of the novel technology[6]. Second, we use the actual cumulated distribution of each novel technology to fit the corresponding trend function. The trend function represents an algebraic approximation of the diffusion curve of the technology. Following the literature, we opt for an S-curve characterized by a slow initial growth and by an asymptotic convergence to a ceiling level (Griliches 1957; Geroski 2000). The use of an estimated diffusion curve, instead of using the actual diffusion data, allows us to obtain a ceiling and legitimation period also for those technologies that do not exhaust their innovative potential during the observation period of 20 years. In fact, technologies with particularly slow legitimation might not show any asymptotic convergence to a ceiling level after the 20 years covered by the observed data. One example of a technology with a particularly long diffusion process is the light bulb that was introduced in 1909 and that continued to generate additional patented inventions until 1955 (Abernathy and Utterback 1978). Nevertheless, more than two third, sixty-eight percent, of the novel technologies included in our sample reaches the estimated ceiling level after 20 years[7]. Appendix C provides further evidence

[6] We consider all the novel technologies until the patent cohort of 1996 to leave a 20-year window forward to this last cohort.

[7] We consider a novel technology reaching the ceiling level if the difference between the actual cumulated number of patents after 20 years and the estimated value of the ceiling is less than 15% of the ceiling value.

on how well the S-shaped curve fits our actual data. We also check robustness of our results using the actual diffusion data in Appendix E.

The technological diffusion S-curve can be expressed by the following equation and identified by three parameters, i.e, *Midpoint*, *Alpha* and *Ceiling*:

$$\hat{Y}_t = \frac{Ceiling}{1+e^{\left(-\frac{(t-Midpoint)}{Alpha}\right)}} \qquad \text{(Equation 1)}$$

where $\hat{Y}_t$ is the cumulated number of re-using patents predicted at time *t*; *t* is the number of years elapsed since the appearance of the novel technology; the parameter *Ceiling* is defined as the upper asymptote of the S-curve; *Midpoint* is the required time to reach the fifty percent of the ceiling; and *Alpha* is the inverse of the curve slope at the *Midpoint*. An increase of the value of *Alpha* leads to flatter diffusion curves while a decrease of the value of *Alpha* leads to steeper diffusion curves.

We use the three parameters describing the S-curve as proxies of the concept of *Legitimation* and *Technological impact*. We consider a technology to be legitimated when the technology reaches the ten percent of its ceiling (Griliches 1957). *Legitimation* can be calculated as a linear combination of *Midpoint* and *Alpha*, *Midpoint-ln(9)*Alpha* (See Equation 1 and Appendix D for the mathematical details). It is expressed in number of years since the appearance of the novel technology until its legitimation (Griliches 1957). We measure the full impact of a technology as the maximum number of inventions that the novel technology is expected to generate. The value of the full technological impact is captured by the *ceiling* parameter. While our main results concentrate on *Legitimation* and *Technological Impact*, we extend our discussion of the characterization of the S-curve in section 4.3 when we also present results on *Alpha* and *Midpoint*, which allows to check further characteristics of the diffusion curve.

3.4. Determinants of the diffusion curve of novel technologies

We conduct a set of regression exercises aiming to estimate the two critical parameters characterizing the diffusion curve of our sample of novel technologies: *Legitimation* and *Technological Impact*. As pivotal determinants of these parameters, we look at the characteristics of the novel technology which is being re-used, more particularly the technological and cognitive characteristics of the newly combined components. Specifically, we estimate with an Ordinary Least Squares (OLS) the following two equations[8]:

$$\textbf{\textit{Legitimation}} = \beta_0 + \textbf{\textit{Component characteristics}}*\boldsymbol{\beta_1} + \textbf{\textit{Inventors' characteristics}}*\boldsymbol{\beta_2} + \textbf{\textit{Applicants' characteristics}}*\boldsymbol{\beta_3} + \textbf{\textit{Other controls}}*\boldsymbol{\beta_4} + \varepsilon \quad \text{(Equation 2)}$$

$$\textbf{\textit{Technological impact}} = \beta_0 + \textbf{\textit{Component characteristics}}*\boldsymbol{\beta_1} + \textbf{\textit{Inventors' characteristics}}*\boldsymbol{\beta_2} + \textbf{\textit{Applicants' characteristics}}*\boldsymbol{\beta_3} + \textbf{\textit{Other controls}}*\boldsymbol{\beta_4} + \varepsilon \quad \text{(Equation 3)}$$

Component characteristics is a vector of variables that includes the characteristics of the components in the new combination characterizing the novel technology. We consider the technological *Similarity* and *Science-based* nature of the newly combined components as well as the *Familiarity* of the follow-on inventors' community with the newly combined components. All

[8] We run also a set of regressions where we estimate Equation 2 and 3 simultaneously using a Seemingly Unrelated Regressions (SUR) method. However, having the same set of regressors in the two equations guarantees estimates that are as consistent and efficient as the estimation equation-by-equation using a standard OLS (Davidson and MacKinnon; 1993).

of these characteristics are measured at the time of the initial patent which introduces the new combination for the first time ever.

To measure *Similarity* between the newly combined components, we exploit the hierarchical structure of the IPC code classification where each additional digit denotes a higher degree of refinement of the technological classification. More precisely, we define two components as being similar when they have the first three digits of their IPC codes in common. To construct the variable *Science-based content*, we consider the patent applications which include any of the two combined components over a rolling time window from *t-1* to *t-4*. For each component, we compute the average number of references to the non-patent literature per patent application (Meyer-Krahmer and Schmoch 1998). We calculate the *Science-based content* variable as the average number of references per patent for the two components. We use Fleming (2001) to construct our measure for *Familiarity*. We count the number of patent applications in a four-year rolling window for each of the two combined components and calculate the average number of patents of the two components. In the regression model, we consider the logarithm transformation of the variables *Science-based content* and *Familiarity,* to interpret the estimated effects as semi-elasticities.

Several characteristics of the components might bias our estimations of the technological and cognitive characteristics. To correct for this, we include a series of controls. To control for time-invariant and technology specific unobserved characteristics of the components such as the unmeasurable propensity of the component to generate additional inventions and its technological complexity, we include as controls a set of three-digit technological class dummies referring to the combined components (*Dummy Technology class*). A technology class dummy equals one if the first three digits of the IPC codes of at least one of the two combined components equal to the three-digit identifying the technological class dummy, zero otherwise. We also include a set of time-dummies representing the calendar year when the novel technology appeared (*Technology entry year*).

Other characteristics that may influence the diffusion pattern of the novel technology relate to the characteristics of the applicants and/or inventors of the initial patent. As applicants' characteristics, we consider applicants' experience, type (university or public research center versus private company), being a single applicant, and location country[9]. The *applicants' experience* is computed as a count variable that records applicants' previous patented inventions. *University applicant* is a dummy that equals one if at least one of the applicant is a university or a research center, zero otherwise. *More than one applicant* is a dummy that equals one if there is more than one applicant, zero otherwise. *Applicant's country* is a set of dummies, one for each applicant's country reported on the patent documents, that identifies the geographical location of the applicant(s). As controls for inventors' characteristics, we consider *inventors' team size*. The team size is the number of inventors appearing in the patent filed during the first year when the novel technology appears, i.e. the initial patent.

4. DATA AND RESULTS

Before we present the descriptive results on our full sample in section 4.2 and the econometric results in 4.3, we illustrate the main idea of our paper through the case of a renowned invention and the diffusion of its embodied novel idea: the 'onco-mouse' embodying the new 'transgenic mammal technology'. The 'onco-mouse' provides an illustrative example of how we

[9] If there are multiple patents embodying the novel technology the first year when it appears, we aggregate the characteristics of these patents in a representative unique initial patent.

identify the diffusion curve of a novel technology and of how we calculate the characteristics of the combined components.

