KU LEUVEN

FACULTY OF ECONOMICS AND BUSINESS

Measuring technological novelty with patent-based indicators

Dennis Verhoeven, Jurriën Bakker and ReinhildeVeugelers

MSI_1509

Measuring technological novelty with patent-based indicators

Dennis Verhoeven[a], Jurriën Bakker[a] and Reinhilde Veugelers[a,b,c,d]

[a] KULeuven, Management, Strategy and Innovation Department, Leuven
[b] Senior Fellow; Bruegel, Brussels; [c] Research Fellow; CEPR, London [d]ECOOM, Leuven

This version: April 2015

Abstract

This study provides a new, more comprehensive measurement of technological novelty. Integrating insights from the existing economics and management literature, we characterize inventions ex ante along two dimensions of technological novelty: Novelty in Recombination and Novelty in Knowledge Origins. For the latter dimension we distinguish between Novel Technological and Novel Scientific Origins. For each dimension we propose an operationalization using patent classification and citation information. Results indicate that the proposed measures for the different dimensions of technological novelty are correlated, but each conveys different information. We perform a series of analyses to assess the validity of the proposed measures and compare them with other indicators used in the literature. Moreover, an analysis of the technological impact of inventions identified as novel shows that technological novelty increases the variance of technological impact and the likelihood of being among the positive outliers with respect to impact. This holds particularly for those inventions that combine Novelty in Recombination with Novelty in Technological and Scientific Origins. The results support our indicators as ex ante measures of technological novelty driving potentially radical impact.

Keywords: Novelty, patent based indicators, radical inventions;

JEL-Classification: H23, O31, O38

Corresponding author's contact details:

Reinhilde Veugelers, K.U.Leuven, Department of Managerial Economics, Strategy and Innovation, Naamsestraat 69, 3000 Leuven, Belgium. reinhilde.veugelers@kuleuven.be

Acknowledgment: The authors are thankful to seminar participants at MSI-KU Leuven, OECD-WPIA workshop in Paris, the EPIP conference in Brussels, the FINKT & CISS PhD summer schools for useful comments, in particular to Ashish Arora, Sam Arts, Rene Belderbos, Antonio Della Malva, Bart Leten, Stuart Graham, Dominique Guellec, Katrin Hussinger, Mariagrazia Squicciarini and Bart Van Looy. Financial support from KULeuven (GOA/12/003) and FWO (G.0825.12) is gratefully acknowledged. The authors thank ECOOM for allowing access to the patent and publication data. Publication data are sourced from Thomson Reuters Web of Science Core Collection.

1. Introduction

Technology is generally believed to develop along well-defined and predictable trajectories, occasionally interrupted by discontinuities introduced by paradigm shifts (Dosi, 1982). At the heart of such technological discontinuities are inventions that introduce a novel technological approach (Arthur, 2007, 2009) with a potentially game changing impact on industries and markets (Cooper & Schendel, 1976; Henderson & Clark, 1990). As an example, Arthur describes how the turbojet engine introduced the concept of generating thrust by expelling particles to create an opposite force to accelerate an airplane. This was a novel technological approach compared to typical propeller engines that generate a drag in order to drive the plane. In the course of the following decades, a series of follow-on incremental improvements refined this novel approach in terms of its functioning and performance. All this led to an unprecedented performance increase of jet engines, allowing for a tremendous growth of the aviation industry.

Radical inventions introducing technological novelty, like the turbojet engine, are, more than run-off-the-mill improvements, subject to uncertainty with respect to potential technological and commercial performance (Fleming, 2001; Hall & Lerner, 2010). Beyond their potential for high impact, they have the ability to disrupt existing competences (Tushman & Anderson, 1986), and eliminate existing players from the market (Christensen, 2003). Firms aiming at technologically novel inventions, might require substantially different competences and search strategies (Ahuja & Lampert, 2001). Because of their distinct profile, unpacking the drivers and effects of radical innovations is of major interest to scholars studying the economics and management of innovation.

Technological novelty can be an important driver of radical technological innovation like in the case of the turbojet engine (Arthur, 2007). However, it is clear that not all technologically novel inventions will result in successful innovations with profound technological and economic impact, as they are typically subject to higher uncertainty (Fleming, 2001). And even if eventually successful, the process of development and realization of impact is generally a lengthy one (Rosenberg, 1976), where the invention may pass through several different agents before fruition. Hence, to fully grasp the mechanisms underlying the origins, diffusion and effects of radical innovations, it is important to characterize as early as possible these innovations on their technological novelty. Yet, in most empirical studies, radical or breakthrough inventions are identified only by their ex post large impact on future technological development (Ahuja & Lampert, 2001; Schoenmakers & Duysters, 2005), product performance (Leifer et al., 2001) or market structure (Mascitelli, 2000). Only considering inventions which have been highly impactful typically introduces a success bias. Many inventions with the potential to have radical impact, may not realize this potential and are therefore missed in the analysis of successful inventions only. Furthermore, an approach which is truly novel often times needs considerable refinements. The invention identified as impactful might not be the one having introduced the novelty, but rather one that builds further on the novel approach introduced by another invention. Hence, using an ex post definition and operationalization based on direct impact performance, does not take into account unsuccessful novelty and novelty with indirect impact. These limitations motivate us to argue that, to gain more insight in the mechanisms driving radical technological change, it is important to characterize ex ante technological novelty distinctly from ex post technological and economic impact. This will help us to better understand the process from creation of novel ideas to possible successful implementation and learn how to improve this process.

In this paper, we contribute to a more comprehensive measurement of technological novelty, characterizing radical inventions ex ante. We build on the work of Arthur (2007, 2009) to identify two important dimensions of technological novelty – Novelty in Recombination and Novelty in Knowledge Origins. Novelty in Recombination reflects the extent to which an invention is novel in the way it recombines components and principles to serve its purpose. Novelty in Knowledge Origins reflects the extent an invention draws knowledge from previously unconnected fields of knowledge. We distinguish between knowledge developed from scientific work (Novelty of Scientific Origins), and knowledge developed through previous technological effort (Novelty of Technological Origins). We use classification and citation information on patent documents to operationalize these dimensions for all patented inventions since 1980.

We perform a number of analyses to assess the validity of the proposed measures. First, we illustrate our measures with the patent for the 'onco-mouse' and a number of other well-known novel technologies. Second, we compare our measures to measures of related constructs commonly used in the literature, more particularly the "originality" measure introduced by Trajtenberg et al. (1997) and the "radicalness" measure employed by Shane (2001). Our technology measures correlate with these existing measures of related constructs, but exhibit distinct patterns which are more closely associated with technological novelty. Third, we perform two larger validation exercises using external information about novelty of inventions. First, is a set of inventions that were awarded an R&D prize by 'R&D Magazine' for being among the most technologically significant inventions of the year. Second, we use a sample of patents that were refused by the European Patent Office (EPO) because they lacked novelty or inventive step to assess false positive bias. We find that inventions identified as novel are overrepresented in the group of award-winning inventions, and underrepresented in the group of inventions that were refused a patent for lack of novelty. Finally, we analyze the technological impact generated by inventions and find that inventions identified through our indicators as being technologically novel, have a higher dispersion in terms of forward citations received, and are more likely to end up among the set of highly cited patents, confirming their higher risk profile and their higher probability to be the antecedent of a radical breakthrough. Overall, the results support our indicator as ex ante measure of technological novelty with the potential to drive radical technological change.

The remainder of the paper is structured as follows: In the second section we discuss the concepts and measurement introduced in prior literature. In the third section we conceptualize technological novelty and propose our new patent-based indicators to identify technological novelty. In the fourth section we perform descriptive analyses of the relatedness between the different dimensions of novelty we identify and compare the indicators to measures of related constructs. In the fifth section we perform a number of validity checks using external information on (lack of) novelty of inventions. The sixth section provides an analysis on the impact of patents characterized by technological novelty. The final section discusses the implications of the results, and avenues for future research.

2. Background

Reflecting the multitude of angles from which technological change is studied, considerable variety exists in the definition and measurement of concepts related to what can be broadly termed 'radical invention'. A range of labels (radical, discontinuous, breakthrough, new, etc.)[1] is given to phenomena touching upon different dimensions of inventive outcomes. A nevertheless common theme across the different angles is the notion of a break from the past and/or a large impact on the future along some technological or economic dimension(s). This section structures the different meanings the literature associates to 'radicalness' of inventions and provides an overview of different practices to operationalize some of these concepts. It will mainly take a technological perspective rather than an economic perspective to identify the characteristics of "radical inventions".

2.1. Concepts

2.1.1. Ex post technological impact of invention

Scholars within the perspective of technology trajectories (Dosi, 1982) define radical invention in terms of their impact on future technological development. Their search is for those inventions that introduce new paradigms that define patterns of problems and ways to address them. These new approaches open up avenues for further technological developments, starting new trajectories. In this spirit, radical inventions

[1] Because of the lack of consensus in terminology for different constructs related to technological change, we choose to loosely adopt the term 'radical invention' in its broad meaning for the sake of readability. It is to be noted that this term covers a wide range of different constructs, including novelty and impact, as well as technological and economic characteristics of technologies which are reviewed in this section.

are described as inventions that serve as the basis for many subsequent technological developments, representing a significant leap forward in the technological frontier. An invention on which many future inventions build is argued to be a breakthrough (Fleming, 2001; Ahuja & Lampert, 2001) or deemed radical (Schoenmakers & Duysters, 2010).

Radical inventions have been defined in terms of the profound impact they have on firms, industries and markets. Anderson & Tushman (1990) distinguish between competence-enhancing and competence-destroying technological discontinuities. Utterback (1996) defined radical invention or discontinuous change as "change that sweeps away much of a firm's existing investment in technical skills and knowledge, designs, production technique, plant and equipment," and Henderson (1993) described an invention as being radical in the organizational sense when it renders a firm's information filters and organizational procedures (partially) obsolete.

2.1.2. Ex ante characteristics of invention

Rather than looking ex post at impact, a number of scholars define radical invention in terms of the characteristics of their underlying technology. In this ex ante perspective, radical inventions are often characterized as incorporating technologies that move away from existing practices (Ettlie et al., 1984; Mascitelli, 2000; Shane, 2001; Dahlin & Behrens, 2005), embedding novel knowledge (Dewar & Dutton, 1986; Lettl et al., 2006; Carlo et al., 2012) and being based on different scientific and engineering principles compared to existing technology (Henderson & Clark, 1990). Radical inventions are considered to combine existing or new components in an unprecedented manner (Nooteboom, 2000; Gassmann & von Zedtwitz, 2003; Nemet, 2009; Story et al., 2011). Some scholars argue that radical inventions do not build on any existing technology (e.g. Ahuja & Lampert, 2001; Banerjee & Cole, 2011).

Applying a firm-level rather than an invention-level lens, some studies consider the degree of 'newness to the firm's competences and activities' as a component of the radical invention construct (e.g. Chandy & Tellis, 1998; Garcia & Calantone, 2001). Necessary knowledge or information leading to or resulting from radical invention is considered to reside outside the knowledge base present within firms (Dewar & Dutton, 1986; Hill & Rothaermel, 2003). As a consequence, radical inventions are believed to require significant adaptation of firms' existing routines and activities (Damanpour, 1996; Nahm et al., 2003).

