Measuring the Relative Importance of Reusing, Recombining, and Creating Technologies in the Process of Invention

Deborah Strumsky
José Lobo
Sander van der Leeuw

SFI WORKING PAPER: 2011-02-003

SANTA FE INSTITUTE

Measuring the Relative Importance of Reusing, Recombining and Creating Technologies in the Process of Invention

Deborah Strumsky*
University of North Carolina at Charlotte
(dstrumsky@uncc.edu)

José Lobo
Arizona State University
(jose.lobo@asu.edu)

Sander van der Leeuw
Arizona State University
Santa Fe Institute
(vanderle@asu.edu)

January 2011

Much work on technological change agrees that the recombination of new and existing technological capabilities is one of the principal sources of technological novelty. But there have been instances in the history of technology of new technologies appearing with few antecedents and which originate technological pathways rather than extend existing ones. The many discussions about the relative roles of recombination and origination as sources of technological novelty have not provided much by way of quantification, which is not surprising given the difficulties in discretizing technologies and classifying technological novelty. In the present discussion we use the patent technology codes employed by the U.S. patent office to classify patents to identify four sources of technological novelty. The resulting taxonomy is based on the newness of the technological capabilities constituting an invention. We then use this taxonomy to quantify the relative importance of recombining or reusing existing technological capabilities, and of originating new capabilities, as a source of technological novelty. Our results clearly show that the process of invention, as recorded by patents granted by the U.S. Patent Office, as been primarily a combinatorial process with a very limited role for the development of original technologies. The importance of reusing existing technological capabilities to generate inventions has been steadily rising and recently overtook recombination as the source of most new patents.

Keywords: patent technology codes, technological novelty, recombination

*: Corresponding author. Department of Geography and Earth Sciences, University of North Carolina at Charlotte, McEniry 429, 9201 University City Boulevard, Charlotte, NC 28223 USA, Phone: 704-687-5934, Fax: 704-687-5966.

The work presented in this paper was largely done during a three week research stay at the Santa Fe Institute during the summer of 2010—we gratefully acknowledge the Institute's hospitality.

1. Introduction

Much work on technological change agrees that the recombination of new and existing technological capabilities is one of the principal sources of technological novelty (see, for example, Allen, 1977; Auerswald et al., 2000; Basalla, 1988; Fleming, 2001; Fleming and Sorenson, 2004; Frenken, 2006; Hargadon and Sutton, 1997; Henderson and Clark, 1990; Kauffman et al., 2000; Kogut and Zander, 1992; Levinthal, 1997; Rivkin, 2001, Rosenberg, 1982; Tushman and Rosenkopf, 1992; Usher, 1954; Von Hippel, 1988; Weitzman, 1998). But there have been instances in the history of technology where new technologies appeared with few antecedents and originated technological pathways rather than extend existing ones (Arthur, 2009; Christensen, 1997; Mokyr, 1992). The many discussions about the relative roles of recombination and origination as sources of technological novelty have not provided much by way of quantification, which is not surprising given the difficulties in discretizing technologies and classifying technological novelty.

Patented inventions are one manifestation of technological novelty, and what is novel about an invention is described by its inventors, in technical and precise detail, in the patent's claims. The United States Patent and Trademark Office (USPTO) categorizes inventions by encoding the information contained in the claims using a system of numerical codes, the patent technology codes. At any given time the existing set of technology codes available to a patent examiner is essentially a description of the current set of technological capabilities. With each new patent application a patent examiner must decide which codes to use to capture the technological components in a patent, and whether or not new technology codes are required to capture the invention's novelty. The introduction of a new technology code sets in motion a retroactive reclassification of all previous patents that may have embodied the newly recognized technological capability. The Patent Office's technology codes thus constitute a set of consistent definitions of technological capabilities spanning over 200 years of inventive activity. We believe that the technology codes are an underutilized data resource for identifying technological capabilities, marking the arrival of technological novelty, and assess the relative importance of recombination in propelling technological change.[1]

In the present discussion we utilize patent technology codes to identify four sources of technological novelty (as instantiated in patented inventions). The resulting taxonomy is based on the newness of the technological capabilities constituting an invention. We then use this taxonomy to quantify the relative importance of recombining or reusing existing technological capabilities, and of originating new capabilities, as a source of technological novelty. The discussion is organized as follows. The next section sets the conceptual framework and poses the research questions. Section three describes the use of patent technology codes to classify the technologies embodied in patented inventions, emphasizing their discriminating nature. Section four presents the empirical results while section five concludes.

[1] A detailed discussion of how technology codes are utilized by the U.S. Patent Office and what information they provide about the technologies embodied in a patent see Strumsky, Lobo and van der Leeuw (2010).

2. Conceptual Framework and Research Questions

Inventions—new artifacts, devices, processes, materials or compounds—embody technological novelty. Some inventions, namely those that are patented, leave behind a documentary trail and as a consequence patents have become a widely used metric in studies of the "knowledge economy" (e.g., Acs and Audretsch. 1989; Griliches, 1990; Jaffe et al., 1993; Jaffe and Trajtenberg, 2002). Although it is the case that most patents have been granted to inventions involving machines or the transformation of one physical substance into another, business methods, computer programs and algorithms can also be patented.[2] We share the view that patents are a good proxy measure for generic inventive activity (Acs, Anselin and Varga, 2002; Bettencourt, Lobo and Strumsky, 2007; Marx, Strumsky and Fleming, 2009; Lerner and Kortum, 2000; Jaffe, Trajtenberg and Hall, 2006).

The research literature on patenting has focused on using the patented inventions as instantiations of technology (e.g., Hall, Jaffe and Trajtenberg 2001; Khan, 2005; Youtie, Iacopetta and Graham, 2008), and on using citations to "prior art" as putative channels for knowledge flows.[3] But inventions can alternatively be seen as distinct technological capabilities combined to bring about a specific outcome (Rosenberg, 1982, 1984). Technological capabilities form part of the stock of knowledge; often these capabilities are already in existence and sometimes patented inventions necessitated the creation of new capabilities. The laser presupposes the ability to construct highly reflective optical cavities, create light intensification mediums of sufficient purity, and supply light of specific wavelengths; the polymerase chain reaction requires the ability to finely control thermal cycling (which involves the use of computers), and isolate short DNA fragments (which in turn applies techniques from chemical engineering); the inkjet printer involves the ability to position extremely small bits of matter with extreme precision (an ability that has been used in a variety of applications other than printing documents). Seeing patented inventions as technological capabilities brought together for achieving specific tasks allows us to conceive of technological novelty as resulting from the development of new technological capabilities or from the combination of existing capabilities in ways that had not been previously witnessed in the patenting record. A patented invention might be judged to be very novel in so far as it has no patent antecedents, but its constituent technologies might already be part of the common stock of knowledge.[4] Furthermore, much profitable technological change consists of refinements of existing technological capabilities

[2] The U.S. Supreme Court recently reaffirmed the patentability of business methods, computer programs and mathematical algorithms (561 U.S. Supreme Court, 2010). Note, however, that patenting has never been restricted to devices. Consider, as an example, the patent granted in 1917 to Claurence Saunders for a "self serving store" which became the model for all subsequent self-service stores and the modern supermarket (patent 1,242,872).