4.1 A first illustrative case: the “onco-mouse”

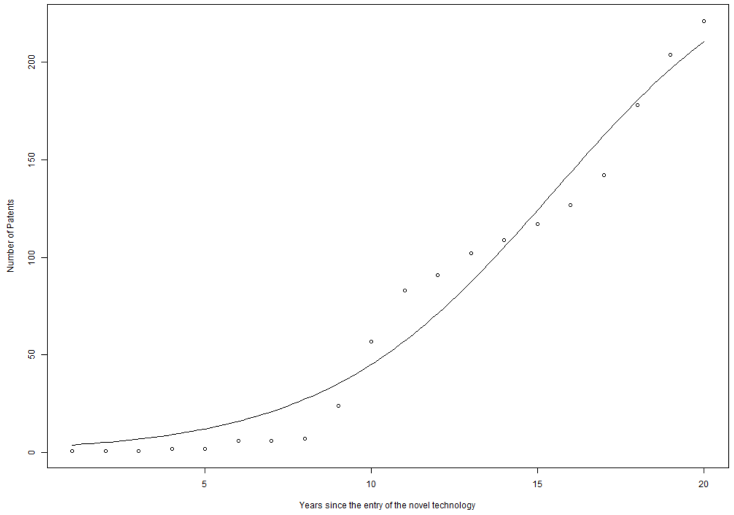
The “onco-mouse’’, a mouse widely used in laboratory cancer research, is an example of a well-recognized breakthrough invention embodying a novel technology. In the mid-80s, Philip Leder and Timothy Steward at Harvard laboratories had the revolutionary idea of isolating cancer-related genes and injecting them into a mouse egg transplanted into a female mouse generating the first transgenic mammal likely to develop a specific disease (Murray 2010). The novel technology to mimic diseases (transgenic mammal technology) uses an unprecedented combination of two existing technological components: “Gene isolation” and “Injection of material into animals”. The “onco-mouse” was the first patented invention embodying the novel technology for cancer, but a variety of follow-on inventions followed, represented by other transgenic mammals used in labs to experiment treatments for a variety of diseases. Geneticists patented at least 110 transgenic mammals designed to develop diseases from Alzheimer to cystic fibrosis (Murray 2010). For instance, the “diabetic mouse” created in 1996 by Seo Jeon Sun, a professor from Seoul National University is a follow-on patented invention building up on the novel transgenic mammal technology.

We look at the patents’ technological content and we mark as transgenic mammal novel technology the combination of the IPC codes “Introducing […] material into […] the body of animals” (IPC code A01K67) and “[…] DNA or RNA concerning genetic engineering […]” (IPC code C07H21) that appeared for the first time in the onco-mouse patent, characterizing its novelty. We reconstruct the diffusion curve of the transgenic mammal technology to mimic diseases, by tracing the re-use of the novel combination “Injection of material into animals” A01K67 and

C07H21 “Gene isolation” in follow-on patents. We observe the novel transgenic mammal technology for a 20-year window. After 20 years, we observe 220 patents re-using the A01K67-C07H21 novel combination of IPC classes introduced by the onco-mouse patent. It is interesting to note that among these 220 patents only six cite the original patent and only 34.5% cite at least another patent belonging to the same novel technology.

Figure 2 shows the actual cumulated distribution (Y_t) of the user patents embodying the combinations of the IPC codes A01K67-C07H21 (dotted line). Referring to the patent distribution, we estimate the three parameters of the corresponding S-curve by using a maximum likelihood estimation methodology. The solid line in Figure 2 represents the fitted S-curve curve ($\hat{Y}_t$). This curve is identified by an estimated *Ceiling* equal to 267.38 patents, a *Midpoint* of 15.49 years, and an *Alpha* of 3.45.

Figure 2: Diffusion curve of the novel onco-mouse technology



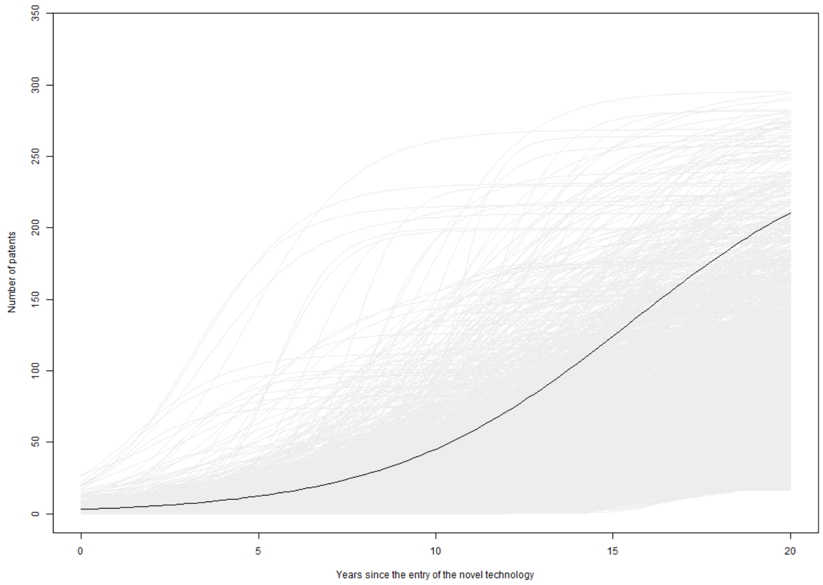
The use of a specific functional form allows us to predict the diffusion of a novel technology at any point in time by means of the three estimated parameters. The ceiling parameter of 267.38 patents corresponds to the full technological impact reachable by the technology. The legitimation of the “transgenic mammal technology, reaching the acceptance rate of 10% of its ceiling, is reached after 7.91 years since its appearance. In our example, by substituting the estimated parameters in Equation 1, we can predict that, after 10 years since the transgenic mammal technology appearance ($\hat{Y}_{10}$), the cumulated number of re-user patents embodying the transgenic mammal technology equals 45.24 ($\frac{267.38}{1+e^{(-\frac{(10-15.49)}{3.45})}}$=45.24). In the same way, we can predict that after 30 years from its appearance ($\hat{Y}_{30}$), the cumulated number of user patents embodying the transgenic mammal technology would be equal to 263.45 ($\frac{267.38}{1+e^{(-\frac{(30-15.49)}{3.45})}}$=263.45). The transgenic mammal technology is thus an example of a breakthrough novel technology with a high technological impact, but which took a rather long time for legitimation.

To explain this diffusion pattern of the transgenic mammal technology, our main hypotheses develop around the characteristics of the two technological components combined to generate the novel technology. Concerning the characteristics of the combined components generating the transgenic mammal technology, these components where strongly science-based (*Science-based content* = 9.51, almost 3 times higher than our sample average). Furthermore, the two components combined for the first time belong to two different technological domains, making them *dissimilar*. And finally, the “Gene isolation” component was rather new within the technological community, as well as the “Injection of material into animals” component at the time when they were recombined for the first time in the “onco-mouse” (*Familiarity=160.5*, which is less than half the sample average value). The transgenic mammal technology being a science-based, dissimilar and unfamiliar novel technology has a high technological impact and a long time to legitimation, confirming our hypotheses.

4.2. Some first descriptive results.

Our study sample includes 10,782 successful novel technologies (i.e. novel technologies with at least 20 follow-on patents in the following 20 years) which generated 249,103 distinct follow-on patents. We first present some descriptive results on the diffusion patterns for these 10,782 observations. Figure 3 plots all the estimated diffusion curves of the novel technologies in our sample. The black line highlights the "transgenic mammal technology" diffusion curve. Table 1 reports the descriptive statistics on the estimated parameters of the diffusion curves.