2.2. Measurement of technological dimensions characterizing radical inventions

2.2.1. Qualitative assessment

Following the ambiguity in defining radical invention, a range of different methods has been used in the literature to empirically measure the technological dimensions associated with radical invention. The empirical literature on technological trajectories engages in mapping one or more performance criteria of technologies over time in order to identify those inventions that instigated a drastic leap in technological performance (Tushman & Anderson, 1986; Anderson & Tushman, 1990), or to compare competing technological solutions over time as they proceed through their technological life cycle (Christensen & Bower, 1996). Large-scale empirical analysis is hindered by the time-intensive nature of the method of systematically analyzing and mapping technological trajectories.

A more direct measurement of the dimensions of radical invention is pursued by using surveys on managers or industry experts in order to attain an expert judgment about the radical features of a certain technology or product (e.g. Dewar & Dutton, 1986; Pavitt et al, 1987; Acs & Audretsch, 1990; Chandy & Tellis, 2000). With this approach, a considerable subjectivity and hindsight bias might exist, favoring those inventions publically identified as important, the ones the respondent is most familiar with or those that are best recalled by the respondent. The problem of establishing a large enough qualified panel for evaluation, often restricts the set being assessed to a handful of inventions in a particular area.

2.2.1. Large scale quantitative assessment using patent information

Patents are an alternative data source, allowing large scale assessment of technological developments. Patents leave an open paper trail of forward citations and backward citations. This trail allows tracing the origin of ideas and where ideas go when they are cited in the future. Patent citations also give an indication of the economic value of patents and there are strict procedures for citations to be issued (Griliches, 1990; Jaffe & Trajtenberg, 2002). Patent information has been widely used to assess ex post technological impact (Carpenter, 1981; Fleming, 2001) and (market) value of inventions (Gambardella et al., 2008; Hall et al., 2005), using information on the number of citations received.

A number of studies use patent information to measure ex ante characteristics which can be associated with radical inventions. These studies typically rely on backward citation information to assess the extent to which an invention builds on existing knowledge. Having no backward citations is argued to be a characteristic of radical invention (Ahuja & Lampert (2001); Banerjee & Cole, 2011). Although it is sensible that an invention that cites no prior art provides novelty in its functionality, this measure does not include inventions that apply new principles and components that were in existence by previously unrelated technologies. Schoenmakers & Duysters (2010) for instance, find that radical patents, which they define as highly cited patents, have more rather than less backward citations.

Rather than looking at the number of backward citations, a series of works look at the characteristics of the backward citations. Trajtenberg et al. (1997) look at the spread of backward citations across technology classes and postulate that the more backward citations are spread, the more "original" the patent is. Rosenkopf and Nerkar (2001) and Shane (2001) look at the number of technological classes a patent cites outside its own technology classes, as measure of radicalness. Age of the patents cited, spread in age (Nerkar, 2003) and the number of references to scientific publications (Gittelman & Kogut, 2003) are argued to be a reflection of novelty of knowledge sources of inventions. The number of citations to patents residing from other firms and the number of technological fields occurring for the first time in the patent portfolio of the firm are used to reflect the extent to which a firm is sourcing knowledge outside its own boundaries (Rosenkopf & Nerkar, 2001; Ahuja & Lampert, 2001). These measures compare inventions of firms to their technologies already in place, and therefore provide a picture on whether technologies are new to the firm. This does however not need to imply that the firm introduced technological novelty on an invention level. None of these measures compares the invention under analysis to previous technological developments. They measure knowledge intensity and type of knowledge used for a particular (set of) invention(s) rather than the construct of technological novelty. Moreover, few attempts are made to directly assess the validity of these indicators using information not residing from patent documents. At best, face validity is established by the arguments made to motivate the measures and descriptive analyses showing plausible patterns.

The most comprehensive attempt to map radical inventions using patent citation information is Dahlin & Behrens (2005). For them, an invention is radical when it passes the criteria of novelty, uniqueness and impact. Here, novelty is measured by backward citation patterns of an invention: if a patent has a low overlap in cited documents compared to patents in previous years, it is considered to be novel. Uniqueness is measured similarly by comparing the backward citations of the focal patents to the patents in the same year. Impact is measured by the amount of overlap of patents in the years following the invention and the focal patent. The latter measure is an addition to the construct of technological impact, capturing the novel domains of impact. The main advantage of this approach is that it explicitly conceptualizes radical inventions along a combination of dimensions and attempts to create indicators for each of them separately. The main disadvantage is the computational complexity introduced by comparing sets of citations of each patent to the citations used by every other patent in the population. In their contribution, the measures are validated using the small case of US patents in the tennis racket industry as a reference group to compute backward citation overlaps of each patent in this set. Their proposed measures have so far not been implemented on a large scale because of problems selecting the comparison group used to assess overlap patterns in backward citations when assessing a multitude of technological fields.

Finally, other patent information beyond citation information can be used to ex ante characterize technological inventions. To measure the newness of the components and combination of components embodied in the invention, Fleming (2007) looks at the occurrence of new combinations of subclasses to which a patent is assigned. He uses this as a measure of generative creativity of inventors on the patent. In Fleming (2001), he measures how often and how recent a certain component/combination of components is used before. He uses this measure as measure of familiarity of (combinations of) components, driving recombinant uncertainty in the creation of breakthrough patents.

3. Our approach to characterize technological novelty

As argued in the introductory section, it is important to make a distinction between the ex ante and ex post characteristics of radical inventions. In this contribution, we focus on ex ante technological novelty of inventions. Most of the existing literature agrees on the conception that radically novel inventions incorporate a departure from existing practice in terms of knowledge embodied in the invention and/or technological components used in it. However, most definitions remain implicit as to how to measure technological novelty. In this section, we provide a clear definition of technology which allows us to define and construct measures of technological novelty.

3.1. Defining technology

Following the work of Arthur (2007, 2009) we define technology as being a means to fulfill some purpose. A technology embodies principles and consists of components in an architecture that work in relation to each other to exploit the principles of working that meet the purpose at hand. As such, technologies differ from each other in their ***recombination*** of existing components and principles (Fleming, 2001; Hargadon, 2002)[2]. In addition, technologies can differ from each other in terms of the domains of knowledge they draw from to guide the choice of components and principles used to serve its purpose. Different domains of knowledge can be sourced. ***Scientific knowledge origins*** source knowledge accumulated around scientific exploration while ***technological knowledge origins*** source from domains of knowledge built up by development of technologies itself.

[2] Only very rarely completely new components or principles arrive. Most of the time the novelty arises from combining existing components in a new fashion.

3.2. Defining technological novelty

Assessing whether something is novel implies comparing to what existed before along certain attributes. We rely on the characterization of technology outlined above to define the attributes that can be used to assess technological novelty of an invention. The basis of comparison is the whole of previous technologies in place.

> ***Novelty in recombination****:* An invention is identified as having novelty in recombination if the combination of components and principles of working applied to serve its purpose are different from those embodied in previous technologies.
>
> ***Novelty in technological knowledge origins****:* An invention is identified as having novel technological knowledge origins if it draws technological knowledge from domains that were previously not used in the technological domain of the invention.
>
> ***Novelty in scientific knowledge origins****:* An invention is identified as having novel scientific knowledge origins if it draws scientific knowledge from domains that were previously not used in the technological domain of the invention.

The constructs defined above cover distinct but related characteristics of technological novelty. While novelty in recombination concerns the difference between approaches to serve a given purpose, the construct of novelty in knowledge origins aims at capturing change in knowledge clusters drawn from to serve the purpose of the approach. The dimensions proposed are distinct from each other since novel recombination can be achieved within the bounds of the knowledge pool relied on before. Vice versa, one could imagine technologies tapping into novel domains of knowledge without providing drastic novelties with respect to the principles and components used for the technology. Nevertheless, despite being distinct, both dimensions are related. Incorporating novel principles and technological elements to serve a purpose will often require drawing knowledge from different domains compared to previous approaches. Moreover, drawing from different knowledge pools might inspire inventors with respect to the technological elements/principles used to solve problems encountered.

3.3. Measuring technological novelty

Although a number of interesting patent measures have been proposed in the literature (cf. supra), no comprehensive operational measurement of the construct of technological novelty as conceptualized in the previous section, is available. In this section we provide measures for novelty in recombination and novelty in knowledge origins using classification information of patents and scientific literature.

3.3.1. Novelty in recombination

As detailed in section 3.2, we identify inventions having novelty in recombination as inventions that apply novel combinations of components and principles of working to serve a certain purpose.

We proxy the recombination embodied in a technology by looking at the components and principles constituting the technology. Like Fleming (2007), we rely on the (pairwise) combinations of technological classifications assigned to a patent to proxy the body of components and principles used to serve the purpose of the underlying invention. More specifically, we rely on the group-level IPC-codes (International Patent Classification)[3] to which a patent is assigned. For each pair of IPC-codes, we assess previous existence of the pair in the body of patents before the application year of the patent under consideration, providing us with a large pool of pairs as a comparison group for assessing novelty. Employing this method, we define our measure as follows:

> **A patent has "Novelty in Recombination (NR)" when it contains at least one pair of IPC groups that were previously unconnected.**

Fleming (2007) has applied a similar approach using the US Patent Office Classification system (USPC) to proxy inventors' generative creativity. We depart from this approach by employing the IPC classification scheme for three reasons. First, the IPC system is more comprehensive. In fact, every patent application in the population receives an IPC classification while USPC classes are only assigned to patents filed for at the USPTO. Hence, using the USPC would only compare patents filed for at the USPTO to the group of USPTO patents. Second, as argued by Gruber et al. (2013), the IPC classification is more suitable when assessing the recombinant nature of an invention because it classifies according to the complete technological information contained in the application, whereas the USPC system classifies according to the scope of protection. Third, for our purpose, the USPC provides an overly disaggregated level of classification (about 160 000 subclasses). Hence, because of the very sparse nature of the co-occurrence matrix, employing our measure using this disaggregated level results in the identification of almost half of the patents as novel.

---

[3] The International Patent Classification (IPC) was established in 1971 by the Strasbourg Agreement and provides a hierarchical system to classify patents according to the technological areas they belong to. It uses 5 layers of detail to classify a patent documents labeled respectively 'Section' (8), Class (±130), 'Subclass' (±630), 'Group' (±8 000) and 'Subgroup' (±70 000). See http://www.wipo.int/classifications/en/ for more information.

3.3.2. Novelty in knowledge origins

For our second construct, novelty in knowledge origins, we identify patents that rely on technological or scientific knowledge as prior art from domains that were previously not used for the purpose of the invention.

We use references provided in patent documents to denote the knowledge origins inventions draw from. Technological and scientific classifications are used to reflect the domains of knowledge of cited prior art. The approach followed departs from existing measures using number of backward citations and technological classification of the backward citations (Trajtenberg et al., 1997; Shane, 2001; Nerkar, 2003; Gittelman & Kogut, 2003; Schoenmakers & Duysters, 2010) by allowing to identify those inventions that make *novel* connections between fields of knowledge, a characteristic which none of these existing measures account for.