[3] "Prior art" constitutes all information that has been made available to the public in any form before a given date that might be relevant to a patent's claims of originality. It is the applicant's responsibility to list any related previous patents of which he or she is aware of; patent examiners usually add other relevant citations. Patent citations narrow the reach of a new patent by placing cited patents outside the realm of the current patent.

[4] The technological novelty of a patent is therefore distinct from the originality of a patent, a metric which relies on the citations made by an invention to prior patents (Hall, Jaffe and Trajtenberg, 2001; Jaffe, Hall, and Trajtenberg, 2005; Trajtenberg, 1990; Tranjtenberg, Henderson and Jaffe, 1997). Consider the case of one of the most cited inventions, the polymerase chain reaction (patent # 4,683,202) which did not cite any prior inventions (but which has been cited 2,135 times). This patent did not utilize prior inventions but it surely combined existing technological capabilities in a novel way (Pisano 2006).

rather than developing wholly new technologies (Abernathy and Utterback, 1978; Rosenbloom and Christensen, 1994).

The USPTO effectively defines inventions as bundles of technological capabilities, and identifies the technologies constituting inventions through an elaborate system of technology codes which categorize the technological features of an invention. These codes represent a catalogue of the technological capabilities available at any given time: inventions both result from using these capabilities and serve to expand the catalogue. We put forth the following research questions which we believe can be tackled by using patent technology codes:

Q1. Can frequently theorized sources of technology novelty be measured empirically through the use of patents?
Q2. What has been the relative importance of recombination in the development of new inventions?
Q3. Have the rates at which the various sources of technological novelty operate been changing?

3. Methods

3.1 Data: Patent Technology Codes

A patent is an intellectual property right granted by the Government of the United States of America to an inventor; it excludes others from making, using, offering for sale, or selling the invention throughout the United States, or importing the invention into the United States, for a limited time (generally 20 years from the filing date) in exchange for public disclosure of the invention when the patent is granted. Section 101 of U.S. Patent Law specifies four categories of inventions or discoveries that are eligible for the protection of a patent: processes, machines, manufactures and compositions of matter. The United States Patent and Trademark Office (USPTO) grants three types of patents: utility patents—also referred to as "a patents for invention"—are issued for the invention of "new and useful" processes, machines, artifacts, or compositions of matter; design patents, which are granted for the ornamental design of a functional item; and plant patents which are conferred for new varieties of plants or seeds (92% of the patents granted by the USPTO are utility patents). Though it is the case that most patents have been granted to inventions involving machines or the transformation of one physical article into another, the U.S. Supreme Court has reaffirmed the patentability of business methods, computer programs and mathematical algorithms (561 U.S. Supreme Court, 2010).

By definition a patent is granted by the USPTO to an invention that is both novel—i.e., not obvious to others skilled in the same field—and useful. A patent is intended to be limited to only one invention consisting of several closely related and indivisible (i.e. integrated) technologies that, acting together, accomplish a specified task (in patent law this is know as the "unity of invention"). In plain terms what this means is that a jet engine cannot be patented but the numerous components of the jet engine can. In the case of inventions which accomplish multiple and separable tasks the inventors are required to file separate patent applications for each distinct task.

What is new in a patentable invention is specified by the patent's authors in the numbered claims which also serve to define the scope of the legal protection granted by the patent. The claims state, in technical and precise terms, the subject matter of the invention (or discovery), as well as its purpose, principal properties, how it operates and the methods it employs. For patent examiners, the claims are the principal factor in determining whether a proposed invention infringes on an existing patent, and whether it meets the criteria of novelty and non-obviousness. Claims have been a necessary part of U.S. patent applications since the enactment of the Patent Act of 1836. Inventors are obviously motivated to seek the broadest set of monopoly rights, and therefore to be expansive with respect to how many claims they make in their applications.

The USPTO uses a system of numerical technology (or classification) codes to categorize the technologies invoked by approved claims. The codes constitute a classification system that groups patents together according to similarly claimed subject matter (allowing for patents to be searchable), and are used by patent examiners when searching for relevant prior art during a patent application examination process. Multiple dependent claims may fall under a single technology code, though the code will only appear once. The classification system is legally mandated to provide an exhaustive reference to all patentable subject matter; the present USPTO patent classification system is based on the general structure of the first classification schema created in 1900.

A technology (or classification) code consists of two parts: a technology class and a technology subclass. A class is a major category of patentable technology while a subclass is a refined subset within a class. Subclasses have very detailed definitions and some subclasses are nested within hierarchical relationships to other subclasses. There are currently 481 classes and about 150,000 subclasses used by the USPTO; the active technology codes number 214,091. Once codes have been assigned to a patent, one of the classifications must be designated as the original or primary classification. If all the codes occur in the same class, then the most specific or detailed code is chosen. If codes are in different classes, the primary classification is generally determined by the most comprehensive claim. Just as the minimum criterion for a patent is a single claim of novelty, a patent must have at least one code, but there are no limits to how many codes may be assigned to a patent. Other than the code selected as the original classification, the order in which technology codes are listed on a on a patent is arbitrary. The passage from claims to codes is not the result of applying precisely-defined and algorithmic-like procedures, but rather depends on the expertise and experience, that is on the craft, of the patent examiners. We emphasize that the reduction of the number of claims to a more limited number of codes is not informative or revealing about the nature of the technologies constituting an invention (for a detailed discussion of the role of claims in a patent and how technology codes codify claims see Strumsky, Lobo and van der Leeuw (20100).

Once each claim in a patent has been assigned a code, one of the classifications must be designated as the "original" or" primary classification" for the patent. A patent must have at least one code but there are no limits to how many codes may be assigned to a patent, and other than the original classification, the order is arbitrary. Once the classification is completed a patent's codes fully encapsulate those aspects of novelty set forth in the claims; therefore the list of technology codes assigned to a patent specifies the technological capabilities embodied in the

[6] These results are available upon request.

invention. We refer to the set of n technology codes categorizing a patent as an *n-tuple*. (For the results presented here *n-tuples* are treated as combinations and not as permutations.)

"Classification orders," issued monthly by the USPTO, record the changes made to the classification system over the previous month. Classification orders convey the introduction of new codes, changes in the definitions of existing codes, or (quite rarely) the elimination of codes. The classification manual is updated every two months and archived every June and December. The recognition that a new technological capability has come into existence, a recognition formalized by the introduction of a new technology code, usually occurs some time after the new capability has been in play. When the set of technology codes is revised, the USPTO reviews all granted patents and reclassifies those meeting the criteria of the new codes. As an example of this back-dating process, consider the case of nanotechnology. In October 2004 the USPTO announced the creation of a new class for patents, Class 977, for nanotechnology. The USPTO then reviewed all patents issued before the 977 class was created and re-classified those meeting the 977 criteria; as a result, according to the USPTO the first nanotechnology patent was granted in 1986. As a consequence of the reclassification process the technology codes provide a consistent classification scheme making it possible to compare patent technologies across more than 175 years of inventive activity.