Figure 3: Novel technologies' estimated diffusion curves.



Note: The black curve identifies the onco-mouse technology

Table 1: Descriptive statistics on the estimated diffusion curves.

	Obs.	Mean	Sd	Median	Min	Max
Estimated parameters:						
<u>*Dependent variables*</u>						
Technological impact/Ceiling [# patents]	10,782	66.95	94.58	38.2	20	998.29
Legitimation (10%) [# years]	10,782	5.91	3.13	5.67	0	16.78

Other sigmoid parameters						
Midpoint (50%) [# years]	10,782	12.45	3.29	12.39	1.53	24.89
Alpha	10,782	2.98	1.16	2.88	0.21	7.47

Both Figure 3 and Table 1 illustrate the heterogeneity in diffusion patterns of novel technologies in our sample. On average, a typical diffusion pattern of novel technologies is one of a modest technological impact, as the median *Ceiling* is about 38 follow-on patents in a 20-year window and the average *Ceiling* for a novel technology about 67 patents. However, the sample contains a high variance on technological impact, with a right skew and substantial outlier cases, as Figure 3 shows. With its 267 patents at *Ceiling*, the transgenic mammal technology is one such clear outlier. At the same time, the average/median *Legitimation* is less than 6 years to reach the 10% threshold, substantially lower than the 8 years for the transgenic mammal case. The variables *Ceiling* and *Legitimation* are positively correlated (0.37) meaning that on average a higher technological impact (*Ceiling*) tends to correspond with a longer *Legitimation* time.

Table 2: Descriptive statistics on the component characteristics curves.

	Obs.	Mean	Sd	Median	Min	Max
Component characteristics						
Similarity 3 digits (dummy)	10,782	0.33	0.47	0	0	1
Science-based content	10,782	3.5	7.45	2.25	0.02	283.99
Familiarity	10,782	347.02	429.17	208.5	0.5	5,490.5
Inventor's characterisics						
Inventors' team size	10,782	3.36	2.9	3	1	55
Applicant characteristics						
Applicants' experience	10,782	638.2	1,502.61	33	0	17,279
University applicant (dummy)	10,782	0.05	0.22	0	0	1
More than one applicant (dummy)	10,782	0.25	0.43	0	0	1

Table 2 shows the component characteristics of the novel technology. About one third of our sample of novel technologies newly combines similar components. On “familiarity”, the sample displays a left skew with most observations having a lower than average familiarity of its newly combined components (median smaller than mean). For "science-based content", the sample contains substantial heterogeneity (high standard deviation), also with a left skewed distribution: most observations have a relatively low science-based content of its newly combined components, while there are a few outliers with high science-based content (median smaller than mean).

4.3. Main econometric results

Table 3 shows the results of our econometric exercise, estimating equations 1 and 2. Columns 1 and 2 report the estimation of a baseline model including only the component characteristics and technology entry year dummies. Columns 3 and 4 add as controls the inventors' and applicants' characteristics. Columns 5 and 6 add the technological class dummies. Comparing columns 3 and 4 with columns 5 and 6 shows the importance of adding technology classes as controls. According to our most complete model specification, estimated in columns 5 and 6, we find that combining two similar components reduces significantly the legitimation time. Specifically, a novel technology resulting from the new combination of two similar components has a legitimation time which is 14.4 months[10] shorter than a novel technology resulting from the new combination of two dissimilar components. Compared to the average sample legitimation time of 5.91 years, this implies a 20% longer legitimation time. Component similarity also significantly affects the ceiling, leading to lower levels of full technological impact. Having similar components decreases the ceiling by 8.58 patents. Compared to the sample average of 67 ceiling

[10] 14.4 is obtained multiplying the coefficient of *Similarity* in the regression with *Legitimation* as dependent variable (-1.20) by the number of months in a year (12).

patents, this is a 13% difference. With both results, our first hypothesis on similarity is confirmed, with sizeable effects.

Newly combining components with a higher science-based content generates a novel technology that requires more time to be legitimated. This effect is significant, but not substantial: novel technologies with a 50% higher score on science-based content, have a 0.96 months longer legitimation time (1.4% difference relative to sample average). At the same time, novel technologies combining science-based components have a significantly higher ceiling. A 50% higher score on science-based content increases the ceiling by 8.3 patents, all else equal. These results confirm our second hypothesis on combining science-based components.

A novel technology resulting from the new combination of two familiar components requires a shorter time to be legitimated. Increasing the level of familiarity by 50% decreases the time needed for a novel technology to be legitimated by 1.2 months. The same technology has lower technology impact, with a ceiling smaller by 4.6 patents. These results are in line with our third hypothesis, combining familiar components, be it with small sized effects.

Concerning the controls, we find that the inventors' team size, as well as the experience of the applicants, having multiple applicants, and the presence of a university among the applicants lead to a novel technology with a higher ceiling. The legitimation time is negatively affected by the presence of multiple applicants.

Table 3: Regression results. OLS estimations for Equation 1 and 2.

	(1)	(2)	(3)	(4)	(5)	(6)
	Legitimation	Technological	Legitimation	Technological	Legitimation	Technological
VARIABLES	(10%)	Impact/Ceiling	(10%)	Impact/Ceiling	(10%)	Impact/Ceiling
Component characteristics						
Similarity 3 digits (dummy)	0.066	19.1***	0.11*	18.6***	-1.20***	-8.58***
	(0.066)	(2.00)	(0.065)	(2.00)	(0.10)	(3.22)
log(Science-based content)	0.24***	19.9***	0.30***	18.6***	0.16***	16.6***
	(0.032)	(0.98)	(0.034)	(1.03)	(0.048)	(1.54)
log(Familiarity)	-0.27***	-9.25***	-0.25***	-9.43***	-0.20***	-9.13***
	(0.029)	(0.88)	(0.029)	(0.88)	(0.029)	(0.92)
Inventor’s characteristics						
log(Inventors' team size)			-0.32***	1.14	-0.12***	5.19***
			(0.047)	(1.45)	(0.045)	(1.43)
Applicant characteristics						
log(Applicants' experience)			-0.026**	0.40	-0.014	0.80**
			(0.011)	(0.35)	(0.011)	(0.35)
University applicant (dummy)			-0.047	3.79	0.12	8.60**
			(0.13)	(4.14)	(0.13)	(4.08)
More than one applicant (dummy)			-0.75***	8.17***	-0.87***	5.43*
			(0.095)	(2.90)	(0.088)	(2.83)
Dummy Technology class (3 digits)	No	No	No	No	Yes	Yes
Dummy Applicant’s country	No	No	Yes	Yes	Yes	Yes
Dummy Technology entry year	Yes	Yes	Yes	Yes	Yes	Yes
Constant	7.61***	117***	8.34***	99.6***	10.6***	146***
	(0.17)	(5.10)	(0.19)	(5.78)	(0.25)	(7.90)
Observations	10,782	10,782	10,782	10,782	10,782	10,782
R-squared	0.062	0.060	0.099	0.076	0.240	0.148

In Table 4 we further characterize the diffusion curves of novel technologies, reporting a set of regressions with *Midpoint* and *Alpha* as dependent variables. While full technology impact has a direct correspondence to the ceiling parameter of a sigmoid curve, legitimation is a linear combination of the remaining two parameters, *Midpoint*, and *Alpha*. The coefficients estimated for each variable can be calculated as follow (cf. Appendix B):

$$\hat{\beta}_{Legitimation} = \hat{\beta}_{Midpoint} - \hat{\beta}_{Alpha} * 2.2 \qquad \text{(Equation 4)}$$

Two novel technologies with a similar time to legitimation (reaching 10% of their ceiling) can have different times to midpoint (reaching 50% of their ceiling) when they have a different *Alpha* (i.e. the inverse of the slope at midpoint). The *Alpha* parameter reflects the time it takes from legitimation to midpoint, i.e. to go from 10% to 50% of the ceiling. Looking at Equation 4, we can see that a higher *Alpha* leads to a higher difference in time from legitimation to midpoint and therefore a slower diffusion between legitimation and midpoint. The *Midpoint* results in Table 4 allow to check our results for sensitivity in the time to various shares of the ceiling value reached (10% versus 50%) while the *Alpha* results allow to look at the time from Legitimation (10%) to Midpoint (50%).