3.3.2.1. Novelty in technological knowledge origins

To identify the domains of technological knowledge which an invention uses as prior art, we look at the patents referenced by the focal patent and identify their IPC classification (again at the group level). We construct ‘backward citation pairs’ of IPC-codes, i.e. combinations between distinct IPC-codes from, on the one hand, all patents cited by the focal patent and, on the other hand, all distinct IPC-codes the focal patent belongs to. We compare each of the focal patent’s ‘backward citation pairs’ to all citation pairs previously used to assess whether a certain pair is new (has never occurred before).

> **We identify a patent as having “Novelty in Technological Origins (NTO)” if it makes backward citation pairs that have not occurred in the years previous to the application year of the patent.**

3.3.2.2. Novelty in scientific knowledge origins

To identify the scientific knowledge used as prior art for the invention, we look at the scientific documents referenced in patents. We use the IPC-codes present on the patent document and the Web of Science (WOS) journal classification of the scientific articles cited to create “IPC – scientific field” combinations. To retrieve information of WOS-categories of scientific articles referenced in patent documents, we first extract Non-Patent References (NPR’s). Next, we link[4] the NPR’s to the WOS-subject category of the

[4] NPR’s are classified as being scientific based on the algorithm developed in Callaert et al. (2012) which allows extraction of journal information of NPR’s. We use the Thomson Reuters Web of Science database to extract the scientific field of these journals.

journal in which the reference appeared. Pairs between IPC-classes and scientific categories introduced by the patent applications in our source dataset, are compared to all previously made pairs.

> **We identify a patent as having "Novelty in Scientific knowledge Origins (NSO)" if it makes a connection between an IPC-code and a scientific field that has not occurred in the years previous to the application year of the patent.**

4. Implementing our measures of novel recombination and novel knowledge origins

4.1. Our sample of technological inventions

The sample we use to implement our measures of novel recombination and novel knowledge origins includes all utility patent applications filed between 1980 and 2011 present in the October 2011 version of PATSTAT. As we want to look for novelty in the universe, we include all major patent offices. We use filings at the European Patent Office (EPO), US Patent and Trademark Office (USPTO) and World Intellectual Property Office (WIPO) (the latter including patents that were filed using the Patent Cooperation Treaty, also referred to as the PCT-route)[5]. This total sample contains 8 426 062 patent applications. Since one invention is usually patented through multiple offices, we can have multiple equivalent patents for one invention (a so-called 'patent family'). To remedy this issue, we opt to use the patent family according to the DOCDB definition (Martinez, 2011) as a unit of analysis. We aggregate at the family level, by taking the maximum value for the relevant indicators. When using application years, we take the application year of the first patent in the family. After aggregation we are left with 5 297 283 observations.

For the construction of the novelty indicators we use information on IPC-classification, patent citations and application filing year for *all* patent applications present in PATSTAT. We use citations to other patent documents, as well as references to scientific articles to proxy the technological and scientific knowledge

---

[5] Due to a number of peculiarities in the PATSTAT database, a number of cleaning steps were performed. Duplicates caused by untraceable priorities and citations were removed. Moreover, we removed duplicates caused by a change in practices to make publically available applications from the USPTO in 2001. Since USPTO non-granted applications are only (partly) present from 2001 onwards, we exclude all non-granted USPTO applications. For applications filed at the EPO and through the PCT-route, we also include non-granted patent applications and include a dummy-variable indicating whether an application was granted in order to control for a potential grant bias.

an invention draws from[6]. For classification of technological knowledge, we use the International Patent Classification (IPC) system which has the goal to provide a detailed account of technological content provided in patents to establish an effective search tool for retrieving patent documents. For our measures we will make use of the 'group level' consisting of about 8000 different technological classes. For classification of cited scientific references, we rely on the Thomson Reuters Web Of Science (WOS) categories[7] of journals into scientific domains.

4.2. Illustrating the construction of our novelty indicators: the "Oncomouse"

We illustrate the construction of the indicators using the example of the "Oncomouse" (also known as the "Harvard Mouse"). Developed by molecular geneticists Philip Leder and Thimothy Stewart from Harvard University, the oncomouse was the first patented, genetically altered animal (Arts et al., 2013). The mouse is genetically modified to carry an activated oncogene, which makes it highly susceptible to cancer and therefore most suitable for cancer research. Leder and Stewart were the first ones to isolate a gene that causes cancers and inject it into fertilized mouse eggs to develop a new breed of genetically altered mice highly susceptible to cancer. As such, they were the first to apply the principle of genetically modifying an animal to cancer research. Their invention had large impact on cancer research, and follow-on inventions applied similar techniques to 'construct' suitable animal models for human diseases (e.g. "Xenomouse" or "Alzheimer mouse") (Arts et al., 2013).

Table 1 provides an overview of how we construct our novelty measures, using the information on the patent family containing patent document US 4736866. The "**Novelty in Recombination**" indicator is constructed using the IPC groups assigned to the patent applications in the family. The oncomouse patent is assigned to 8 different groups leading to 28 pairwise combinations (examples of which are given in the upper part of table 1). We compare each of these 28 pairs to all pairs in existence by patents applied for in years previous to the first application year of a patent in the family (in this case: 1985). The oncomouse patent family makes 4 IPC-combinations not previously appearing in any patent family. The new combinations made include the IPC groups for 'new breeds of animals' (A01K 67), 'introducing […] materials into […] the body of animals' (A61D 7) and '[…] instruments or methods for reproduction or

---

[6] We refer to the European Patent Convention (EPC) (see http://documents.epo.org/projects/babylon/eponet.nsf/0/00E0CD7FD461C0D5C1257C060050C376/$File/EPC_15th_edition_2013.pdf) and the Manual of Patent Examining Procedure (MPEP) (see http://www.uspto.gov/web/offices/pac/mpep/index.html) for a detailed account of practices regarding prior art examination in the different patent offices.

[7] Categorization of scientific journals into scientific disciplines (249) using field tag 'WC' in the Thomson Reuters Web of Science database. For more information, see http://images.webofknowledge.com/WOKRS56B5/help/WOS/hp_subject_category_terms_tasca.html.

fertilisation […]’ (A61D 19) as well as ‘[…] DNA or RNA concerning genetic engineering […]’ (C07H 21) and ‘[…] genetic engineering processes for obtaining peptides […]’ (C07K 14). These combinations indeed reflect the novelty in the combination of principles and components which characterizes this invention.

The “**Novel Technological Origins”** indicator looks at “backward citation combinations” and assesses whether they were used for the first time. The oncomouse patent cites two patents, linked to in total 8 IPC groups. To construct the indicator, we make combinations between the IPC groups on the oncomouse patents, and the IPC-groups of its cited patent references. The patent makes 64 citation combinations, 10 of which have never occurred before as citation combinations of previous patent applications. Mainly the new combinations are made by the formerly mentioned IPC-codes (A61D 7 and A61K 67) and their link to genetic engineering such as class C12P 21 (‘[…] preparation of peptides or proteins […]’) and C12N 15 (‘[…] use of medicinal preparations containing genetic material which is inserted into cells of the living body […]’. Indeed, knowledge around genes and genetic engineering was building up. The oncomouse was the first to apply the principle from that domain for the purpose of cancer research.

For “**Novel Scientific Origins**” we apply a similar method, but use the WOS subject categories of the scientific references to make the citation combinations. The patent makes 40 such combinations, of which 21 had never been connected before. The patent relies for the first time on knowledge in the field of biochemistry and cell biology to make possible its recombination (as represented by the classes linked to animals and genetic engineering processes).

The Oncomouse patent family: US4736866, EP169672 , CA1341442, DE3586020, JP5048093, JP61081743, JP2058915

IPC groups (Exhaustive)			**(Examples of) Combinations**	**First Occurrence?**	**Indicator**
A01K 67			A61D 19 - C07H 21	Yes	
A61D 7			A01K 67 - C07H 21	Yes	
A61D 19			A61D 7 - C07H 21	Yes	
C07H 21			A61D 19 - C07K 14	Yes	
C07K 14			A01K 67 - A61D 7	No	NR=1
C12N 5			A01K 67 - A61D 19	No	
C12N 15			A01K 67 - C07H 21	No	
G01N 33			…	No	
				Total=4	
	References	**Classification of References**			
	Patents	**IPC groups**			
	US4579821	C12N 15	A01K 67 - C12N 15	Yes	
			A61D 7 - C12N 15	Yes	
			A61D 19 - C12N 15	No	
	US4535058	C12P 21	A01K 67 - C12P 21	No	
		A61K 39	A01K 67 - A61K 39	No	NTO=1
		C12Q 1	A01K 67 - C12Q 1	Yes	
		C07K 19	…		
		C12N 15	A61D 7 - C12P 21	Yes	
		C07K 14	A61D 7 - A61K 39	No	
		C07K 16	A61D 7 - C12Q 1	Yes	
		G01N 33	…	…	
				Total=10	
	References	**Classification of References**			
	Scientific	**Journal (Subject Category)**			
	Binster et al. (1983) Nature 306, 332 336	NATURE (RO: Multidisciplinary Sciences)	A01K 67 – RO	No	
	Blair et al, Science 212:941 943, 1981	CELL (CQ: Bio-chemistry and molecular biology, DR: Cell biology)	A01K 67 – CQ	No	
	…	JOURNAL OF MOLECULAR BIOLOGY (CO: Bio-chemical research methods)	A01K 67 – DR	Yes	NSO=1
		BIOCHEMISTRY (CO)	A01K 67 – CO	Yes	
		NEUROSCIENCE (RU: Neurosciences)	…	…	
		…	…	…	
				Total=21	

Table 1: Illustration of construction of the indicators based on the Oncomouse patent family

4.3. Descriptive statistics

In a first step, we present descriptive statistics of the indicators proposed for each novelty dimension: the frequency of occurrence of novelty in recombination and novelty in scientific/technological origins, the intensity of novelty (as reflected in the number of new combinations being made) and the communality between the different novelty indicators. We particularly want to see how skewed the phenomenon of novelty is, as measured by our constructs. Scholars studying the evolution of technology note that the bulk of inventions are incremental improvements along trajectories that are only occasionally interrupted by discontinuous steps (Dosi, 1982; Baumol, 2004; Arthur, 2009). We should thus expect only a small set of patents to be truly novel. We also expect our different dimensions of novelty, novelty in recombination and novelty in knowledge origins, to be correlated but nevertheless sufficiently distinct.

4.3.1. Frequency of occurrence of technological novelty

Frequencies of patent families	**# undefined**	**# scoring 0**		**# scoring at least 1**	**Total**
		No combinations made	**Combinations made**		
Novelty in Recombination (NR)	1 697 912 (32.05 %)	0	3 339 596 (63.04 %)	259 775 (4.90 %)	5 297 283
Novelty in Technological Origins (NTO)	0	141 757 (2.68 %)	4 023 549 (75.95 %)	1 131 977 (21.37 %)	5 297 283
Novelty in Scientific Origins (NSO)	0	4 551 880 (85.93 %)	673 457 (12.71 %)	71 946 (1.36 %)	5 297 283

Table 2: Distributions of patent families by scoring on the indicators and by making (citation) combinations

A first important point to make is that a substantial number of patents belong to only 1 IPC group and therefore cannot score on our Novel Recombination indicator. Conceptually, we do not allow for the possibility that a technology is not recombinant in nature. Hence, we assume that this problem occurs because within the single IPC-code assigned, there are different components or principles used for the technology[8]. In our analyses, we treat these cases as missing values when we analyze the effects of the NR

[8] Going to a more disaggregated IPC group level does not solve the problem: when employing the IPC subgroups (lowest level of aggregation: 69 884 classes), still about 21 percent only belong to 1 IPC-code

indicator in separation from the other indicators. For the novelty in knowledge origins indicators (NTO and NSO) making no combinations means that these patents do not contain any backward references. This happens seldom for NTO[9], but it is common for NSO, as about 86% of the sample patents have no scientific reference. Since this might be a true reflection of not drawing on any (novel) knowledge, we do not exclude these cases. In the multivariate analysis, we control for whether or not and how many backward citation combinations are being made.