We present some basic empirics for technology codes and the technology *n-tuples*. We use data for utility patents and count them based on the year patents were successfully applied for—the application year—so as to count patents as close as possible to the time they were invented. To dampen noise, time-series results are presented for five year windows starting in 1835 and ending in 2009. The year 1835 was chosen as the starting date due to the paucity of patent observations from 1790 to 1834. (Construction of the patent database is discussed in Appendix A.)

Figure 1 shows the mean and median numbers of technology codes assigned to utility patents (calculated within each five-year window). The mean number of codes categorizing patents has increased over the past 175 years, a reflection of the steadily growing technological complexity of inventions. Since 1945 the mean number of codes has hovered around 4 while the median has been around 3 (the 75th percentile has consistently had a value of 5). There has been significant, and increasing, variability as indicated by the difference between mean and median values, and a coefficient of variation of about 0.75 for every time window (much of this variability is due to some inventions having upwards of 20 codes assigned to them).

Figure 2 shows the cumulative distribution of patents by *n-tuples* for all utility patents granted between 1835 and 2009. Sixty-two percent of all patents have been categorized by no more than three technology codes and four codes have sufficed to describe almost 75% of all patents granted over the 175 years covered by the data. Clearly the codes provide the vocabulary for precise and parsimonious descriptions of inventions. Figure 3 depicts the change over time for the proportion of patents categorized by *n-tuples* of different sizes (the percentages are calculated over five-year windows). Whereas one technology code is enough to categorize almost 60% of patents granted between 1835 and 1870, by 2009 only 10% of patents could be described with only one code. The proportion categorized by two technology codes decreased by much less, from 31% to 25%. The proportions accounted by descriptions using 3 and 4 codes

both increased (to 21% and 14%, respectively). Over the past 50 years a third of all inventions have required 5 or more codes indicating that a significant portion of inventive activity consists of combining several distinct technologies. The *n-tuples* are very discriminating descriptions. Figure 4 shows the percentage of differently sized *n-tuples* which are unique, that is, have been used only once to categorize a patent over the period 1835 to 2009. Almost 80% of code combinations consisting of two codes (2-tuples) have only been used once; in the case of 3-tuples the percentage of unique usages is an astounding 92%.

3.2 Sources of Technological Novelty

Since technology codes denote specific technological capabilities, multiple codes are combined by patent examiners in order to generate comprehensive depictions of inventions. The technology codes used to categorize an invention are very precise descriptors and as a consequence most *n-tuples* are unique. Simply noting whether a patent's *n-tuple* has been seen before in the patent record would therefore not be an effective way of identifying technological change. A revealing analysis of technological change using patents needs to take note of the manner in which technological capabilities are brought together to create a new composition, device, or process. We focus on the pairings (binary combinations) of capabilities which can be generated from an *n-tuple*—pairings which are inherent in the combination of technologies which constitute an invention—and use these pairings to assess inventions' technological novelty. We explain the procedure with a stylized example.

Consider a patent categorized by a 3-tuple, {a, b, c}, with a, b and c denoting three technological capabilities. This 3-tuple generates the following pairings of technologies: ab, ac, and bc (recall that the order in which technology codes are arranged in an n-tuple does not matter). If the three codes have been used before to categorize patents, and if the paring ac has not been used in a patent before whereas the pairings ab and ac have, the 3-tuple is classified as a recombination. If the three codes have been used before and if the three pairings have been seen in the categorization of previous patents, the 3-tuple is labeled a refinement. If the three codes used in the 3-tuple are new (that is, have not been used before to categorize any patent), the invention is considered to be original. If a and b are old codes while c is a new code, and if the paring ab has been seen before whereas the pairings ac and bc have not, the invention is labeled a novel combination.

The following definitions specify the four ways in which we use patents' technology codes to identify the sources of technological novelty.

1. *Recombination*: the *n-tuple* consists of new binary combinations of technological capabilities used in previous patents ("new" means not used in a previous period).
2. *Novel Combination*: the *n-tuple* consists of new binary combinations with at least one new code.
3. *Origination*: all the codes used in the *n-tuple* are new.
4. *Reuse* (*refinement*): all of the codes used to classify a patent have been used in previous patent descriptions, and the binary pairings have all been used previously to categorize patents.

Of the 7,706,496 patents included in our database, only 0.65% of them are categorized as *originations*, while *novel combinations* represent 0.87% of all patented inventions. Most patents are either *recombinations* (57% of the total) or *refinements* (41%).

Recombination as defined here is close to the notion of technological recombination often found in the literature (e.g., Fleming, 2001). Reuse involves the grouping of previously existing capabilities or functionalities in order to generate an invention which often amounts to a refinement of existing artifacts, processes or materials. Originations and novel combinations are the two types of inventions which involve not only combining technologies in ways not seen before but also the use technologies not previously in existence—these are the most novel types of invention. Appendix B provides examples of patents which illustrate the four sources of technological novelty (as well as the specificity of technology codes).

4. Results

We now present our main results. Figure 5 shows the time series (expressed in natural logarithms for ease of comparison) for the cumulative number of total patents and cumulative totals for each source of new patents (data show for five-year windows). Several features stand out. The large majority of patented inventions consist of recombinations and refinements, and these two sources of technological novelty have fueled the increase in invention observed over the past three decades. From 1980 to 2009, a period that saw the granting of almost 50% of all patents granted between 1835 and 2009, recombinations and refinements accounted for 44% and 53% of all patents, respectively. Figure 6 zooms in on the period between 1950 and 2009 which makes it easier to note the remarkably steady rate at which new technological capabilities, recorded as originations in our technological novelty taxonomy, developed over nearly sixty years. Not until around the year 2000 did the rate of origination increase noticeably (and concomitantly, so did the rate of novel combinations). It is likely that this burst in origination and novel combination patents was the result of a large backlog of patent applications for business methods, computer programs and mathematical algorithms that were successfully processed and granted after a major U.S. Supreme Court decision in 1996.

Figure 7 plots the percentage of new patents, granted during a given time-window, which represent the recombination and reuse of existing technological capabilities, as well as the use of new technologies (as signified by the use of new codes to categorize patents). Clearly recombination was the primary source of new inventions—representing around 65% of all new inventions during 1835 to 1990. Strikingly, around 1990 the reuse of existing technological capabilities became the predominant source of new patents. This regime shift is not an artifact of aggregating patents across all industries and technologies: the same pattern is observed when patents are grouped into 56 technology areas (such as communications, computer software, drugs, optics, medical devices, etc.).[6] The role of new technological capabilities, either originations or novel combinations, as a source of new inventions was quite minimal, accounting for no more than 2% of new patents over the 1900 to 2005 period. Somewhat surprisingly, the rate at which new technological capabilities were developed remained relatively stable over a period which saw dramatic transformations in the economy and society.