Table 4: OLS estimation using as dependent variables Midpoint and Alpha and adopting the same specifications reported in Equation 1 and Equation 2.

	(1)	(2)
VARIABLES	Midpoint (50%)	Alpha
Technological component characteristics		
Similarity 3 digits (dummy)	-0.47***	0.33***
	(0.098)	(0.037)
log(Science-based content)	-0.26***	-0.19***
	(0.047)	(0.018)
log(Familiarity)	-0.14***	0.028***
	(0.028)	(0.011)
Inventor's characterisics		
log(Inventors' team size)	-0.11***	0.00032
	(0.044)	(0.016)
Applicant characteristics		
log(Applicants' experience)	-0.033***	-0.0087**
	(0.011)	(0.0040)
University applicant (dummy)	-0.31**	-0.20***
	(0.12)	(0.047)
More than one applicant (dummy)	-0.29***	0.26***
	(0.086)	(0.032)
Dummy Applicant's country	Yes	Yes
Dummy technology entry year	Yes	Yes
Dummy Technology Class (3 digits)	Yes	Yes
Constant	16.7***	2.76***
	(0.24)	(0.090)
Observations	10,782	10,782
R-squared	0.342	0.262

The results concerning *Legitimation* (time to reach 10%) and *Midpoint* (time to reach 50%), appear consistent (see Table 3 column 5 versus Table 4 column 1). Indeed, "Similarity" and

"Familiarity" show qualitatively the same impact on *Legitimation* and *Midpoint* with less time needed to reach not only 10% (legitimation), but also 50% (midpoint) of the ceiling. Their significant positive effect on *Alpha* signals that similar or familiar novel technologies, take a longer time to go from *Legitimation* to *Midpoint* as compared to dissimilar or non-familiar novel technologies. This suggests that the speed advantage associated with similarity or familiarity matters particularly in the initial phase of diffusion, i.e. before *Legitimation.* But after legitimation, dissimilar and familiar novelty are able to catch up, but only to a limited extent. For "Science-based content", we find different results for *Legitimation* and *Midpoint*. While a higher science-based content of the combined components slightly increases the legitimation time, it decreases the time needed to reach the *Midpoint*. The results on the parameter *Alpha* explains this finding. A high science-based content of the combined components leads to a diffusion curve that grows slowly in its very first part (as shown by the positive coefficient in the *Legitimation* regression), but after this first part, catches up quickly, due to its steepness (as shown by the negative coefficient in the *Alpha* regression), thus reaching the midpoint faster (as shown by the negative coefficient in the *Midpoint* regression). These results, therefore, show that the science-based content of the novel technology retards only the very initial diffusion speed before legitimation.

To visualize the impact of the component characteristics on the novel technology diffusion, we consider in Figures 4 and 5 three cases of novel technologies, which we compare with a baseline novel technology characterized by dissimilar components, a low science-based content (equivalent to the first quartile (Q1) value in our study sample), and a low familiarity value (Q1 value in our study sample)[11]. The three cases are: 1) the novel technology differs from the baseline for having

[11] To characterize the baseline technology, we assign to the remaining control variables some representative values as follow: log(Inventors' team size)=log(3); log(Applicants' experience)=log(38); University applicant=0;

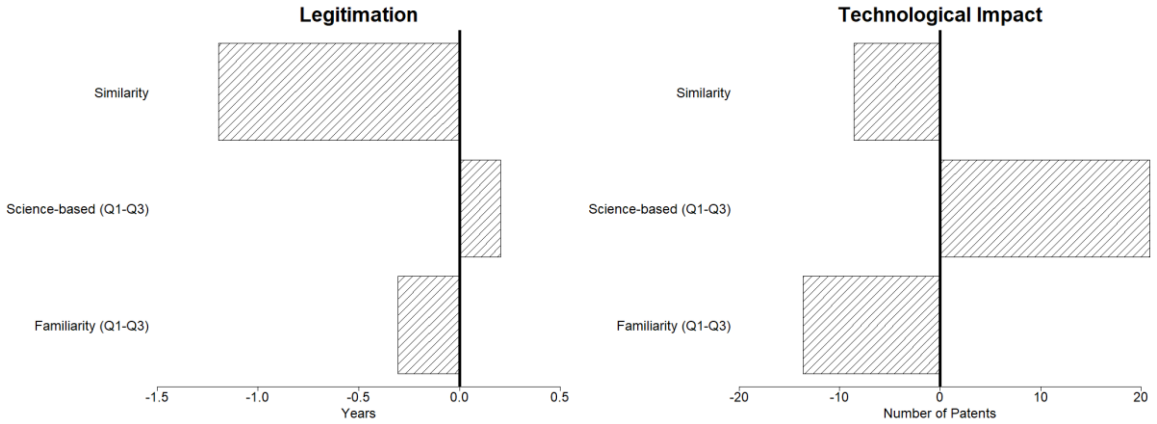
similar components; 2) the novel technology differs for being highly science-based (Q3 value); and 3) the novel technology differs for having a high familiarity (Q3 value). The three cases show a trade-off when changing the values of each component characteristics (Similarity, Science-based content, Familiarity). In building the different cases, we use the estimated coefficients of Table 3, columns 5 and 6.

Looking at Figure 4, we observe that by augmenting the value of the variable "Science-based content" from Q1 to Q3, the corresponding novel technology will have a higher ceiling. However, it will require a longer time to be legitimated. On the contrary, augmenting the value of "Familiarity" from Q1 to Q3 or having a "similar" novel technology leads to a shorter legitimation time, but, as a drawback, leads to a lower ceiling.

Although each component characteristic shows a trade-off, the direction and the magnitude of these trade-offs differs significantly according to the characteristic considered. Augmenting the "Science-based content" increases to a considerable extent the ceiling, while augmenting the legitimation time, but only slightly. Both "Familiarity" and "Similarity" reduce the legitimation time, but at a cost of a lower ceiling. While for "Familiarity" the gain in legitimation time is only modest and the loss in ceiling is substantial, the gain in legitimation time for "Similarity" is much more substantial, while its loss in ceiling is more modest.

More than one applicant=0; Dummy Applicant's country (US=1, zero all the other country dummies); Dummy Technology class (A01=1 and C07=1, zero all the other technology class dummies); Dummy technology entry year (Dummy 1985=1).

Figure 4: Variation in novel technologies' legitimation time and ceiling when changing the values of the technological component characteristics.



The legitimation time of the baseline technology equals to 5.56 years; The average number of ceiling patents is 45.70

4.4. Robustness checks

To test the reliability of our results we implemented a set of robustness checks which are reported in Appendixes E - H.