Frequencies of patent families	**# scoring 1**	**# scoring 2 – 5**	**# scoring >5**	**Modus**	**Median**	**Mean**	**Total Scoring >0**
Novelty in Recombination (NR)	146 026 (56.21 %)	96 375 (37.10 %)	17 374 (6.69 %)	1	1	2.29	259 775
Novelty in Technological Origins (NTO)	409 928 (36.21 %)	479 716 (42.38 %)	242 333 (21.41 %)	1	2	4.58	1 131 977
Novelty in Scientific Origins (NSO)	32 354 (44.97 %)	31 640 (43,98 %)	7 952 (11.05 %)	1	2	2.85	71 946

Table 3: Conditional on scoring on the indicator, the distributions by number of new (citation) combinations made

260 000 patent families (about 7 percent of those that make combinations) make at least one new combination of technology classes and can therefore be classified as having **Novelty in Recombination** (NR) (Table 2). This corroborates the skewedness of the novelty phenomenon. Of these NR patents, the majority make only 1 new combination, being also the mode and median (Table 3). Only 7% of NR patents (or less than 0.5% of those that make combinations), make more than 5 new combinations. It therefore seems that most of the heterogeneity in novelty in recombination lies in whether or not there is a novel combination, rather than in the number of novel combinations being made.

Most of the novelty in knowledge origins is driven by novelty in technological origins. When looking at **Novelty in Technological Origins (NTO)**, a less skewed novelty phenomenon is picked up. More than 1 million patent families, i.e. 22% of all patent families, make at least 1 new combination between citing and cited technology class. Although also for novel technological origins, one new combination is the most frequent case, the mean number of new combinations is above 4, more than 20% of NTO patent families

[9] Not drawing on previous technologies is seen by some as a characteristic of radicalness (see f.e. Banerjee & Cole, 2011),

have at least 5 new combinations. NTO is therefore more frequent and with more heterogeneity in intensity than NR and NSO.

When looking at **Novelty in Scientific Origins (NSO),** most of the patent families cite no scientific references as prior art and can therefore not make new combinations. Of those few who do, only 72 000 patent families (or 11%) use at least one new scientific origin, i.e. refer to a scientific publication as prior art from at least one scientific class previously not being cited by its technology class (Table 2). If new scientific origins are being sourced (Table 3), this is most often only 1 new combination (modus).

4.3.2. Communality between novelty in recombination and novelty in knowledge origins

Since we outlined novelty in recombination and novelty in technological and scientific origins as distinct, but related dimensions of technological novelty, it is interesting to look at their observed communality. Figure 1 shows the frequency of occurrence across the three novelty dimensions while Table 4 summarizes a series of contingency tables cross-classifying the observations according to whether or not they score on (a combination of) the indicators proposed.

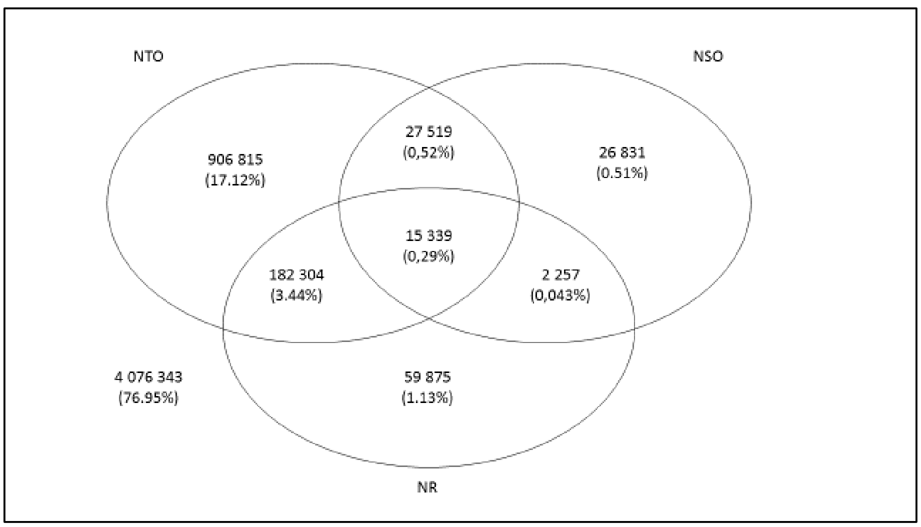

Figure 1: Patent families scoring on different novelty dimensions

About 25% of all patent families score at least once on one of our three novelty indicators. Within the patent families with novelty, 80% score on only one novelty indicator, most typically NTO; 19% score on 2 dimensions. Only 15 339 patent families, i.e. less than 1%, score on all 3 dimensions. Combining NR, NTO and NSO is therefore a very skewed phenomenon.

	NR=1&NTO=1&NSO=1	NR=1&(NTO=1|NSO=1)	NR=1&NTO=1&NSO=0	NR=1&NS0=1&NTO=0	NTO=1&NSO=1&NR=0
Observed frequency	15 339	199 900	182 304	2 257	27 519
Observed frequency relative to expected frequency	5.19***	2.80***	2.88***	3.85***	2.73***

Table 4: Summary of results from contingency tables.

Note: The columns including Novelty in Recombination (NR) require at least 1 combination made on the patent. *** indicates a p-value for the Chi-sq test <0.001

Although most novel patents score on only 1 novelty indicator, the different novelty concepts are significantly correlated. Scoring on all three novelty dimensions is about 5 times as likely compared to what is expected under the hypothesis of being unrelated (p-value chi-squared test <0.001). Patent families with novelty in recombination are significantly more likely to have novelty in knowledge origins and vice versa: 199 900 patent families, or, three out of 4 NR patents, also have novelty in knowledge origins, technological and/or scientific. This is 2.80 times more than expected. Nevertheless, that still leaves 80% of patent families with novelty in technology origins to have no novelty in recombination and 90% of patent families with novelty in scientific origins to have no novelty in recombination. This confirms that although NR and NTO and NSO are significantly correlated, they capture different dimensions of novelty.

Overall, these first statistics confirm the skewed nature of technological novelty: relatively few patent families display technological novelty as measured through our indicators. Particularly very few patent families combine novelty in recombination with novelty in knowledge origins. The descriptive statistics show that although the different dimensions of novelty are significantly correlated, a substantial number of novel patent families score on only 1 dimension, confirming that each of the measures is related but distinct in the information it conveys.

In the remainder of the analysis, we will use each of the indicators for technological novelty (NR, NTO, NSO). As most of the heterogeneity is in whether or not a novel combination is made, rather than in the number of new combinations, the analysis will be on the dichotomous distinction. To capture the distinct but nevertheless correlated nature of the different dimensions, we will also work with the 7 exclusive combined categories (**Only NR; Only NTO; Only NSO; Only NR&NTO; Only NR&NSO; Only NTO&NSO; NR&NTO&NSO**); We are particularly interested in the few cases that combine NR, NTO and NSO **(NR&NTO&NSO)**.

4.3.3. Relationship with other ex ante patent indicators

The literature commonly uses other related indicators characterizing patents based on backward citation information. Trajtenberg et al. (1997) measure the spread of backward citations over technology classes in their "originality" measure, arguing that higher scores of "originality" reflect inventions characterized by 'synthesis of divergent ideas', and thus 'basicness'. This measure of spread in knowledge sourcing could be expected to be associated with technological novelty, but it does not identify novelty per se. Using a more diverse knowledge sourcing strategy might not lead to necessarily novel approaches, for instance when it is customary for the field to be sourcing more broadly. The approach taken by Shane (2001) to characterize "radicalness" and by Rosenkopf & Nerkar (2001) to measure 'technological boundary spanning' is to look at the degree to which an invention 'sources in' knowledge from other fields than its own, by counting the number of patent classes in the focal patent's backward citations to which the patent itself does not belong (or the number of backward citations that fulfill this requirement). Again, sourcing outside one's own technology area may be a characteristic which novel patents are more likely to display, but it does not identify them. Citing from 'outside' fields of knowledge is not a sufficient condition to actually apply a *novel* approach since a large number of previous patents might have already sourced knowledge from these 'outside' fields before. Although the ex ante characteristics of spread in sourcing and external sourcing can be expected to be positively correlated with novelty, we expect our novelty indicators to convey different information about inventions. Hence, we expect a moderately positive relation between both indicators and the indicators proposed in this paper. In what follows, we look at how both measures for "originality" and "radicalness" i.e. measuring whether sourcing is more spread or external, correlate with our measures of technology novelty.

	NR	NTO	NSO	NR&NTO&NSO	NR|NTO|NSO	External Sourcing
External Sourcing	0.0934*	0.3552*	0.0744*	0.0497*	0.3434*	1
Spread in Sourcing	0.0980*	0.2958*	0.0600*	0.0406*	0.2934*	0.7310*

Table 5: Correlations between related measures and our novelty indicators. * indicates p-values of <0.05 in significance tests of pairwise correlation coefficients

As shown in table 5, there is a significant positive correlation both for "external sourcing" and "spread in sourcing" with our novelty indicators. This correlation holds most strongly for novelty in technological origins, which is not surprising as all of these indicators are based on patent references. Note that, as diversity of sourcing is associated with venturing outside one's own domain, both indicators are highly correlated (correlation coefficient of 0.73).

		NR|NTO|NSO	Only NR	Only NTO	Only NSO	NR&NTO&NSO
	Occurrence in total Sample	24%	1.1%	18%	0,5%	0.4%
External Sourcing-4th quartile	*Occurrence in all 4th quartile patents*	45%	0.8%	34.5%	0.6%	0.8%
	Observed occurrence/Expected occurrence	1.91***	0.73***	1.97***	1.22***	2.59***
Spread in Sourcing-4th quartile	*Occurrence in all 4th quartile patents*	46%	1.2%	34%	0.6%	0.8%
	Observed occurrence/Expected occurrence	1.95***	1.07***	1.94***	1.24***	2.75***

Table 6: Results from contingency analyses cross-tabulating related measures and our measures.

Note: First row is the share of patents in the 4th quartile on the alternative indicator which score on our novelty indicators; Second row is the ratio of the observed to expected frequency of occurrence. *** indicates a p-value for the Chi-sq test <0.001

As shown in table 6, a total of 9 773 patent families score on all of our three novelty indicators and are also in the top quartile on the External Sourcing index, an observed frequency that is 2.59 times larger than expected. In the other quartiles of the external sourcing measure, the observed frequencies of co-occurrence are all lower than expected (not reported). Similar results arise for the Spread in Sourcing index: co-occurrence of scoring on all three novelty indicators and belonging to the 4th quartile for spread in sourcing occurs 2.75 times more often than expected.