5. Discussion

Our results clearly show that the emphasis the literature has placed on recombination as a source of technological novelty is quite deserved. It would not be an exaggeration to describe invention, as recorded by patents granted by the U.S. Patent Office, as a combinatorial process with a very limited role for the development of original technologies. The process of origination, which under the taxonomy presented marks the arrival of previously unavailable technological capabilities, has played a very small role as a source of patented inventions—its heyday was during the mid-nineteenth century, and for most of the 20th century originations represented no more than 0.5% of all new patents per five-year window. But we must be careful when gauging the importance of origination. As long as there is some origination occurring, providing new capabilities as inputs to the inventive system, refinements and recombination can generate inventions at an increasing rate.

Starting at around 1960 the share of new inventions which were the result of recombination began to decline, and concomitantly, the importance of refinements began to increase. Using the language introduced by March (1991), inventors have been increasingly relying on the exploitation (in contrast to the exploration) of the known technology space as the strategy for finding technological novelty. What was behind the shift in inventive search strategy? The heightened importance of reusing existing technological capabilities as a source of new inventions coincided with the dramatic increase in the rate of patenting: 41% of all patents granted to date by the USPTO have been granted since 1990. This increase in patenting activity did not result exclusively, or even primarily, from an increase in inventive capacity but also reflected the increasing role of patents as a signaling device for attracting interest and investment to firms, and as means for evaluating the performance of scientists and engineers (Gompers and Lerner, 2001; Jaffe and Lerner, 2006). Under pressure to generate patents in copious amounts, inventive agents might have turned to refinements as a surer path to obtaining them. In summary, recombination and recombination are far and away the most important sources of technological novelty.

An obvious question has to do with the relationship between the technological novelty of a patented invention and its usefulness, a question which we aim to devote much attention to in future investigations. The evaluation of the usefulness of a patent is fraught with empirical difficulties, but in the absence of readily-available data on commercialization success or licensing revenues, a common approach is to use citations by other patents—how useful a given invention is to other inventions—as a proxy measure of overall usefulness.[7] Do patent citations differ across the four sources of technological novelty discussed above? Which type of patent is cited most often? Is a high level of technological novelty a predictor of future usefulness? Answering these and related questions constitute an ongoing research effort, but we can report on some preliminary findings.

[7] Another possible, if largely unexplored, indicator of the usefulness of a patented invention is whether its author or owner pays the patent's maintenance fee. All utility patents are subject to maintenance (or renewal fees) which must be paid to maintain the granted patent in force (i.e., maintain ownership). These fees (which are in the order of $12,000) are due 3 1/2, 7 1/2 and 11 1/2 years from the date of the original patent grant. Payment of the maintenance indicates that at least someone thinks that the patent is worth protecting. We intend to examine the relationship between the technological novelty taxonomy presented here and the decision to pay a maintenance fee.

In order to use electronically-available data on patent citations, we restrict our attention to patents successfully applied for between 1975 and 2005; the end year chosen so as to give recently granted patents enough time to build-up citations. Patents were categorized based on the sources of technological novelty, and the number of citations each patent received was counted five and eight years after the date the patent was granted so to control for older patents accumulating more citations simply as a result of longevity. Summary statistics were then calculated for citations per patent for each of the four categories of technological novelty. We find, using eight years as the counting period, that the highest number of citations accrues to patents which are instances of novel combinations (mean = 14.1, median = 8) followed by patents which represent technological origination (mean = 9.8, median = 5). For the two most common sources of technological novelty, recombination (mean = 5.2, median = 3) and refinement (mean = 3.8, median = 0), we find significantly less citations. (The story is unchanged if the average number of citations is normalized by the mean number of citations for each technology class.[8]) Restricting our attention to the past thirty-five years, the most novel patents are cited more often suggesting that true novelty is more useful (if decidedly less common).

There is much else that can be studied using the taxonomy of technological novelty presented here. Does recombination generate "breakthrough" inventions? And what about the institutional profile of the different sources of technological novelty? Are originations primarily associated with small organizations (as popular lore would lead us to expect)? Is there a characteristic type of firm involved in recombination? Are the firms that create the originations the same firms that follow it through and also create the refinements, or do copy cat firms do that? We hope to tackle some of these questions next—and we hope too that other researchers find patent technology codes and our taxonomy of technological novelty useful probes with which to study technological change.

Appendix A: Patent Data

The USPTO makes available data on the technology codes assigned to every patent it has granted in optical disc format (the Cassis Patent Assignments File); the data covers the period from 1790 to the most recent month that data has been archived. The data provided by Cassis file includes the date the patent was granted but not information on the patent filing date. So as to count patents close to the time they were invented we choose to count them in the year they were successfully applied. In order to count patents by application year the patents in the Cassis file have been matched with the patents covered in a database constructed by one of us (Strumsky) using data provided by the United States Patent and Trademark Office (USPTO). This patent database, which covers almost 8 million patent records, includes information on filing data for every utility patent granted by the USPTO since 1963.

For utility patents granted prior to 1963 the application year has been obtained from citation records and direct searches. This still left about 13% of all patents, applied for between 1790 and 1960, without an application year. For these patents the year the patent was granted year was used to replace missing applications years. Prior to the 1980s, the time between the

[8] Different technologies display different propensities for citing prior art: medical devices and combinatorial chemistry, for example, cite prior inventions more often than computer software of communications.

filing of an application and the granting of a patent was significantly less than now (usually taking place within 24 months), and since we use five-year windows to report our results, we don't believe that using granted year for as filing year in those patents introduced any significant bias.

Appendix B: Examples of Technological Novelty in Patents

The patents described next exemplify the sources of technological novelty discussed in section 3.2. Patent number 4,639,925 (filed for on April 24, 1984 and granted on January 27, 1987) is for a semiconductor laser; this patent is classified by two codes: 372/46.013 and 372/704. Code 372/46.013 refers to a coherent light generating system in which part of the control system is oxidized when its valence changes from lower to higher, while code 372/704 refers to a semiconductor base structure provided with an etched channel. At the time this patent was published, neither of the two technology codes had been previously seen in the patent record – therefore patent 4,639,925 is an example of *origination*. Patent 7,670,090 (filed on August 12, 2008 and granted on March 2, 2010) for a "system for transporting an airfoil having a root end, a tip end, and a midsection over a railroad, utilizing a first railcar coupled to a second rail car" is another example of a patent resulting from origination: its three codes (410/44, 410/45, 410/53) had not been previously used to describe an invention.