The assumption that the diffusion process follows an S-curve might raise concerns. We report in Appendix E a variant of the econometric exercise of Table 3, using as dependent variables the actual *Legitimation* time and the actual *Technological Impact*, rather than the estimated values from fitting an S-curve. Specifically, we calculate the legitimation time as the time required to reach the ten percent of the actual *Ceiling*. We define the actual *Ceiling*, as the cumulated number of patented inventions using the novel technology 20 years after its appearance. Results are consistent with those reported in Table 3 when using actual *Legitimation* time and actual *Technological Impact*.

Another concern with our S-curve approach is that we force the diffusion pattern into a 20-year window. For those novel technologies which peak after 20 years, the maximum value within the 20-year window will not coincide with their true ceiling. Figure F1 in Appendix F plots the actual cumulated number of patents at the end of our observation period (20 years) against the estimated technological impact (the *Ceiling* of the S-curve). It shows that although about 32% of the cases do not reach their *Ceiling* (+/- 15%) at the end of the 20-year period, most of the deviations are small in size. Overall, the 10,782 novel technologies considered in our study are at 88.5% of their ceiling value at the year 20. Nevertheless, in Appendix F we rerun our econometric exercises on two sub-samples. The first sample includes only novel technologies reaching their full technological impact within the 20 years observed, while the second sample includes the technologies not reaching their full technological impact within the 20-year period. Results from both sub-samples are similar and consistent with those obtained in Table 3. Only for *Similarity* the negative effect on *Technological Impact* disappears in the second subsample.

In the main analysis, reported in Table 3, we consider the diffusion of successful novel technologies and, as a consequence, we limit our observations to all the novel combinations that reach at least 20 follow-on inventions within the 20-year window considered. Excluding novel technologies that do not or only marginally diffuse may generate a selection bias problem. To correct for this selection bias we adopt a 2-step Heckman estimation strategy. Appendix G reports the results of our estimations without and with the correction for the selection bias. We find a substantial coherence between both results. This allows us to claim that analyzing successful novelty and restricting our sample to 10,782 novel combinations does not introduce any serious selection bias in the estimated impact of the characteristics of the combined components on diffusion.

5. Conclusion

Despite the high interest of scholars in identifying successful inventions and their diffusion, little attention has been devoted to investigate how the novel ideas embodied in original inventions are re-used in follow-on inventions and how these diffusion patterns are affected by the antecedent characteristics of the novel technology.

We address these limitations by using patent data to empirically map and characterize the trajectory of novel technologies' re-use in follow-on inventions on a large scale. We identify a novel technology as an unprecedented combination of existing technological components and trace all the patents who are re-using this new combination in their follow-on inventions. We analyze how the full technological impact reached by the novel technologies and the time needed to be legitimated, are affected by the antecedent characteristics of the novel technology, more specifically, the science-based nature and similarity of the newly combined components as well as the familiarity of the follow-on inventors' community with the newly combined components.

Our large scale empirical study, using patent data from 1985 till 2015, allows to trace the diffusion of 10,782 successful novel technologies which generated 249,103 distinct follow-on patents. Using the common S-shaped diffusion curve, we obtain an estimated ceiling and legitimation period for each novel technology. Our results suggest a typical diffusion pattern of relatively fast legitimation but modest technological impact. However, the sample contains a high variance on technological impact, with a right skew and substantial outlier cases, the "onco-mouse" being one such example. Most high outliers have a longer than average legitimation time, like the transgenic mammal technology. In general, we find that higher ceiling values are positively correlated with longer legitimation time.

Our analysis on antecedent characteristics of the novel technology associated with differences in diffusion patterns shows that a high degree of similarity between the newly combined components and a high familiarity of inventors with the newly combined components shorten the legitimation time, but also reduce the technological impact. The science-based nature of the combined components plays an opposite role, increasing the technological impact of a novel technology, but extending the legitimation time. Although each component characteristic shows a trade-off between legitimation time and technological impact, the direction and the magnitude of these trade-offs differs significantly according to the characteristic considered. Augmenting the science-based content of the newly combined components improves to a considerable extent the full technological impact, at a cost of only slightly augmenting the legitimation time. In fact, soon after the legitimation time, the speed of diffusion takes up, such that mid-term is reached faster than for non-science based novel technologies. Augmenting familiarity leaves only a modest gain in legitimation time, while the loss in technological impact is substantial. Newly combining similar rather than dissimilar components leaves a substantial gain in legitimation time, while the loss in technological impact is more modest.

Our analysis contributes in several ways to the literature, improving our understanding of the diffusion trajectory of novel technologies. First, we identify the diffusion trajectory of novel technologies carefully. Rather than using the classification in the same technology class as link between the users, or a patent citation link, we identify "users" as those follow-on inventions which are re-using the new combination which characterized the initial novel technology. A topic modelling validation analysis confirmed that this re-use link captures a content link, while it poorly correlates with patent citation links. Second, as we are using patent information, we can identify diffusion trajectories for a large sample of novel technologies. While the literature so far has mostly looked at the diffusion of specific cases of technologies and concentrated on characterizing the adopters, our large scale approach allows to identify the characteristics of the novel technologies as determinants of the diffusion patterns.

Our approach leaves some important findings. Perhaps the finding with the most important implications is the apparent trade-off between legitimation time and technological impact. Those novel technologies with bigger technological impact seem to need a longer time before they are legitimatized. Particularly the riskier types of novel inventions, i.e. those that newly combine dissimilar, unfamiliar and science-based components, while having a larger technological impact, require a longer legitimation time. This raises questions for further analysis on whether the higher technological impact that these riskier novel inventions generate, could have been reached with shorter legitimation. What causes a longer legitimation time for bigger impact novel inventions? Can re-use be speeded up for these “hits”? And what kind of policy intervention could help? Could we learn from the exceptional cases in the sample that became big hits fast?

Our research opens up other areas of future research. Other determinants of the diffusion patterns could be looked at, such as the characteristics of the incumbent technology areas affected by the novel technology. Our current analysis showed the importance of controlling for technology areas, but we would like to know which characteristics matter, such as its concentration in few incumbent users or its diversity of different types of incumbent users. Characterizing the “re-users” is another interesting area of further research. Who are the applicants and inventors in the follow-on inventions and how close are they to the original novel technology and to each other? And finally, although the S-shaped diffusion pattern fitted well the sample of novel technology diffusion trajectories on average, there are nevertheless outlier cases which could be interesting to study in more detail.

References

Abernathy, W.J. and Utterback, J.M. 1978. Patterns of Industrial Innovation. ***Technology Review,*** 80: 40–47.

Achilladelis, B. 1993. The dynamics of technological innovation: The sector of antibacterial medicines. ***Research Policy***, *22*: 279–308.

Amabile, T. M. 1996. Creativity in Context. Boulder: Westview Press.

Andersen, B. 1999. The hunt for S-shaped growth paths in technological innovation: a patent study. ***Journal of evolutionary economics*** 9.4 (1999): 487-526.

Andrews, D., Criscuolo, C., & Gal, P. 2015. Frontier firms, technology diffusion and public policy: micro evidence from OECD countries (Vol. 2). OECD Publishing.

Arthur, W. B. 2009. The Nature of Technology: What It Is and How It Evolves. Simon and Schuster.

Arts, S., Appio, F. P., & Van Looy, B. 2013. Inventions shaping technological trajectories: do existing patent indicators provide a comprehensive picture? ***Scientometrics***, *97*: 397–419.

Breschi, S., Malerba, F., & Orsenigo, L. 2000. Technological Regimes and Schumpeterian Patterns of Innovation. **The Economic Journal**, 110(463), 388–410.

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

Curran, C.-S. 2013. The Anticipation of Converging Industries. London: Springer London.

Caviggioli, F. 2016. Technology fusion: Identification and analysis of the drivers of technology convergence using patent data. ***Technovation***, 55–56.