Nevertheless, despite these positive correlations, most of the inventions fall in the off-diagonal cases. Somewhat more than half of the patent families in the top quartile of scoring on External Sourcing and Spread in Sourcing do not score on any of our novelty indicators. More than 99% of all patent families in the top quartile of these distributions do not score on all three of our novelty indicators. When looking at the individual novelty indicators, the correlation is weakest with the novelty in recombination indicator, confirming the correlation analysis. In fact, inventions that only score on the NR indicator, are less likely to be in the top quartile for external sourcing.

Overall, the analysis shows that other backward-citation based indicators, measuring the diverse and external nature of technology sourcing, pick up different dimensions compared to our novelty indicators. Nevertheless, the analysis also shows that technologically novel patents are sourcing more outside their own field and have a more diverse sourcing of technological knowledge.

5. Validation exercises

How valid are our indicators to proxy for technological novelty? A first simple exercise is to look at some cases of patented inventions that are well known as novel inventions. The “onco-mouse” used in section 3 to illustrate our indicator construction is already one such example. It scores positively on all three novelty indicators. It is therefore an illustration of the highly skewed case of combining all dimensions of technological novelty. But there are others, listed in Box 1.

These examples, being all “famous” cases suffer from an ex post success bias. As we also want to see whether our indicators are able to pick up cases of ex ante novelty, which may not have been ex post “famous”, we perform two larger validation exercises.

First, we use a set of inventions that were awarded an R&D prize by ‘R&D Magazine’ for being among the most technologically significant inventions of the year. These prizes are known as the “Oscars of Inventions”. Other studies using these awards include Carpenter et al., 1981; Scherer, 1989; Block & Keller, 2009 and Fontana et al., 2012. They are awarded after review by a panel of technical experts and reflect an almost contemporaneous assessment of the inventions considered since the goal is to award the top technology products in a given year. Every year the US-based magazine selects about 100 winners out of all applicants. Although there is no strict novelty requirement as selection criterion stipulated, we expect the patents scoring on our indicators to be more likely to be among the awarded inventions.

Second, we use a sample of patents that, after examination by patent officers, were refused by the European Patent Office (EPO) because they lacked novelty or inventive step. We expect the patents scoring on our indicators to be less likely to be among the inventions being refused a patent for lack of novelty. Both external validation exercises are informative, but cannot be conclusive, as both of them are not explicit validations of the concepts of technological novelty as identified by us.

BOX 1: Some famous inventions and their novelty scores.

Polymerase Chain Reaction (PCR). In 1985, Kary Mullis and his team invented a technique for multiplying DNA sequences in vitro at Cetus Corporation in Berkeley, California (Arts et al., 2013). PCR is considered as one of the most revolutionary inventions in molecular biology in the 1980's. Cetus applied for a patent at the USPTO (US4683202) in 1985 and sold the patent to Hoffman-La Roche, Inc for $300 million. When looking at the novelty scores of the patent family containing this patent, it is apparent that it is among the very skew of the novelty distribution as assessed by our indicators. It scores on Novelty in Recombination with 8 new combinations (out of 28 combinations made), on Novel Technological Origins with 27 new combinations (out of 136 combinations made) and on New Scientific Origins with 8 new combinations (out of 24 combinations made).

Rapid Prototyping. As suggested by Shellabear and Nyrhillä (2004), Chuck Hull was the first to propose 'layerwise building' of a three-dimensional object (popularly known as '3D printing') using a technique called Selective Laser Sintering (SLS). He commercialized this technology through the company '3D systems', and EOS bought the patent portfolio of 3D systems in 1997. One of these patents is US4575330, applied for in 1984. When looking at the indicator scores of its family, we observe it scores on all three of our indicators (NR, 31 new combinations out of 91; NTO, 86 new combinations out of 406 and NSO, 22 out of 56).

HIV Protease-inhibitors. Pioneered between 1989 and 1994 by several inventors at Hoffmann – La Roche, Abbott Laboratories and Merck & Co, protease-inhibitors are used to treat patients with AIDS by lowering the viral load carried by patients. Although Hoffmann – La Roche launched the first product in December, 1995 ('Invirase'), the first patent describing an HIV protease inhibitor was filed by Merck & Co in 1989 (EP0337714). This patent led to the introduction of Merck's protease-inhibitor product (called 'Indinavir'). The patent family scores heavily on all three novelty indicators introduced (NR, 13 new combinations out of 435; NTO, 33 out of 1410 and NSO, 34 out of 90).

Lab-on-a-Chip. Michael Ramsey and Stephen Jacobson received an R&D100 Award (cfr. infra) for developing a chip which is able to perform rapid, high-resolution chemical separations of e.g. amino acids, peptides and proteins. This technology enabled bio-chemical researchers to perform separations 10 to 100 times faster than conventional methods using sample volumes that are 100 to 10000 times smaller (Jacobson et al., 1994)[6]. The first patent describing the technology (US5858195, 1995) was used to calculate the indicators. Again, this breakthrough technology scores on all indicators proposed (NR, 4 new combinations out of 55; NTO, 49 out of 484, and NSO, 1 out of 55.

[10] See also http://www.ornl.gov/our-people/corporate-fellows/j--michael-ramsey

5.1. R&D100 Awards

5.1.1. Data and methodology

To obtain the information on the award-winners, we collected data from the R&D100 awards from 2000 until 2007 from the R&D magazine webpage[11]. We manually extracted the names of the inventors and organizations provided by the webpage (about 50 percent of the prizes contain information on the name of one or more inventors) and the description of the prize. We perform a name matching of the inventors using a database containing disambiguated inventor names (Li et al., 2014), and a name matching of the organizations with PATSTAT applicants (Du Plessis et al., 2010). We require that at least one of the inventors and at least one of the associated organizations are linked to a patent with an application year ranging between the year of the prize and previous 5 years. If multiple patents satisfy this requirement, we perform a text similarity analysis between the abstracts of the patents and the description given with the invention (number of words in common, normalized by the length of both texts) and select the patent having the largest similarity score. After these steps we are left with a sample of 196 patent families linked to an award. As a control group, we select all patents having IPC subclasses (4 digit), all occurring within the list of subclasses of the prize winners during the time period defined by the application years of the prize-winning patents. Moreover, we restrict the sample to all patent families applied for by firms that received the prize at least once in our sample. This ensures that all patents in the sample were filed by a company that knew about the existence of the prize. This leaves a control group of 57 590 patent families.

In the descriptive analysis, we perform cross-tabulations of scoring on the novelty indicators and being among the award-winning patents to look for associations. In the econometric analysis, we perform logistic regressions predicting the probability to be among the award winners by scoring on our indicators, controlling for a number of characteristics of the inventions other than novelty that may influence the likelihood to win the award.

5.1.2. Descriptives

The contingency analysis shows that 84 R&D prize patents out of the total 196 (or 43%) score on at least one of our novelty indicators. This co-occurrence is 3 times higher than expected. 76 of them score on novelty in technological origins, 12 on novelty in scientific origins and 25 on novelty in recombination. Although only 5 out of the 196 R&D prize patents score on all of our 3 indicators (NR&NTO&NSO), this is more than 8 times the expected frequency.

[11] http://www.rd100awards.com/rd-100-awards-history . Last accessed on 09/12/2014

196 R&D prize winners	**NR or NTO or NSO**	**NTO**	**NR**	**NSO**	**NR & NTO & NSO**
Occurrence indicator	14.03%	12.92%	3.09%	1.55%	0.33%
Observed co-occurrence	84	76	25	12	5
Observed/expected co-occurrence	3.06***	3.01***	5.56***	4.00***	8.33***
Observed co-occurrence as share of all Prize winners	42.9%	38.78%	17.12%	6.12%	2.55%

Table 7: Results cross-tabulations of winning an R&D100-award and the novelty indicators.

Note column 3: Because of undefined observations for the NR indicator, only 146 patents are in the award-winning group, and 43 305 in the control group.

5.1.2. Results logit regressions

The regression analysis largely confirms these descriptive statistics. The regression analysis predicts the likelihood to be an R&D prize winner, based on the novelty characteristics of the patent family together with a set of control variables. Control variables used in the analyses are: dummies for the application year of the first patent in the family, number of combinations made for each of the dimensions, for each of the indicators a dummy with value 1 if the patent did not make any combination, dummies indicating whether the patent family, in addition to its US application, has a member in the EPO and whether an application through the PCT-route was filed, a dummy indicating whether at least 1 of its IPC groups was a newly introduced class after 1980, number of applications filed for in the family (DOCDB definition), and dummies for belonging to each of 35 technological areas ('fhg-codes' as described by Schmoch (2008))[12].

[12] These controls are selected based on the criterion of being naturally correlated to the measures introduced. **Application year dummies**: Because the number of pairwise combinations that have never been used before decreases over time, the measures have a downward sloping trend over time. **Number of combinations made** and **dummies no combinations made**: because some patent applications are assigned to more IPC-codes than others (because the classification is more fine-grained or because it involves a larger number of components), the measures naturally correlate to the number of combinations made for each dimension. Moreover, some patents can by definition not score on the indicators because they do not make any combination. **Application office dummies**: Because classification and citation practices might differ between offices (for example, the USPTO uses a concordance scheme to translate USPC classes in IPC classes and USPTO patents usually have more backward references), the measures correlate to the membership in different offices. **Dummy IPC newly introduced after 1980**: Since the IPC is reclassified periodically, new classes can be added. Because the World Intellectual Property Organisation (WIPO) states all previous patents are reclassified when this happens, this should not be a major concern. Still, a reclassification can introduce artificially new pairs, because the process is automated. To mitigate this concern, we include a dummy which takes the value of 1 if the patent contains any class that was introduced after 1980 (about 1.5 percent of the patent families score 1 on this dummy). **Number of applications in family**: Because some patents are applied for in multiple offices (and as such might increase the number of IPC-codes and citations related to the family), the number of applications in a family correlates to the indicators. **Technological**

The analysis includes as observations the set of prize winning patents together with the control group of patents from the same time period and IPC subclasses. The odds ratios (exponentiated coefficients) for the measures of interest are displayed in table 8. As our novelty indicators should be significantly more associated with R&D prize winners, we expect coefficients significantly larger than 1 for our novelty indicators in predicting R&D prize awards.

	(1)	(2)	(3)	(4)
Novelty in Recombination (NR)	1.790*			
	(0.500)			
Novelty in Scientific Origins (NSO)		1.076		
		(0.370)		
Novelty in Technological Origins (NTO)			1.580*	
			(0.299)	
NR only				4.756**
				(2.277)
NTO only				1.698*
				(0.349)
NSO only				1.618
				(0.981)
NR and NTO				2.113*
				(0.734)
NTO and NSO				1.344
				(0.774)
NR and NTO and NSO				2.205
				(1.119)
Pseudo R-squared	0.164	0.143	0.145	0.149
N	43301	57583	57583	57583

Table 8: Results from logistic regressions on novelty patents winning an R&D100 award

Note: Controls included (but not reported): application year dummies, dummies for deciles on #combinations made, dummies for no combinations made, # applications in family, EPO filing in family, PCT filing in family, technological area dummies, dummy IPC newly introduced after 1980. Robust standard errors in parentheses, * p<0.05, ** p<0.01, *** p<0.001

---

area dummies: Because technological areas might differ from each other in terms of patenting density and classification and citation practices, we observe significant variations in scoring on the indicators depending on the technological area belonging to.