At the opposite end of the novelty spectrum is patent number 6,286,416 (filed on April 6, 2000 and granted on September 11, 2001) for a Coffee-making apparatus to make coffee and steam for a second time, directly after making a cup of coffee; this patent is categorized with three codes: 99/293; 99/290; 323.1. Code 99/293 refers to an infusion apparatus of the type in which the infusible material is subjected to the action of a vaporized liquid, usually steam; code 99/290 denotes an apparatus for subjecting a material containing an extractable or infusible beverage substance to the action of an extracting fluid which is usually at an elevated temperature; and code 323.1 indicates that the invention involves an electrical system wherein a single electrical source is coupled to a single electrical load circuit and means are provided which control the magnitude or level of the current or voltage. In the case of this coffee-making invention, all three codes used to categorize it had been used in previous patent descriptions, and the binary pairings of the codes have all been used previously to categorize patents. Patent 6,286,416 is therefore an instance of reusing existing technological capabilities. Another example of reusing existing technological capabilities to create an improved invention is provided by patent 2,347,115 (filed in October 18, 1941 and granted in April 18, 1944) for a method to improve the construction of derricks for oil extraction; the patent is described by two codes (52/117 and 52/119). Patent 717,590 (filed for in April 24, 1902 and granted in January 6, 1903) for an improved ship's rudder is another example of an invention which results from using existing technologies to create an improved device. (This patent's description utilizes just two codes: 114/164 and 114/165.)

Patent number 6,408,805 (filed on April 12, 2001) and granted on June 25, 2002 is for a two-stroke cycle engine using a preceding air-layer for scavenging; the patent is categorized by three codes—123/73PP, 123/65P and 123/654—which had been used before in patent classifications, however the three possible pairings of codes have not been seen in a previous patent. This patent is an example of what we term recombination. The patent's three codes refer,

respectively, to a two-cycle recompression internal-combustion engines having a cylinder closed at one end and an air-tight casing inclosing the connecting-rod and crank and forming a closed crank-case in open communication with the other end of the cylinder, an internal-combustion engine having a single single-acting working cylinder and a single working piston reciprocating, and a device having an ignition system with an interrupter located in series with the power supply and starting and stopping current flow through the primary coil. Patent 6,474,599 (applied for on December 11, 2001 and granted a year later) for an aircraft security system comprising a transparent bullet-proof security chamber interposed between the pilot area and the passenger area, is another example of recombining existing technological capabilities (codified by seven technology codes) to inventing a new device.

Next consider patent number 4,190,105 (filed on September 9, 1977 and granted on February 26, 1980) for a heat exchange coil used in engine radiators. This patent is described by three technology codes: 165/179 (an apparatus for transferring heat in which the heat transmitters project from the body wall), 138/38 (using a pipe for changing the direction or rate of flow of a fluid for facilitating the transfer of heat between fluids inside and outside of the pipe), and 165/DIG.536 (an apparatus comprising a non-fluent heat storing mass that in alternation has heat added to it by one fluid stream and removed from it by another). Of these three codes, code 165/DIG.536 had not been used previously to classify a patent. Patent 4,190,105 is therefore an instance of what we term novel combination. Another illustration of a novel recombination is patent 2,155,331 (applied for on December 16, 1938 and granted on April 18, 1939) for a safety apparatus for divers to "ward off creatures of the deep" by means of an electric shock. This invention combines five distinct technological capabilities of which one had not been utilized in prior inventions.

References

Abernathy, W.J., Utterback, J.M. 1978. Patterns of industrial innovation, *Technology Review*, 80, 40-47.

Acs, Z.J., Anselin, L., Varga, A. 2002. Patents and innovation counts as measures of regional production of new knowledge, *Research Policy*, 31, 1069-1085.

Acs, Z., Audretsch, D. 1989. Patents as a measure of innovative activity, *Kylos*, 42, 171-180.

Allen, T. 1977. *Managing the Flow of Technology*. MIT Press, Cambridge, MA.

Arthur, B. 2009. *The Nature of Technology: What It Is and How It Evolves.* Free Press, New York.

Auerswald, P, Kauffman, S., Lobo, J., Shell, K. 2000. The production recipes approach to modeling technological innovation: an application to learning by doing, *Journal of Economic Dynamics and Control*, 24, 389-450.

Basalla, G. 1988. *The Evolution of Technology*. MIT Press, Cambridge, MA.

Bettencourt, L.M.A., Lobo, J. Strumsky, D. 2007. Invention in the city: increasing returns to patenting as a scaling function of metropolitan size, *Research Policy*, 36, 107-120.

Christensen, C. 1997. *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail.* Harvard Business Press, Cambridge, MA.

Fleming, L. 2001. Recombinant uncertainty in technological search, *Management Science,* 47, 1019-1039.

Fleming, L., Sorenson, O. 2001. Technology as a complex adaptive system: evidence from patent data, *Research Policy*, 30, 117-132.

Fleming, L., Sorenson, O. 2004. Science as a map in technological search, *Strategic Management Journal*, 25, 909-928.

Frenken, K. 2006. *Innovation, evolution and complexity theory.* Edward Elgar Publishing, Inc., Northampton, MA.

Gompers, P.A., Lerner, J. 2001. *The Money of Invention: How Venture Capital Creates New Wealth.* Harvard Business press, Cambridge, MA.

Griliches, Z. 1990. Patent Statistics as Economic Indicators: A Survey, *Journal of Economic Literature,* 28, 1661-1707.

Hall, B.H., Jaffe, A., Trajtenberg, M. 2001. The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools. NBER Working Paper 8498. National Bureau of Economic Research, Cambridge, MA.

Hargadon, A, Sutton, R. 1997. Technology brokering and innovation in a product development firm, *Administrative Science Quarterly,* 42, 716-749.

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

Jaffe, A., Hall, B.H., Trajtenberg, M. 2005. Market value and patent citations, *Rand Journal of Economics*, 36, 16-38.

Jaffe, A., Lerner, J. 2006. *Innovation and its Discontents: How Our Patent System is Endangering innovation and Progress and What to do About It.* Princeton University Press, Princeton.

Jaffe, A.B., Trajtenberg, M. 2002. *Patents, Citations, and Innovations: A Window on the Knowledge Economy*. MIT Press, Cambridge, MA.

Jaffe, A.B., Trajtenberg, M., Hall, B. 2006. Market value and patent citations: a first look, in Cantwell, E. editor, *The Economics of Patents*. Edward Elgar Publishers, Northampton,

MA.

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

Kauffman, S., Lobo, J., Macready, W.G. 2000. Optimal search on a technology landscape, *Journal of Economic Behavior and Organization,* 43, 141-166.

Khan, Z. 2005. *The Democratization of Invention: Patents and Copyrights in American Economic Development, 1790-1920.* Cambridge university Press, Cambridge.

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

Lerner, J., Kortum, S. 2000. Assessing the contribution of venture capital to innovation, *Rand Journal of Economics*, 31, 674-692.

Levinthal, D. 1997. Adaptation on rugged landscapes, *Management Science*, 43, 934-950.

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

Marx, M., Strumsky, D., Fleming, L. 2009. Mobility, skills, and the Michigan non-compete Experiment, *Management Science*, 55, 875-889.

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

Pisano, G.P. 2006. *Science Business; the promise, the Reality and the Future of Biotech.* Harvard Business School Press, Boston.

Rivkin, J.W. 2001. Reproducing knowledge: replication without imitation at moderate complexity, *Organization Science*, 12, 274-293.