Dosi, G. 1982. Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. ***Research Policy***, *11*:147–162.

Dosi, G. 1991. The research on innovation diffusion: An assessment. In Diffusion of technologies and social behavior (pp. 179–208). Springer.

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

Fleming, L., Mingo, S., & Chen, D. 2007. Collaborative brokerage, generative creativity, and creative success. ***Administrative Science Quarterly***, *52*: 443–475.

Geroski, P. A. 2000. Models of technology diffusion. ***Research Policy***, *29*: 603–625.

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. **Proceedings of the National Academy of Sciences**, 101(suppl 1), 5228–5235.

Griliches, Z. 1957. Hybrid Corn: An Exploration in the Economics of Technological Change. ***Econometrica***, *25*: 501.

Hall, B. H. 2004. Innovation and diffusion. Cambridge, Mass.: National Bureau of Economic Research.

Jaffe, A. 2002. Building Programme Evaluation into the Design of Public Research-Support Programmes, ***Oxford Review of Economic Policy***, 18 (1): 22–34.

Klevorick, A. K., Levin, R. C., Nelson, R. R., & Winter, S. G. (1995). On the sources and significance of interindustry differences in technological opportunities. **Research Policy**, 24(2), 185–205.

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

Meyer-Krahmer, F., & Schmoch, U. 1998. Science-based technologies: university–industry interactions in four fields. ***Research Policy***, 27 835–851.

Murray, F. 2010. The onco-mouse that roared: Hybrid exchange strategies as a source of distinction at the boundary of overlapping institutions. ***American Journal of Sociology***, 116(2), 341–388.

Nelson, R. R., & Winter, S. G. 1982. ***An evolutionary theory of economic change***. Cambridge, Mass.: Harvard Univ. Press.

Rogers, E. M. 1983, The Diffusion of Innovations, 3rd Edition, New York: The Free Press.

Rosenberg, N. 1976. Perspectives on technology. CUP Archive.

Rosenberg, N. 1982. Inside the black box: technology and economics. Cambridge University Press.

Schumpeter, J. A. 1939. ***Business cycles***. New York: McGraw-Hill.

Schmoch, U. 2008. Concept of a Technology Classification for Country Comparisons. Final Report to the World Intellectual Property Organization (WIPO), Fraunhofer Institute for Systems and Innovation Research, Karlsruhe.

Strumsky, D., & Lobo, J. 2015. Identifying the sources of technological novelty in the process of invention. ***Research Policy***, *44*: 1445–1461.

Stoneman, P., & Battisti, G. ,2010. The diffusion of new technology. In Handbook of the Economics of Innovation (Vol. 2, pp. 733-760). North-Holland.

Verhoeven, D., Bakker, J., & Veugelers, R. 2016. Measuring technological novelty with patent-based indicators. ***Research Policy***, *45*: 707–723.

Appendix A: Cross-patent citations in the same novel technology diffusion curves

Table A1 and A2 show how patents belonging to the same novel technology diffusion curve cite each other. On average, a negligible share of patents in the diffusion curve, 2.7%, cites the first patent including novelty. Only 27% of the patents cite other patents in the same diffusion curve, while 29% are cited by other patents. The limited number of cross-citations among patents in the same diffusion curve shows that citations capture only partially the re-use of a specific technology by technologically related inventions.

Table A1: Descriptive statistics.

	Obs	Mean	Sd	Min	Q1	Q2	Q3	Max
Patents in the diffusion curve	10,782	57.4	80.78	20	25	33	55	1004
Patents in the diffusion curve citing the first patent including the novelty	10,782	1.13	2.63	0	0	0	1	48
Patents in the diffusion curve cited by other patents in the diffusion curve	10,782	16.61	27.91	0	5	9	16	454
Patents in the diffusion curve citing other patents in the diffusion curve	10,782	18.3	32.07	0	5	9	18	516
Share of patents in the diffusion curve citing the first patent including the novelty	10,782	0.027	0.06	0	0	0	0.03	0.68
Share of patents in the diffusion curve cited by other patents in the diffusion curve	10,782	0.27	0.13	0	0.17	0.25	0.35	0.85
Share of patents in the diffusion curve citing other patents in the diffusion curve	10,782	0.29	0.15	0	0.18	0.27	0.38	0.92

Appendix B: Validation of the novel technology measure with topic modelling

A possible concern with the identification of a novel technology as the combination of two different components (represented by the combination of two IPC codes) is that there is no proof that the resulting combination is meaningful. It could be that the novel combination is a mere artifact of our way of identifying novel technologies without any meaningful content. To investigate the existence of a meaningful content, we implement a topic modelling analysis that verifies the coherence of the content of the patents embedding the same novel technology. If the patents associated with the same novel technology are coherent in terms of content, we can assume that they embed a meaningful common content, i.e. the novel technology.

First, we collected the titles of all the patents included in our study sample (250,979 patents. We inferred the topics of each patent by using the Latent Dirichler Allocation (LDA) method. The idea behind LDA is that each document is the results of a set of (latent) topics. Topics contain words with a probability distribution (Griffiths and Steyvers, 2007). For example, for a document on the topic "transgenic modification methods", the words that are likely to appear in the document will be "protein", "human", "dna", "transgenic". We used Gibbs's algorithm to estimate the LDA parameters and to classify each patent according to 20 topics. We chose the number of topics according to the optimization method proposed by Griffiths and Steyvers (2007) which aims at finding the number of topics that maximizes the log-likelihood value of Gibb's estimation of the LDA parameters. The results of this optimization method are reported in Figure B1. Figure B1 shows that the maximum value of log-likelihood is obtained for 20 topics. Figure B2 illustrates the most likely four words generated by each of the 20 topics.

Figure B1. Optimal topic selection number of topics identifying the patent titles text in our sample.

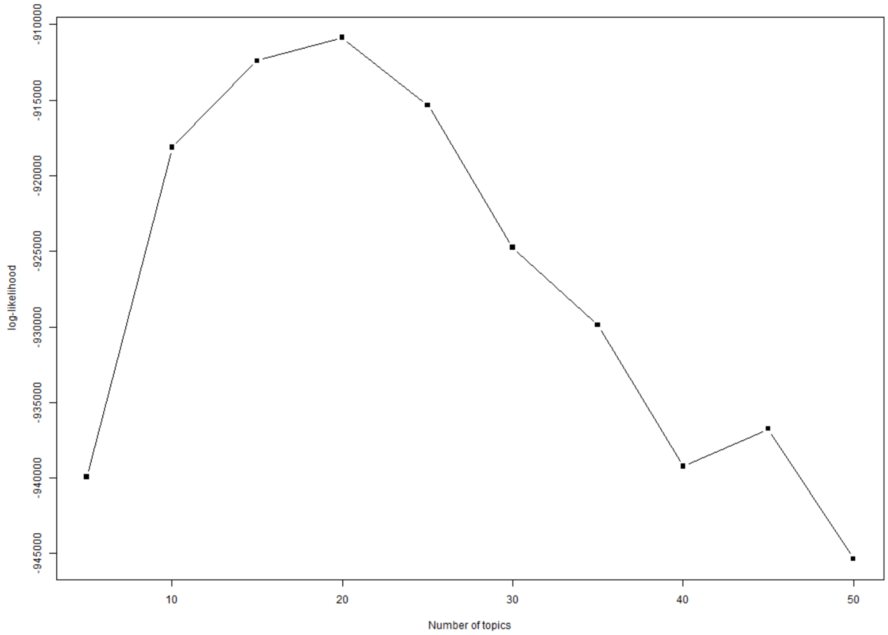


Figure B2: The most likely four words generated by each topic.