Novelty in recombination (NR), as well as novelty in technological origins (NTO) have a significant positive predictive power on winning an R&D award. Novelty in scientific origins (NSO) seems to be an overly skewed phenomenon to give significant results[13]. The positive predictive power of NR and NTO comes mostly from those novel patents with 1 new combination (results not shown). When looking at the exclusive categories of novelty, we find that all categories including NR and/or NTO score significantly higher than one, with the highest significant coefficients for the categories including NR[14]. Although the descriptive statistics gave much less significance to NR as a predictor of an R&D prize, the regression analysis controlling for other influencing factors seems to suggest that it is particularly novelty in recombination that is most powerful in predicting R&D prizes, solely or in combination with novelty of technological origins. When using the measures for spread of sourcing (Trajtenberg, 1997) or the measure of external sourcing (Shane, 2001; Rosenkopf & Nerkar, 2001) as predictor of R&D prizes, together with the same set of controls, none of these indicators have a significant power to predict an R&D prize. When including these indicators next to our novelty indicators, they still remain insignificant while our novelty indicators remain significantly associated with higher R&D prize winning probabilities (results not shown).

Overall, although the R&D prize winning analysis cannot provide definite validation of our indicators, it does show a significant correlation in support of our measures.

5.2. Patent refusals

As a second validation exercise we look at a group of patents which were refused because the patent examiner did not deem the invention inventive enough for a patent to be granted. When such a patent application would score on our indicators, it would indicate that the new combination(s) identified by our indicators do not reflect real novelty. This might happen when the classifications of the focal patent, or its citations, do not precisely reflect the invention's functioning.

[13] Only 24.2% of patent families in this sample cite a scientific source (and hence make an IPC-scientific field combination). Belonging to this group (which is included as control variable in the analysis) positively affects the probability of winning the award. Only 1,8% of patent families in the sample score on NSO, 12 R&D prize winning patent families score on NSO.
[14] The combination of NR&NTO&NSO is significant at the 12% level.

5.1.1. Data and methodology

We rely on the EPO worldwide legal status database (INPADOC)[15] to select a random sample (about one third) of refused patents in the fields of Biotechnology and Digital Communication Processes (as defined by Schmoch, 2008). Since information on reasons for refusal[16] is not available in a structured form, we manually identify the reason for refusal using the EPO website (using the European Patent Register (EPR) files[17]) using the documents provided in the EPR[18]. We use the patent applications that were not granted because of 'lack of novelty' or 'lack of inventive step' (the concepts related to technological novelty). As such, we identified 390 patent applications refused because of lack of novelty or inventive step. These are all applications filed at the EPO, but we perform analyses on the family level (as before). The comparison group consists of all granted EPO patents (again, we expand to the family level) in the fields of Biotechnology and Digital Communication Processes. We perform contingency analyses in which we cross-tabulate being among the refused patents with scoring on the indicators of novelty.

5.2.2 Descriptives

The contingency analysis shows that only 19 out of the total 390 refused patents score on either one of our novelty indicators, a ratio that is significantly lower than expected. Only 3 patents score on the combined category. A multivariate analysis (results not reported) confirms the descriptive analysis with patents scoring on novelty indicators significantly less likely to be refused because of lack of novelty/inventive step.

Of all 390 patent refusals	**NR or NTO or NSO**	**NTO**	**NR**	**NSO**	**NR & NTO & NSO**
Observed co-occurrence	19	17	5	7	3
Observed/expected co-occurrence	0.25***	0.28***	0.21***	0.42*	0.55

Table 9: Results from cross-tabulations between the indicators and being refused for lack of novelty/inventive step.

Note column 3: because of undefined observations for the NR indicator, only 339 patents are in the refusal group, and 59 524 in the control group.

[15] See http://www.epo.org/searching/subscription/raw/product-14-11.html for more information

[16] We refer to the European Patent Convention for a detailed account of reasons for refusal: http://www.epo.org/law-practice/legal-texts/epc.html

[17] See http://www.epo.org/searching/free/register.html

[18] The coding was performed by 2 students, independent cross-checks resulted in a 100% overlap in terms of the conclusion.

In summary, the validation exercises show the expected patterns when comparing our novelty indicators with externally assessed information on technological novelty of inventions. They also confirm that each of the indicators, NR, NTO and NSO, have a distinct profile, but that especially the more skewed cases of novelty in recombination combined with novelty in sourcing indicate technological novelty.

6. Analyses on ex post technological impact

Further validation of our constructs should come from the use of the indicators in analysis. In this contribution, we look at the technological impact generated by patent families identified as technologically novel. As argued in the introductory section, technological novelty is seen as a source of uncertainty on the impact which the invention will have. Novel inventions should, therefore, show a larger variability in technological impact. Following Fleming (2001), we use generalized negative binomial models to analyze both the average and dispersion of forward citation counts to assess the uncertainty surrounding the impact of inventions. Moreover, since novelty is often argued to be the ultimate source of breakthrough invention, we analyze the relationship between our novelty indicators and the probability of being among the outliers in terms of forward citations. Confirmation of these suppositions may strengthen our confidence in the constructed novelty indicators as being associated with pertinent features of radical inventions.

6.1. Measures for technological impact

We use the number of patent citations received as a proxy of technological impact. To calculate the number of forward citations received, we apply a full family correction using the DOCDB family definition (Martinez, 2011; Bakker et al., 2014). This implies that we count the number of distinct families that cite at least one family member of the focal application. To allow for a minimum amount of time for the patents to be cited, we only use patent families up until the year 2003. For the count models, we use a fixed window of 7 years. To calculate the outlier performance measures, we construct dummies which take a value of 1 if the patent family is among the top 2 standard deviation (respectively 5 standard deviation) outliers in the distribution of forward citations of at least one of its technological classes and in the same application year. In total there are 120 366 patent families (respectively 18 229) in our sample which qualify for being 2 SD (respectively 5 SD) outliers in impact (i.e. about 3%, respectively 0.5%, of the total sample).

6.2. Results average and dispersion forward citation counts

Table 10 shows that, for every novelty indicator, novel patent families on average receive significantly more forward citations compared to non-scoring patent families. However, the differences are relatively small for the NR and NTO indicators (about half a citation). Interestingly, the premium amounts to more than 4 citations on average for the NSO indicator. The highest premium however is found for patents scoring on all novelty indicators. This is also the category that has the highest standard deviation in citations received.

	NR or NTO or NSO		NTO		NR		NSO		NR & NTO & NSO	
	0	1	0	1	0	1	0	1	0	1
Mean	6.03	6.62	6.05	6.61	6.60	6.66	6.11	10.74	6.17	12.96
S.D.	9.83	11.02	9.91	10.89	10.71	11.57	9.93	18.99	10.09	21.72
N	2 663 975	942 938	2 716 079	890 834	2 170 336	172 210	3 547 092	59 821	3 594 442	12 471

Table 10: Descriptive statistics of Forward Citation counts by scoring on different novelty indicators.

Note: P-values of independent sample t-tests of difference in means: <0.001.

Table 11 shows the coefficients of interest resulting from the generalized negative binomial count models estimating the effects on the mean (left) and dispersion (right). The analysis includes a similar set of controls as used supra[19]. Scoring on the NR or NTO indicators (models 1 and 3) significantly decreases the average, but increases the variability in the number of forward citations received. This is not true for the NSO indicator (model 2), which shows a positive effect on the average, with no significant difference in dispersion in number forward citations received.

When looking at the combined categories for the three indicators, we see that the only novelty inventions that have a significantly higher average impact score are the inventions that combine NR, NTO and NSO. All the other novelty categories have a lower (or non-significantly different) average impact factor. At the same time however, inventions that combine NR, NTO and NSO have a significantly higher dispersion in technology impact. A significantly higher dispersion holds also for the other categories of novelty, with the exception of the NSO only and the NR&NSO category.

The positive effect of NSO on the average as reported in model 2 is therefore only present from those NSO inventions which are also scoring on both NR and NTO. Scoring on NSO only, or NSO and one of the other

[19] Application year dummies, number of combinations made for each the dimensions, 3 dummies with value 1 if the patent did not make any combination for all three indicators, 3 dummies indicating whether the patent family has a member in the USPTO, the EPO and whether an application through the PCT-route was filed , dummy indicating whether at least 1 family member was granted, number of applications filed for in the family (DOCDB definition), dummy for IPC newly introduced after 1980, and dummies for belonging to each of 35 technological areas (‘fhg-codes’ as described by Schmoch (2008)).

indicators leads to a lower average. Scoring on only NSO decreases the dispersion, while combining NSO with NR has no effect on the dispersion, and combining NSO with NTO or both NTO and NR increases the dispersion. These mixed results explain the overall non-significant difference in dispersion effect of NSO reported in model 2.

	(1)		(2)		(3)		(4)	
Effects on:	***mean***	***dispersion***	***mean***	***dispersion***	***mean***	***dispersion***	***mean***	***dispersion***
Novelty in Recombination (NR)	-0.0957***	0.0942***						
	(0.00346)	(0.00581)						
Novelty in Scientific Origins (NSO)			0.0469***	0.00740				
			(0.00577)	(0.00851)				
Novelty in Technological Origins (NTO)					-0.127***	0.0419***		
					(0.00175)	(0.00298)		
NR only							-0.0905***	0.0651***
							(0.00807)	(0.0129)
NTO only							-0.129***	0.0321***
							(0.00181)	(0.00313)
NSO only							-0.0401***	-0.0293*
							(0.00860)	(0.0127)
NR and NTO							-0.173***	0.122***
							(0.00394)	(0.00660)
NR and NSO							-0.0988**	-0.00801
							(0.0344)	(0.0502)
NTO and NSO							-0.0153	0.0297*
							(0.00835)	(0.0123)
NR and NTO and NSO							0.0463***	0.0909***
							(0.0131)	(0.0191)
Pseudo R-squared	0.0419		0.0429		0.0432		0.0433	
N	2340607		3604932		3604932		3604932	

Table 11: Results from generalized negative binomial regressions estimating the effects of the indicators on the mean and dispersion parameter.

Note: Controls included for # applications in family, USPTO filing in family, EPO filing in family, PCT filing in family, Dummy at least 1 grant in family, Dummy IPC newly introduced after 1980, technology field dummies, application year dummies, dummies for deciles on #combinations made, dummies for no combinations made. * p<0.05, ** p<0.01, *** p<0.001

6.3. Results outlier performance

The higher variance in technology performance associated with most cases of technological novelty are suggestive of higher probabilities for outlier performance, and hence technological breakthroughs. We examine the probability for positive outlier performance explicitly in this section.

6.3.1 Descriptives

The descriptive contingency analysis (table 12) shows a clear and significant co-occurrence of outlier performance and technological novelty. About half of the 2 SD impact inventions score on at least one of our novelty indicators. Compared to all patents, this observed frequency occurs 86% more than expected. 2 SD impact inventions mostly score on new technological origins, but 6% have new scientific origins and this is almost 4 times more than expected. Most distinguishing is the scoring on combining all 3 novelty indicators, with an occurrence of over 6 times more often than expected.