Rosenbloom, R. S., Christensen, C.M. 1994. Technological discontinuities, organizational capabilities, and strategic commitments, *Industrial and Corporate Change*, 3, 655-685.

Rosenberg, N. 1982. *Inside the Black Box: Technology and Economics*. Cambridge University Press, Cambridge.

Rosenberg, N. 1994. *Exploring the Black Box: Technology, economics, and History.* Cambridge University Press, Cambridge.

Rosenthal. S.S., Strange, W.C. 2004. Evidence on the nature and sources of agglomeration economies, *Handbook of Regional and urban Economics*, Vol.4 Henderson, J.V., Thisse,

J.F., eds. Elsevier press, Amsterdam.

Strumsky, D., Lobo, J., van der Leeuw, S. 2010. Using patent technology codes to study technological change. Santa Fe Institute Working Paper 10-11-023.

Trajtenberg, M. 1990. A penny for your quotes: patent citations and the value of innovations, *RAND Journal of Economics*, 21, 172-187.

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, 19-50.

Tushman, M., Rosenkopf, L. 1992. Organizational determinants of technological change: toward a sociology of technological evolution, *Research in Organizational Behavior,* 14, 311-347.

Usher, A.P. 1954. *A History of Mechanical Inventions.* Dover Publication, New York.

Von Hippel, E. 1988. *The Sources of Innovation.* Oxford University Press, New York.

Weitzman, M.L. 1998. Recombinant growth, *The Quarterly Journal of Economics*, 113, 331-360.

Youtie, J., Iacopetta, M., Graham, S. 2008. Assessing the nature of nanotechnology: can we uncover an emerging general purpose technology?, *The Journal of Technology Transfer*, 33, 315-329.

Figure 1. Mean and median number of technology codes assigned to patents, 1835 – 2009 (calculated for five-year windows)

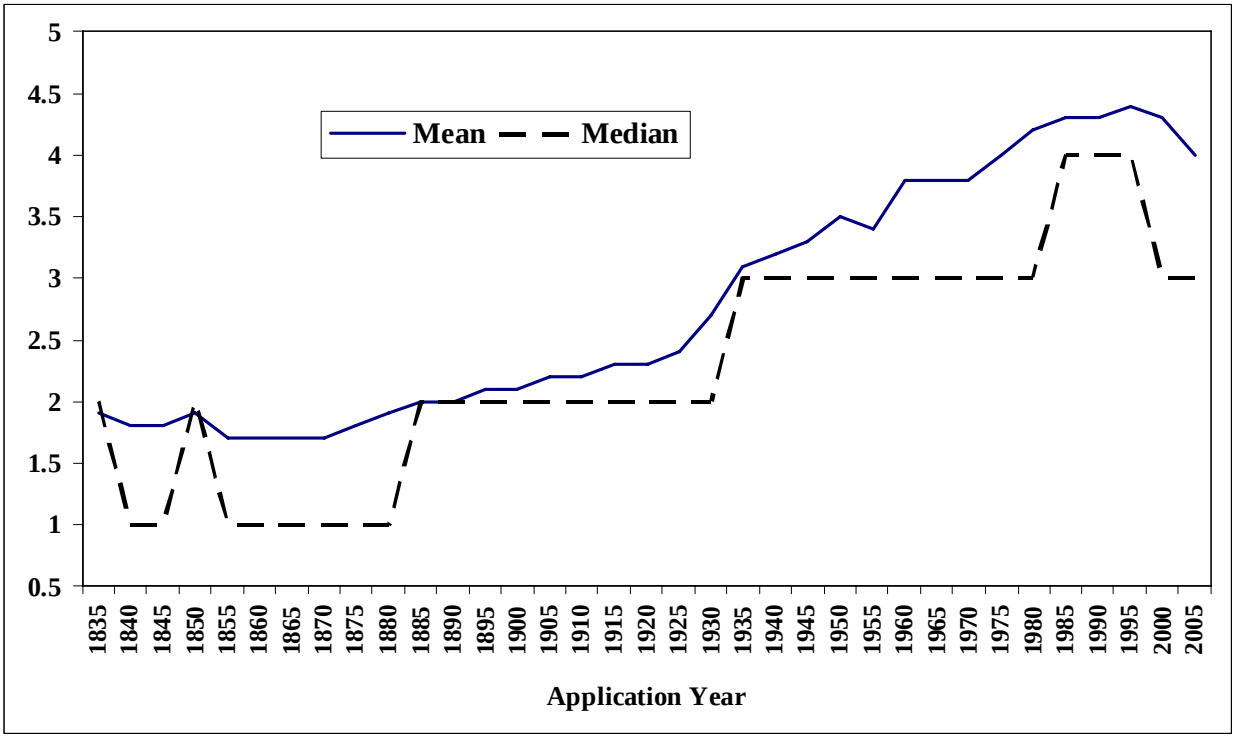

Figure 2. Cumulative distribution by *n-tuples* (the number of technology codes used to describe a patent) of all patents granted between 1835 and 2009.