1	2	3	4	5
circuit	control	material	light	communication
machine	vehicle	producing	electronic	mobile
integrated	power	materials	devices	wireless
printing	motor	product	unit	radio
6	7	8	9	10
treatment	transmission	process	systems	network
treating	data	metal	devices	service
agent	signal	production	monitoring	management
delivery	digital	producing	medical	providing
11	12	13	14	15
engine	process	optical	based	display
combustion	surface	sensor	detection	manufacturing
internal	water	antenna	detecting	structure
protein	producing	component	determining	element
16	17	18	19	20
compositions	information	composition	cell	derivatives
comprising	processing	film	fuel	compounds
preparation	data	coating	type	acid
active	image	forming	heat	inhibitors

Once the 20 topics have been identified, we redefined each patent as a combination of these 20 topics. Figure B3 shows an illustration of how topics contribute to the patent titled “Method and device for preparing fibrous plant bodies.”. The patent is characterized by topic 3 and 12 that have the highest shares.

Figure B3: Share contribution of topics in patent titles.

Ttitle	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Method and device for preparing fibrous plant bodies.	5%	5%	8%	5%	5%	6%	5%	5%	5%	5%	5%	6%	5%	5%	5%	5%	5%	5%	5%	5%

Next, to confirm that each novel technology and its re-users have a meaningful relationship, we need to show that patents belonging to the same group are homogenous in terms of topic content. To do that, we grouped the patents belonging to the same novel technology according to our definition reported in section 2: a patent is assigned to a novel technology if it has a specific new pair of IPC codes in its IPC classification list. Then, we evaluated the level of homogeneity of each topic within the group of patents embedding the same novel technology. To do so, we calculated the standard deviation of each topic share (see Figure B4). Then, we calculated the average of those standard deviations by group (novel technology) (see Figure B5).

Figure B4: Standard deviation of each topic share by novel technology. A patent is assigned to a technology if it has a specific pair of IPC codes in its IPC classification list.

IPC class	IPC class	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	#patents
A01D45	A01D43	2.8%	0.6%	1.0%	0.8%	0.6%	0.6%	0.6%	0.6%	1.3%	0.7%	1.1%	0.8%	0.3%	0.1%	0.8%	0.5%	0.5%	0.4%	0.6%	0.4%	26

Figure B5: Average standard deviation of topics shares by novel technology. A patent is assigned to a technology if it has a specific pair of IPC codes in its IPC classification list.

PC class	IPC class	Average standard devation	#patents
A01D45	A01D43	0.76%	26

To assess if topics are homogeneous or not within each group, as represented by the average standard deviation calculated in Figure B5, we need to have a comparison value. We constructed a comparison sample where each patent is assigned randomly to a group (and not grouped according to its IPC codes). Each group maintains the same number of patents as in our original classification. We computed the average standard deviation of topics shares of each of those groups where patents are randomly assigned (see Figure B6).

Figure B6: Average standard deviation of topics shares by novel technology. A patent is assigned randomly to a technology.

PC class	IPC class	Average standard devation	#patents
A01D45	A01D43	1.02%	26

Finally, we test if the average standard deviations in the two samples, i.e. the original sample and the one with patent randomly assigned to technologies, are statistically different. We find that the average standard deviation of the topics share in the original sample is significantly lower than in the sample with randomly assigned patents (0.008 vs. 0.012, P-value 0.000). This result confirms that topics are more homogeneous in a sample where patents are assigned to the novel technology according to the IPC code combinations rather than when they are assigned randomly. In other words, the method applied to identify the group of patents (re-)using a novel technology seems to link patents with coherent contents.

Appendix C: S-curve fitting

A good fit between the actual data and the estimated S-curve is crucial in our empirics. To test the quality of our estimated curve within our observation period as well as the quality of our predictions after the 20-year window, we extract the cohort of novel technologies who appeared in 1985. The 1985 technology cohort allows us to observe 20 years of actual data on which we estimate our S-curve leaving a buffer of 10 additional years of actual data to use as a benchmark for our predictions. To evaluate the fit of the S-curve to our data, we replicate the idea of the R-squared index in an OLS regression, where the fitting index is 1 minus the ratio of the residual variation (SSR) compared to the total variation (SST). To implement this approach, we first compute the difference between the real cumulated number of patents for the technology i in year t (y_{it}) and the value predicted by the S-curve $\widehat{y_{\iota t}}$. Then, we square and sum the values to obtain SSR=$\sum_{t=1}^{T}\sum_{i=1}^{I}(y_{it}-\widehat{y_{\iota t}})^2$ where T is the time span considered and I is the total number of novel technologies. We calculate the total variation as SST=$\sum_{t=1}^{T}\sum_{i=1}^{I}(y_{it}-\bar{y})^2$ where $\bar{y}=\frac{\sum_{t=1}^{T}\sum_{i=1}^{I}y_{it}}{T*I}$. We are interested in evaluating the goodness of the fitting in three window-period of 10 years (T=10), i.e. 1-10, 11-20, and 21-30, for all the technologies who appeared in 1985 (I=1,461).

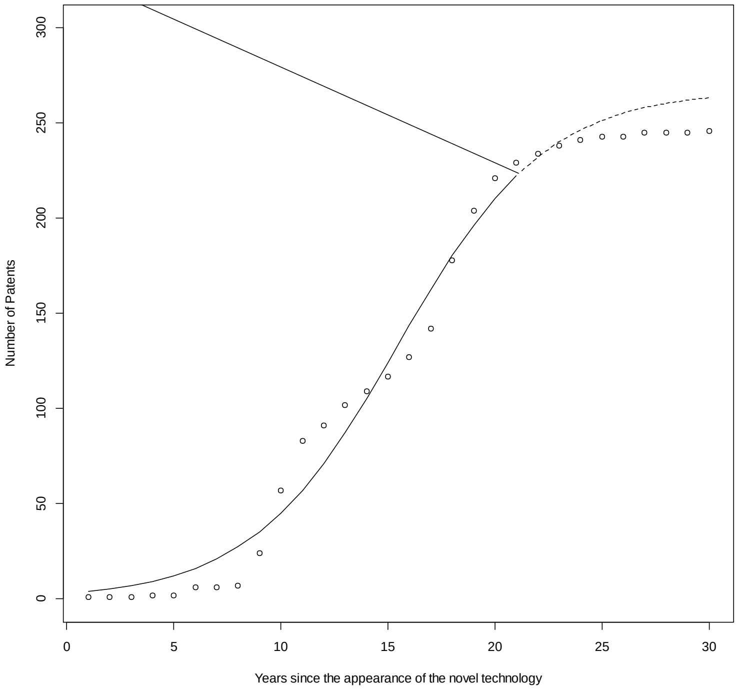
Table C1 shows that our curves have a good fit within each 10-year window. The fitting index ranges from 0.926 to 0.973. Although the values of the last period (21-30), not included in the S-curve estimation, are the lowest, their fitting is not significantly different for the other two periods. This implies that overall the quality of our predictions after the 20th year declines only marginally (7% = 0.997-0.926 in ten years) and that our predictions of the unobserved part of the S-curve, i.e. the cumulate until the *Ceiling*, are reliable.

Table C1: Goodness of fitting for the S-curve technologies of the cohort 1985

Period	SSR	SST	Fitting (1-SSR/SST)
1-10	96,842	3,529,687	0.973
11-20	140,488	41,473,820	0.997
21-30 (predictions)	1,232,1082	167,500,000	0.926

As a graphical example of fitting, figure C1 shows the actual cumulated number of patents re-using the transgenic mammal technology embedded for the first time in the onco-mouse patent in 1985 (dots). In the same figure we draw the estimated S-curve distinguishing the predicted values within the 20-year period (solid line) from the predicted values for an additional 10 years (dashed line).