Looking at the probability for novelty patents to be a breakthrough compared to non-novel patents (last row) shows that all novelty categories have a significantly higher likelihood to score on the 2 SD outlier than the average sample probability which is 3.3%. Particularly, 13% of patents with new scientific origins and 22% of patents with all 3 novelty indicators score within the 2 SD outlier performance.

When we look at a more skewed impact performance with 5 SD outliers, we find even stronger patterns of association with technological novelty: 8% of patents with all novelty indicators score within the 5 SD outlier performance: a frequency of co-occurrence which is 15 times higher than expected. When going still further in the outliers of impact performance, using 10 SD outliers, this leaves only 2968 patents. In this small set of patents 294 score on all 3 of our novelty indicators, which is a frequency of co-occurrence that is 28 times more than expected (not reported).

The contingency results are therefore very supportive of the breakthrough nature of impact of technologically novel patents, if it occurs. Particularly, the more skewed novelty profiles, particularly those combining all novelty indicators, have a significantly higher likelihood to obtain an exceptional outcome.

	NR or NTO or NSO *(26%)*	NTO *(25%)*	NR *(7%)*	NSO *(1.7%)*	NR & NTO & NSO *(0.3%)*
2 SD Outlier Forward Citations (3.3%)					
Observed/expected co-occurrence	1.86***	1.87***	2.11***	3.82***	6.48***
Observed co-occurrence as share of 2 SD Outliers patents	49%	46%	16%	6%	2%
Observed co-occurrence as share of Novelty patents	6%	6%	9%	13%	22%
5 SD Outlier Forward Citations (0,5%)					
Observed/expected co-occurrence	2.26***	2.28***	3.00***	7.03***	15.39***
Observed co-occurrence as share of 5 SD Outliers patents	59%	56%	22%	12%	5%
Observed co-occurrence as share of Novelty patents	1%	1%	2%	4%	8%

Table 12: Results cross-tabulation between 2 and 5 standard deviation outlier in terms of forward citation and the novelty indicators.

Note column 3: because of undefined observations for the NR indicator, only respectively 4.3 and 0.7 percent of the patents score among the 2 respectively 5 sd outlier measure**.**

6.3.2 Results logit regressions

A multivariate analysis, allowing to correct for other factors influencing technological impact[20], confirms these contingency descriptives. Table 13 shows the results from a logit analysis predicting the likelihood of a patent family to be a 2 (left panel) or 5 (right panel) SD forward citation outlier.

Each of our novelty indicator predicts a higher probability for breakthrough impact, with the effects most pronounced in the 5 SD case. This holds particularly for new scientific origins (NSO): patents with NSO have a 52% higher odds to be a 5 10SD forward citation outlier compared to non-NSO, all else equal. When looking at the exclusive categories, it are particularly the patents which combine all three novelty dimensions which have a higher probability for outlier performance. Patent families that combine all three novelty characteristics have 126% higher odds to be a 5 SD outlier. When distinguishing novelty patents based on the number of new combinations being made, the analysis shows a linear effect for each of the

[20] The analysis includes controls similar to the analyses for the R&D100 awards: Application year dummies, number of combinations made for each of the dimensions, 3 dummies with value 1 if the patent did not make any combination for all three indicators, 3 dummies indicating whether the patent family has a member in the USPTO, the EPO and whether an application through the PCT-route was filed , number of applications filed for in the family (DOCDB definition), dummy for IPC newly introduced after 1980, and dummies for belonging to each of 35 technological areas ('fhg-codes' as described by Schmoch (2008)).

indicators: patents that make more new combinations have a higher probability for outlier performance (results not shown).

These positive correlations between novelty and breakthrough performance remain highly significant when including the external and diverse nature of sourcing (results not reported). This confirms that our novelty characteristics are picking up a different critical characteristic of breakthrough impact as compared to the indicators used in the literature. The external nature of sourcing is also significantly positively related to breakthrough impact, especially the patents in the top quartile of this indicator. In contrast, the Spread in Sourcing indicator is significantly negatively related to breakthrough impact , particularly the patents in the top quartile of this indicator (results not reported).

	2 S.D. forward citation outlier				5 S.D. forward citation outlier			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Novelty in Recombination (NR)	1.103***				1.317***			
	(0.0121)				(0.0307)			
Novelty in Scientific Origins (NSO)		1.380***				1.531***		
		(0.0213)				(0.0449)		
Novelty in Technological Origins (NTO)			1.094***				1.326***	
			(0.00797)				(0.0240)	
NR only				1.066*				1.133°
				(0.0333)				(0.0843)
NTO only				1.067***				1.242***
				(0.00828)				(0.0246)
NSO only				1.181***				1.215**
				(0.0353)				(0.0827)
NR and NTO				1.102***				1.464***
				(0.0145)				(0.0431)
NR and NSO				1.331**				1.424°
				(0.130)				(0.276)
NTO and NSO				1.461***				1.736***
				(0.0315)				(0.0720)
NR and NTO and NSO				1.645***				2.255***
				(0.0429)				(0.0995)
Pseudo R-squared	0.134	0.141	0.141	0.142	0.177	0.190	0.190	0.191
N	2340607	3604932	3604932	3604932	2340607	3604932	3604932	3604932

Table 13: Results from logistic regressions with being forward citation outlier (left panel: 2 SD, right panel: 5 SD) as dependent, the indicators and controls as independent variables.

Note: Controls included for # applications in family, USPTO filing in family, EPO filing in family, PCT filing in family, Dummy at least 1 grant in family, Dummy IPC newly introduced after 1980, technology field dummies, application year dummies, dummies for deciles on #combinations made, dummies for no combinations made. Robust standard errors in parentheses, ° p<0.1 * p<0.05, ** p<0.01, *** p<0.001

7. Discussion

7.1 Main insights

This study builds further on the conceptualization of technologies as being recombinant in nature (Fleming, 2001; Hargadon, 2002; Arthur, 2007; 2009). Using this recombinant framework, we provide a more comprehensive measurement of technological novelty. We argue that identifying ex ante technological novelty characteristics of inventions may help our understanding of the origins and effects of radical inventions. Three (related) dimensions of technological novelty - Novelty in Recombination, Novelty in Technology Origins and Novelty in Scientific Origins are identified and operationalized relying on classification and citation information in patents. When operationalizing the indicators on all patented inventions since 1980, we find that technological novelty as measured by our indicators is a skewed phenomenon. Especially inventions combining novelty in recombination and novelty in knowledge origins are rare events.

After illustrating the indicators with a number of well-known novel inventions, we perform two validation exercises relating the indicators to external information on technological novelty. Results show that inventions identified as novel, are overrepresented among a set of award-winning inventions, and underrepresented among inventions that were refused a patent because of lack of novelty.

Comparing our novelty indicators to measures of related construct used in previous studies, shows they are associated with a higher spread in sourcing (Trajtenberg et al., 1997) and with higher levels of external sourcing (Shane, 2001; Rosenkopf & Nerkar, 2001). Nevertheless, they convey substantially different information on the ex ante characteristics of inventions.

When analyzing the ex post technological impact of inventions, we find that inventions characterized by technological novelty have a performance profile which has a significantly higher variance. This holds particularly for the inventions that combine all three dimensions of novelty. But it is only this category of novel patents that succeeds in getting a significantly higher average impact. All other novel patents have a lower average impact. The higher variance of performance for novel inventions finds its corollary in a higher likelihood for outlier performance. This higher probability for radical breakthrough performance holds particularly for the inventions which combine all three dimensions of novelty.

These first results from validating and analyzing the new indicators confirm our comprehensive framework for characterizing technological novelty as ex ante characteristic of radical inventions. Each of the indicators, novelty in recombination and novelty in scientific/technological origins, captures a distinct, but nevertheless related dimension of novelty. Especially inventions that combine novelty dimensions are candidates for radical breakthrough performance.

These results are encouraging for using the proposed indicators in future research on the origins and effects of technological novelty. The question of which (combination) of indicators is the most suitable depends on the specific research question addressed. Novelty in knowledge origins (scientific and/or technological) helps to address research questions pertaining to sources of knowledge used for the invention. Novelty in recombination is more suitable when interested in the elements and principles employed to achieve the functioning of the invention. The choice might be guided by the technological area under consideration. Since technological areas differ in the extent to which they rely on scientific sources and in the number of components embodied in technologies, the indicators might vary in their applicability. For instance, looking for novelty in scientific origins is more likely to be relevant for science based technology areas such as biotechnology. Finally, the choice might be guided by the degree of novelty interested in. When interested in the very skew of the novelty phenomenon, employing a combination of indicators or the indicators showing a higher skew might be most applicable. In general, our results show that the discriminative power of each of the indicator increases with the rarity of occurrence, but imply a trade-off between type I and type II errors.

7.2 Further improvements in indicator construction

Overall, the analysis presented support the validation of our indicators as proxying for technological novelty anteceding potentially radical impact. Nevertheless, more follow-up work needs to be done to further develop the indicators.

The validity of the measures proposed in representing technological novelty relies on two basic assumptions. First, we assume that the combination of patent classifications a patent is assigned to, reliably reflects its recombinant nature. Second, we assume that information with respect to the fields of knowledge an invention draws from is reflected by its backward citations. The analyses provided in this paper are supportive, but do not allow to validate these assumptions directly. A more thorough validation of our indicators would be to engage in in-depth mapping of fields along the dimensions of technological novelty with an extensive argumentation of the validity of the external assessment. Such an exercise falls beyond the scope of this paper, and is left for further research. In addition, a number of candidate-improvements to our indicators suggest themselves. Currently, we do not take into account the 'conceptual distance' between different classes (combining some classes might be more trivial than combining others). Co-citation information of different pairs of classes can for instance be used to weigh measures by 'ease of recombination'. Furthermore, one could introduce a depreciation perspective, by allowing combinations made after a certain time of non-occurrence to be again at risk for novel use. To select the more important novel inventions, one could condition novelty on use of the new combination in the future. A more fine-grained classification of scientific articles would be useful to better track the scientific knowledge origins of technologies.

7.3. Avenues for future analysis

As compared to previous literature on radical innovation which typically looks at only those inventions that have become ex post breakthroughs, looking at ex ante technology novelty as precursor for radical inventions allows to better understand the inventive process that might lead to radical breakthrough performance. A better measurement of ex ante technological novelty might allow us to understand which actors have the capability and economic motivation to pursue novelty with its highly uncertain outcomes. Moreover, it might shed light on factors determining which ones can realize its potential and which ones not. A better measurement of ex ante technology novelty may also allow us to understand better the nature of the impact associated with technological novelty, beyond average impact: the scope for breakthrough impact, the breadth of the impact covering a broader set of applications, the novelty of impact opening up of new fields of applications, the timing of the impact, the disruptive nature of impact displacing traditional sets of applications and the impact novel patents may have indirectly on following patents which can become breakthroughs.

We hope that the conceptualization and the initial validation and analysis of our constructs are already sufficiently convincing at this stage to incite further research in these areas using and further developing our novel technological novelty indicators.

References

Acs, Z.J., Audretsch, D.B., (1990), The Determinants of Small-firm Growth in United-States Manufacturing, *Applied Economics 22*, pp. 143-153.