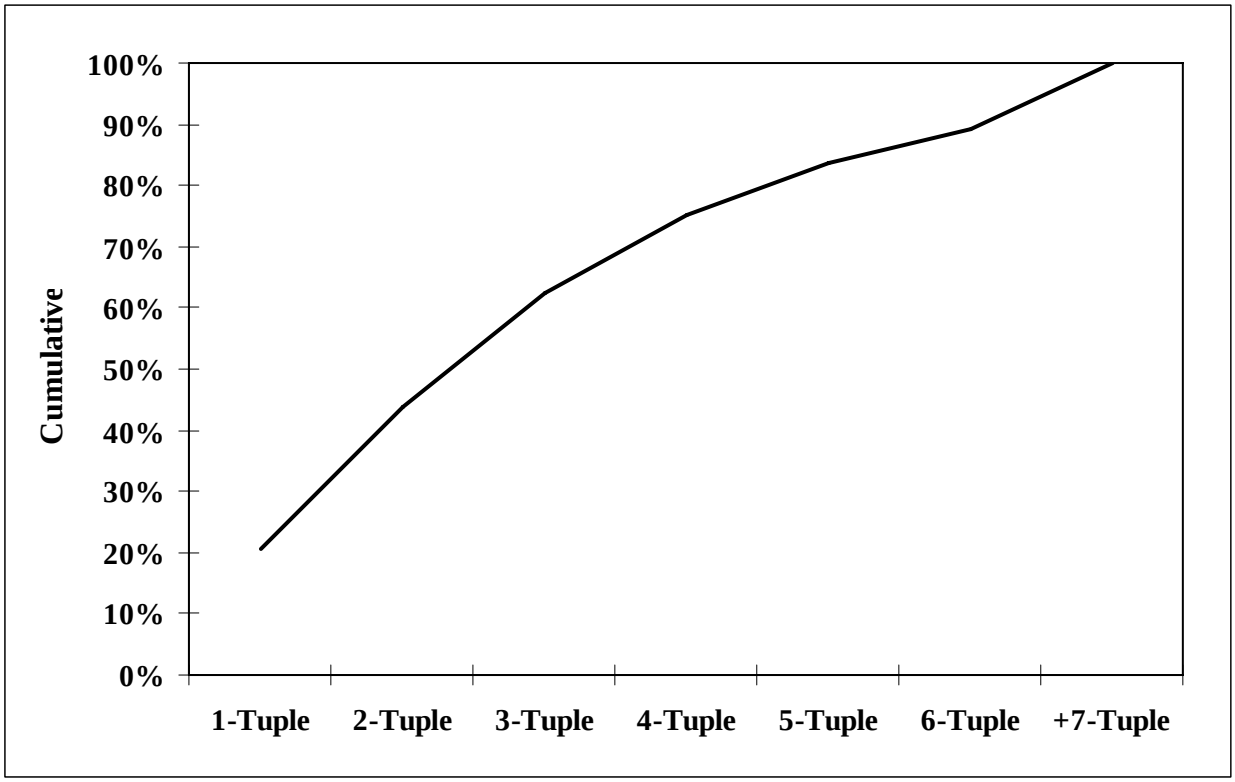

Figure 3. Proportion of patents categorized using 1, 2, 3, 4, 5, and mor etha n five codes (by five year-windows).

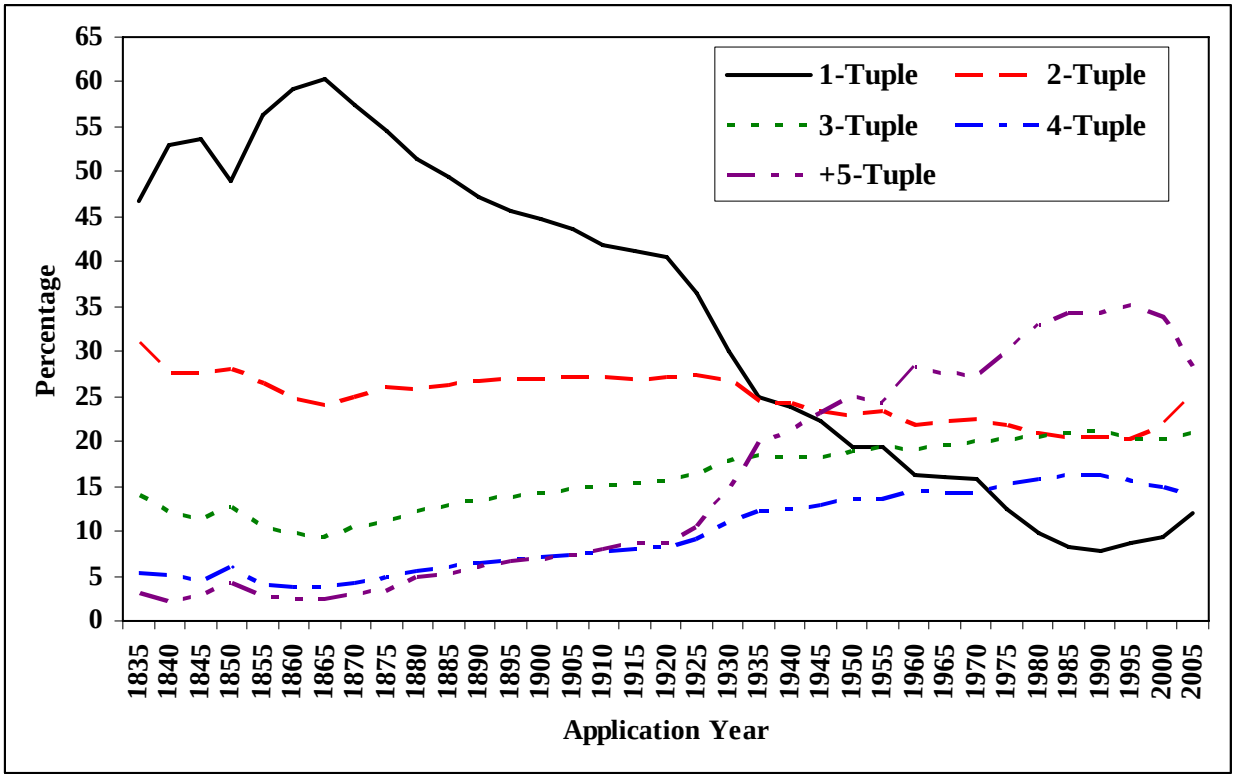

Figure 4. Percentage of differently-sized *n-tuples* which have been used only once to describe a patent's technology.

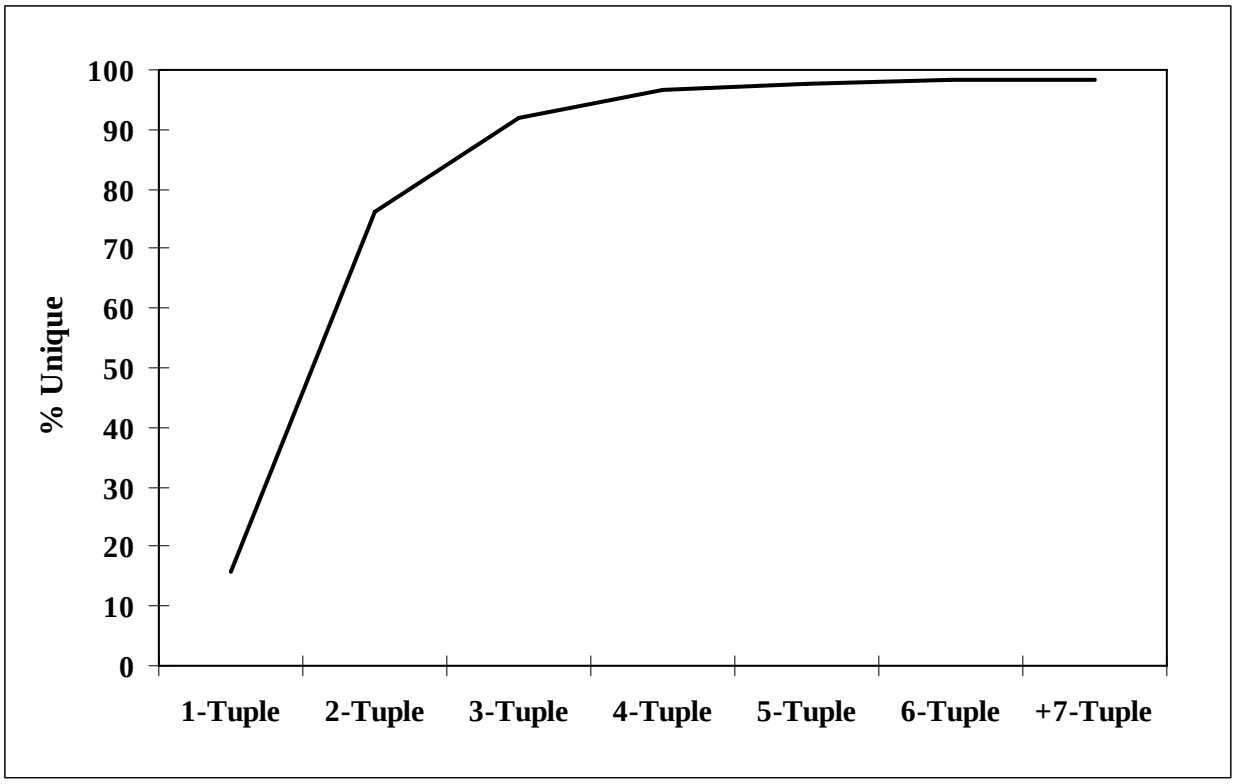

Figure 5. Cumulative number of patents by source of technological novelty, 1835 – 2009.