Figure C1: Graphical representation of the goodness of fitting for the mammal technology curve



Appendix D: Legitimation as a linear combination of midpoint and alpha of an S-curve technology diffusion curve

The mathematical formulation relating legitimation (defined as the time needed for diffusion curve to reach the 10% of its ceiling point) to the midpoint and alpha can be shown as follows. Starting from Equation 1, we first extract *t* :

$$\hat{Y}_t = \frac{Ceiling}{1 + e^{(-\frac{(t-Midpoint)}{Alpha})}}$$

$$1 + e^{(-\frac{(t-Midpoint)}{Alpha})} = \frac{Ceiling}{\hat{Y}_t}$$

$$-\frac{(t - Midpoint)}{Alpha} = \ln(\frac{Ceiling}{\hat{Y}_t} - 1)$$

$$-t + Midpoint = Alpha * \ln(\frac{Ceiling}{\hat{Y}_t} - 1)$$

$$t = Midpoint - Alpha * \ln\left(\frac{Ceiling}{\hat{Y}_t} - 1\right)$$

Next, we set the 10% of the ceiling point, $\frac{\hat{Y}_t}{Ceiling} = \frac{1}{10}$ and compute the corresponding t

$$t_{10\%} = Midpoint - Alpha * \ln(10 - 1)$$

$$t_{10\%} = Midpoint - Alpha * 2.2$$

Appendix E: OLS estimation using actual data

	(1)	(2)
VARIABLES	Actual Legitimation time	Actual Technological Impact
Technological component characteristics		
Similarity 3 digits (dummy)	-0.79***	-7.59***
log(Science-based content)	0.25***	15.4***
log(Familiarity)	-0.27***	-6.81***
Inventor's characterisics		
log(Inventors' team size)	-0.14***	4.24***
Applicant characteristics		
log(Applicants' experience)	-0.010	0.80***
University applicant (dummy)	-0.032	8.50**
More than one applicant (dummy)	-0.97***	4.90**
Dummy Technology class (3 digits)	Yes	Yes
Dummy Applicant's country	Yes	Yes
Dummy technology entry year	Yes	Yes
Constant	9.30***	109***
Observations	10,782	10,782
R-squared	0.226	0.135

Appendix F: Splitting the sample between technologies that reach their ceiling within the first 20 years from their appearance and the others

In Figure F1, we plot the actual cumulated number of patents at the end of our observation period (20 years) against the estimated full technological impact (the ceiling of the s-curve). The plot should be interpreted as follows:

- The closeness of the points to the diagonal indicate that the estimated ceiling is close to the actual number of patents;
- Staying on the diagonal means that the technology reaches its ceiling within 20 years.

Figure F1: Actual cumulated number of patents at the end of the observation period (when t=20 years) against the estimated full technological impact (the ceiling of the s-curve)

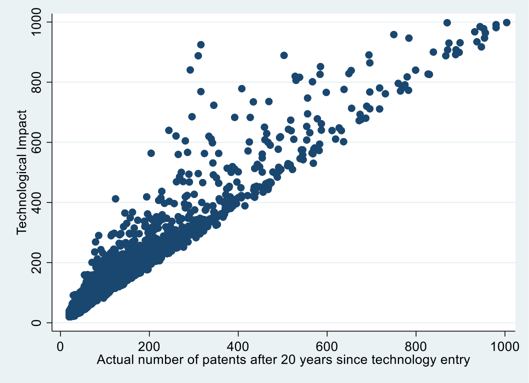


Table F1 apply the regression model reported in Table 2 to two different subsamples of novel technologies. Subsample A includes the technologies reaching their ceiling within the first

20 years (+/-15%), while subsample B includes technologies not reaching their ceiling within the first 20 years.

Table F1: Legitimation and technological impact estimations on two subsamples

	(1)	(2)	(3)	(4)
	Subsample A: Technologies reaching their ceiling within the first 20 years		**Subsample B: Technologies NOT reaching their ceiling within the first 20 years**	
	68% of the cases		**32% of the cases**	
	Legitimation (10%)	Technological Impact	Legitimation (10%)	Technological Impact
Component characteristics				
Similarity 3 digits (dummy)	-1.28***	-11.8***	-0.77***	1.50
log(Science-based content)	0.16***	16.2***	0.25***	17.0***
log(Familiarity)	-0.11***	-5.45***	-0.37***	-15.2***
Inventor’s characterisics				
log(Inventors' team size)	-0.14***	4.98***	-0.082	4.81*
Applicant characteristics				
log(Applicants' experience)	-0.014	0.76*	0.0091	1.39*
University applicant (dummy)	0.17	7.58*	0.071	10.7
More than one applicant (dummy)	-0.81***	4.08	-0.96***	6.15
Dummy Technology class (3 digits)	Yes	Yes	Yes	Yes
Dummy Applicant’s country	Yes	Yes	Yes	Yes
Dummy technology entry year	Yes	Yes	Yes	Yes
Constant	8.63***	106***	11.7***	172***
Observations	7,329	7,329	3,453	3,453
R-squared	0.260	0.131	0.256	0.238

Appendix G: 2-step Heckman selection model

To correct for the selection bias associated with the restriction to successful novelty, i.e. with a minimum level of adoption, we adopt a 2-step Heckman estimation strategy. In the first step, we use all the novel technologies in our sample without any restriction. These cover 191,400 observations. In table G1, we estimate a Probit model where the dependent variable is a dummy that equals 1 if the novel technology diffuses, 0 otherwise. Zero reflects the case of a novel technology with less than 20 re-using patents in our 20-year window. We assume that the probability to diffuse or not depends only on the characteristics of the patented inventions embedding the novel technology in the first year of their appearance. In the second step, we return to our sample of 10,782 successful novel technologies and estimate a model explaining legitimation and technological impact including as explanatory variables the characteristics of the combined technological components and add the Inverse Mill's ratio to correct for the selection bias. Table E2 reports the results of our estimations without and with the correction for the selection bias, respectively columns 1-2 and 3-4.

Table G1: Heckman first step, the probability of a novel technology diffusion

VARIABLES	(1) Probit Diffusion
log(Inventors' team size)	0.086***
log(Applicants' experience)	0.013***
University applicant (dummy)	0.23***
More than one applicant (dummy)	0.14***
Constant	-1.82***
Observations	191,400
Dummy application year	Yes
Dummy technology class (3 digits)	Yes
Dummy applicant's country	Yes
Pseudo-R2	0.035

Table G2: Heckman second step, estimation of legitimation and technological impact equations correcting for selection bias

VARIABLES	(1) OLS Legitimation (10%)	(2) OLS Technological Impact	(3) OLS Legitimation (10%)	(4) OLS Technological Impact
Similarity 3 digits (dummy)	-1.23***	-7.74**	-1.19***	-8.77***
log(Science-based content)	0.14***	18.2***	0.19***	16.7***
log(Familiarity)	-0.21***	-8.98***	-0.20***	-9.12***
Inverse Mill's ratio (nonselection hazard)			2.01***	-52.4***
Constant	10.1***	166***	6.29***	263***
Observations	10,782	10,782	10,781	10,781
R-squared	0.214	0.132	0.229	0.143
Dummy application year	Yes	Yes	Yes	Yes
Dummy technology class (3 digits)	Yes	Yes	Yes	Yes

We find a substantial coherence between the results reported in Table G2, columns 3 and 4 and the results reported in the main text in Table 2, columns 5 and 6. This allows us to claim that restricting our sample to successful novel combinations does not introduce any serious selection bias on the estimated impact of the characteristics of the combined components on diffusion.