Ahuja, G., Lampert C.M., (2001), Entrepreneurship in the Large Corporation: A Longitudinal Study of How Established Firms Create Breakthrough Inventions, *Strategic Management Journal 22*, pp. 521-543.

Anderson, P., Tushman, M.L., (1990), Technological Discontinuities and Dominant Designs – A Cyclical Model of Technological Change, *Administrative Science Quarterly 35*, pp. 604-633.

Arthur, W.B., (2007), The Structure of Invention, *Research Policy 36*, pp. 274-287.

Arthur, W. B. (2009). *The Nature of Technology: What It Is and How It Evolves*. Penguin Books.

Arts, S., Appio, F. P., & Van Looy, B. (2013). Inventions shaping technological trajectories: do existing patent indicators provide a comprehensive picture? *Scientometrics 97*(2), pp. 397-419.

Bakker, J., Verhoeven, D., Zhang, L., & Van Looy, B. (2014). Patent citation indicators: one size fits all?. In *Academy of Management Proceedings* (Vol. 2014, No. 1, p. 14018). Academy of Management.

Banerjee, P.M., Cole, B.M., (2011), Globally Radical Technologies and Locally Radical Technologies: The Role of Audiences in the Construction of Innovative Impact in Biotechnology, *IEEE Transactions on Engineering Management 58*, pp. 262-274.

Baumol, W. J. (2004). Entrepreneurial enterprises, large established firms and other components of the free-market growth machine. *Small Business Economics 23*(1), pp. 9-21.

Block, F., & Keller, M. R. (2009). Where do innovations come from? Transformations in the US economy, 1970–2006. *Socio-Economic Review 7*(3), pp. 459-483.

Carlo, J.L., Lyytinen, K., Rose, G.M., (2012), A Knowledge-based Model of Radical Innovation in Small Software Firms, *MIS Quarterly 36*, pp. 865-895.

Carpenter, M. P., Narin, F., & Woolf, P. (1981). Citation rates to technologically important patents. *World Patent Information 3(4)*, 160-163.

Chandy, R.K., Tellis, G.J., (1998), Organizing for Radical Product Innovation: The Overlooked Role of Willingness to Cannibalize, *Journal of Marketing Research 35*, pp. 474-487.

Chandy, R.K., Tellis, G.J., (2000), The Incumbent's Curse? Incumbency, Size and Radical Product Innovation, *Journal of Marketing 64*, pp. 1-17.

Christensen, C. (2013). *The innovator's dilemma: when new technologies cause great firms to fail*. Harvard Business Review Press.

Christensen, C.M., Bower, J.L., (1996), Customer Power, Strategic Investment, and the Failure of Leading Firms, *Strategic Management Journal 17*, pp. 197-218.

Cooper, A.C., Schendel, D., (1976), Strategic Responses to Technological Threats, *Business Horizons 19*, pp. 61-69.

Dahlin, K.B., Behrens, D.M., (2005), When is an Invention Really Radical? Defining and Measuring Technological Radicalness, *Research Policy 34*, pp. 717-737.

Damanpour, F., (1996), Organizing Complexity and Innovation: Developing and Testing Multiple Contingency Models, *Management Science 42*, pp. 693-716.

Day, G.S., Fein, A.J., Ruppersberger, G., (2003), Shakeouts in Digital Markets: Lessons from B2B Exchanges, *California Management Review 45*, pp. 131+.

Dewar, R.D., Dutton, J.E., (1986), The Adoption of Radical and Incremental Innovations – An Empirical Analysis, *Management Science 32*, pp. 1422-1433.

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

Du Plessis, M., Van Looy, B., Song, X., & Magerman, T. (2010). Data production methods for harmonized patent statistics: Patentee sector allocation 2009. *Eurostat Working Paper-Annual Report 2009*.

Ettlie, J.E., Bridges, W.P., Okeefe, R.D., (1984), Organization Strategy and Structural Differences for Radical Versus Incremental Innovtion, *Management Science* 30, pp. 682-695.

Fleming, L., (2001), Recombinant Uncertainty in Technological Search, *Management Science 47*, pp. 117-132.

Fleming, L., (2007), Breakthroughs and the "Long tail" of Innovation, *MIT Sloan Management Review 49*, pp. 69+.

Fontana, R., Nuvolari, A., Shimizu, H., & Vezzulli, A. (2012). Schumpeterian patterns of innovation and the sources of breakthrough inventions: evidence from a data-set of R&D awards. *Journal of Evolutionary Economics 22*(4), 785-810.

Gambardella, A., Harhoff, D., & Verspagen, B. (2008). The value of European patents. *European Management Review 5(2)*, 69-84.

Garcia, R., Calantone, R., (2002), A Critical Look at Technological Innovation Typology and Innovativeness Terminology – A Literature Review, *Journal of Product Innovation Management 19*, pp. 110-132.

Gassmann, O., Von Zedtwitz, M., (2003), Trends and Determinants of Managing Virtual R&D Teams, *R&D Management 33*, pp. 243-262.

Gittelman, M., Kogut. B., (2003), Does Good Science Lead to Valuable Knowledge? Biotechnology Firms and the Evolutionary Logic of Citation Patterns, *Management Science 49*, pp. 366-382.

Griliches, Z. (1990). Patent statistics as economic indicators: a survey. *National Bureau of Economic Research (No. w3301)*.

Gruber, M., Harhoff, D., & Hoisl, K. (2013). Knowledge recombination across technological boundaries: scientists vs. engineers. *Management science 59*(4), pp. 837-851.

Hall, B. H., Jaffe, A.B, & Trajtenberg, M. (2005). Market value and patent citations. *RAND Journal of economics 36*, pp. 16-38.

Hall, B. H., & Lerner, J. (2010). The financing of R&D and innovation. *Handbook of the Economics of Innovation*, *1*, pp. 609-639.

Hargadon, A. B. (2002). Brokering knowledge: Linking learning and innovation. *Research in Organizational behavior 24*, pp. 41-85.

Henderson, R., (1993), Underinvestment and Incompetence as Responses to Radical Innovation – Evidence From the Photolithographic Alignment Equipment Industry, *RAND Journal of Economics 24*, pp. 248-270.

Henderson, R.M., Clark, K.B., (1990), Architectural Innovation – The Reconfiguration of Existing Product Technologies and the Failure of Established Firms, *Administrative Science Quarterly 35*, pp. 9-30.

Hill, C.W.L., Rothaermel, F.T., (2003), The Performance of Incumbent Firms in the Face of Radical Technological Innovation, *Academy of Management Review 28*, pp. 257-274.

Hurmelinna-Laukkanen, P., Sainio, L.M., Jauhiainen, T., (2008), Appropriability Regime for Radical and Incremental Innovations, *R&D Management 38*, pp. 278-289.

Jacobson, S. C., Hergenroder, R., Koutny, L. B., Warmack, R. J., & Ramsey, J. M. (1994). Effects of injection schemes and column geometry on the performance of microchip electrophoresis devices. *Analytical Chemistry*, *66*(7), 1107-1113.

Jaffe, A. B., & Trajtenberg, M. (2002). *Patents, citations, and innovations: A window on the knowledge economy*. MIT press.

Leifer, R., O'Connor, G.C., Rice, M., (2001), Implementing Radical Innovation in Mature Firms: The Role of Hubs, *Academy of Management Executive 15*, pp. 102-113.

Lettl, C., Herstatt, C., Gemuenden, H.G., (2006), Learning From Users for Radical Innovation, *International Journal of Technology Management 33*, pp. 25-45.

Li, G. C., Lai, R., D'Amour, A., Doolin, D. M., Sun, Y., Torvik, V. I., ... & Fleming, L. (2014). Disambiguation and co-authorship networks of the US patent inventor database (1975–2010). R*esearch Policy 43*(6), pp. 941-955.

Martinez, C. (2011). Patent families: when do different definitions really matter?. *Scientometrics 86*(1), pp. 39-63.

Mascitelli, R., (2000), From Experience: Harnessing Tacit Knowledge to Achieve Breakthrough Innovation, *Journal of Product Innovation Management 17*, pp. 179-193.

Nahm, A. Y., Vonderembse, M. A., & Koufteros, X. A. (2003). The impact of organizational structure on time-based manufacturing and plant performance. *Journal of Operations Management 21*(3), pp. 281-306.

Nemet, G.F., (2009), Demand-pull, Technology-push, and Government-led Incentives for Non-incremental Technical Change, *Research Policy 38*, pp. 700-709.

Nerkar, A., (2003), Old is Gold? The Value of Temporal Exploration in the Creation of New Knowledge, *Management Science 49*, pp. 211-229.

Nooteboom, B., (2000), Institutions and Forms of Co-ordination in Innovation Systems, *Organization Studies 21*, pp. 915-939.

Pavitt, K., Robson, M., & Townsend, J. (1987). The size distribution of innovating firms in the UK: 1945-1983. *The Journal of Industrial Economics*, pp. 297-316.

Reinganum, J. F. (1983). Uncertain innovation and the persistence of monopoly. *The American Economic Review 73*, 741-748.

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

Rosenberg, N. (1976). On technological expectations. *The Economic Journal 86*, pp. 523-535.

Rothaermel, F.T., Hill, C.W.L., (2005), Technological Discontinuities and Complementary Assets: A Longitudinal Study of Industry and Firm Performance, *Organization Science 16*, pp. 52-70.

Scherer, F. M. (1989), 'Comments" on Z. Griliches,"Patents: Recent Trends and Puzzles"', *Brookings Papers on Economic Activity 9*, pp. 291-330

Schmoch, (2008), Concept of a Technology Classification for Country Comparisons, *Final Report to the World Intellectual Property Organisation (WIPO)*

Schoenmakers, W., Duysters, G., (2010), The Technological Origins of Radical Inventions, *Research Policy 39*, pp. 1051-1059.

Schumpeter, J.A., (1939), *Business Cycles: A Theoretical, Historical and Statistical Analysis of the Capitalist Process*. McGraw-Hill, New York

Shane, S., (2001), Technological Opportunities and New Firm Creation, *Management Science 47*, pp. 205-220.

Shellabear, M. & Nyrhilä, O. (2004). DMLS- Development history and state of the art. *Laser Assisted Netshape Engineering 4, Proceedings of the 4th LANE,* pp. 21-24.

Story, V., O'Malley, L., Hart, S., (2011), Roles, Role Performance and Radical Innovation Competences, *Industrial Marketing Management 40*, pp. 952-966.

Trajtenberg, M., Henderson, R., & Jaffe, A. (1997). University versus corporate patents: A window on the basicness of invention. *Economics of Innovation and new technology 5(1),* 19-50.

Tushman, M.L., Anderson, P., (1986), Technological Discontinuities and Organizational Environments, *Administrative Science Quarterly 31*, pp. 439-465.

Utterback, J. M. (1996). *Mastering the dynamics of innovation: how companies can seize opportunities in the face of technological change*. Harvard Business Press, Cambridge, MA.

FACULTY OF ECONOMICS AND BUSINESS
DEPARTMENT OF MANAGERIAL ECONOMICS, STRATEGY AND INNOVATION
Naamsestraat 69 bus 3500
3000 LEUVEN, BELGIË
tel. + 32 16 32 67 00
fax + 32 16 32 67 32
info@econ.kuleuven.be
www.econ.kuleuven.be/MSI