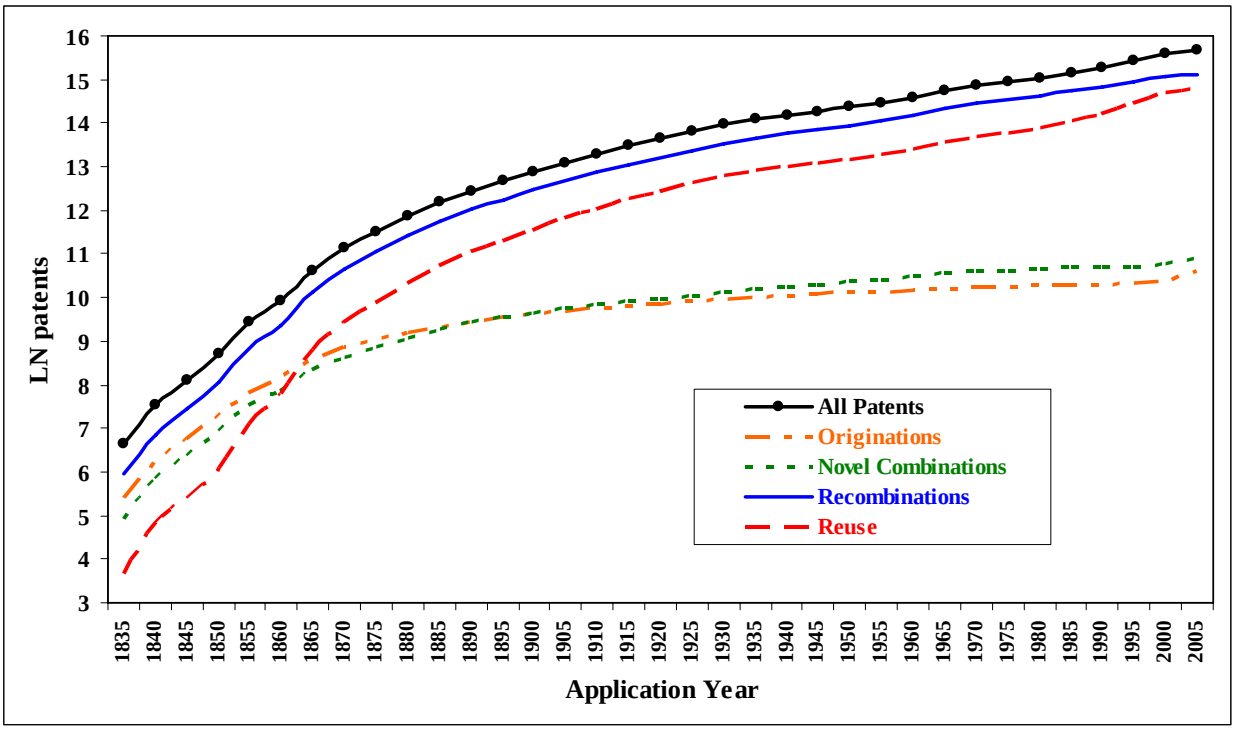

Figure 6. Cumulative number of patents by source of technological novelty, 1950 – 2009.

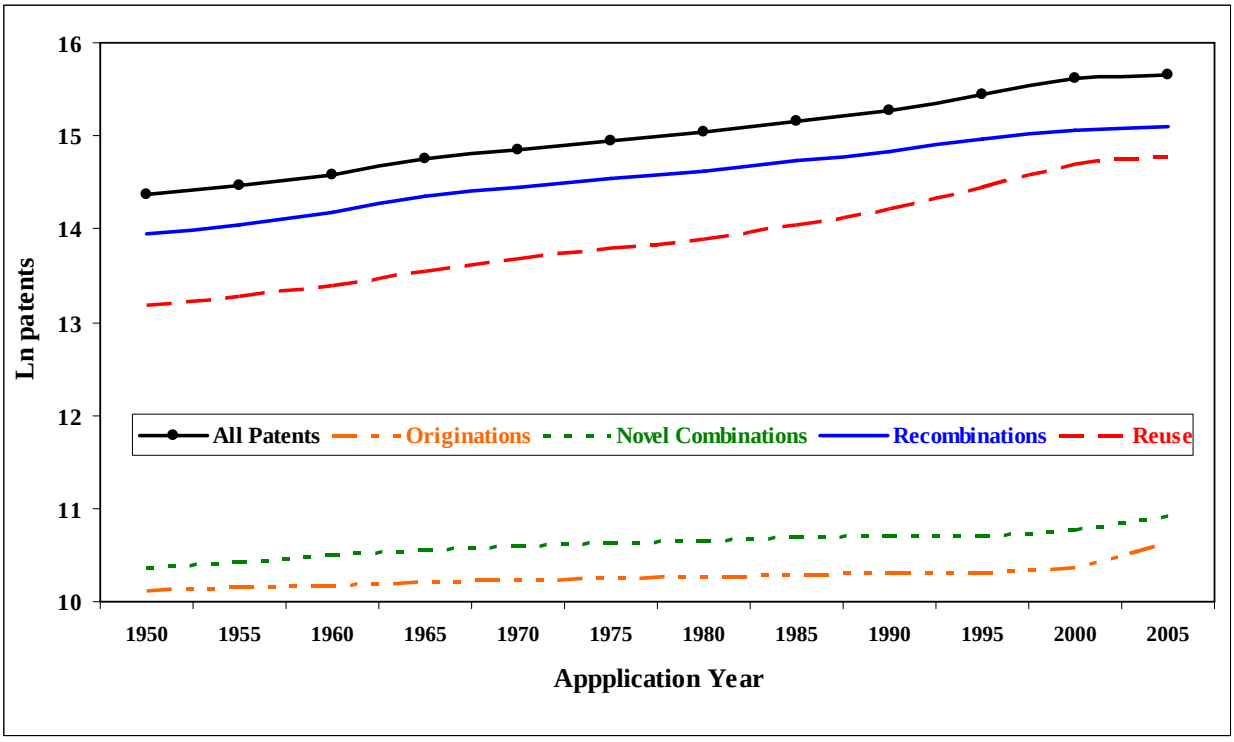

Figure 7. Percentage of new patents, ranted during a given 5-year window, accounted for by the four sources of technological novelty.