Cross-Pollination in Science and Technology: Concept Mobility in the Nanobiotechnology Field

STINE GRODAL
Boston University
School of Management
Boston, MA, 02215, USA
grodal@bu.edu

GRID THOMA
University of Camerino and
CESPRI, Bocconi University
Milan, Italy
grid.thoma@unibocconi.it

April, 2010

This article is forthcoming in Annales d'Economie et Statistique

Acknowledgement: We would like to acknowledge the comments and feedback from participants at DRUID, CINet and the NBER Conference on Emerging Industries - Nanotechnology and NanoIndicators.

1 The authors are listed alphabetically; both authors contributed equally to this work.
Recombination lies at the heart of many innovative processes. It is thus no surprise that a plethora of studies have investigated the impact of cross-pollination on innovation. Yet, these studies have only investigated how cross-pollination affects the creation of innovations, while overlooking how cross-pollination might influence their diffusion. Furthermore, these studies have investigated cross-pollination at the level of the individual, team or through case-studies of individual technologies while assuming that cross-pollination occurred between innovative ideas that these individuals possess. In order to address these gaps in the literature in this paper we move the unit of analysis to the level of the individual concept, and investigate how cross-pollination influences concept mobility. Our setting is the cross-pollination of concepts between nanotechnology and biotechnology, which yielded the new subfield nanobiotechnology. Drawing on a large dataset of publications, patents and press-releases between 1991 and 2005 we track how 133,128 concepts move from science to technology and commercialization. We find strong support for the hypothesis that cross-pollination facilitates concept mobility. Scientists who reside in commercial firms generally assist the mobility of concepts, but hinder the mobility of cross-pollinated concepts. Furthermore, if a patent contains cross-pollinated concepts it is more valuable. This paper contributes to our understanding of how cross-pollination influences the mobility of concepts between institutional contexts, and thus augments our understanding of the commercialization process. We also detail the growth patterns of the emerging nanobiotechnology field.

Keywords: technology; science; commercialization; cross-pollination
INTRODUCTION

The intersection of scientific fields yields a fertile breeding ground for new ideas (Fleming 2001; Fleming et al. 2007; von Hippel 1988). The recombination of previously separate concepts lies at the heart of creativity and novelty (Hargadon and Sutton 1997; Hargadon 2003; Schumpeter 1934). An increasing amount of these scientific discoveries generate innovations with commercial potential (Gambardella 1995; Klepper 2001; Murray 2002; Shane 2001; Shane 2002). The key factor that science contributes to commercial production is knowledge. In the post-industrial society the principal asset in productions has become knowledge and the ability to generate and integrate different knowledge sources within the organization (Machlup 1962; Powell and Snellman 2004; Rosenberg and Steinmueller 1988).

Concepts involved in the commercialization of knowledge exist in three integrated but separate institutional environments: science, technology and commerce (Dasgupta and David, 1994; Rosenberg, 1990). In science researchers generate knowledge, which they disseminate through scientific articles, presentations at research conferences and their informal network of friends and colleagues. Some scientific concepts are translated into technological concepts in the form of technical drawings, documentation and patents. A fraction of these technological concepts are subsequently integrated into actual products; they are commercialized (Agrawal 2006). We define science based on its method of dissemination, i.e. science is the knowledge that is published in scientific articles, where it is made publicly available for others to read and use. Technology is technical knowledge disclosed by inventors in patent documents. An invention needs to be novel and useful in order to be patented, and the public description of the invention enables lawful enforcement of its claims.

Over the last half decade increased efforts have been made to track the flow of knowledge between organizations and across institutional contexts (Powell et al. 1996; Powell

---

2 Research has also highlighted other mechanism for protecting knowledge assets by firms (Levin et al, 1987).
and Snellman 2004; Sorenson et al. 2006). While studies and surveys have suggested that most inventions are patented, many scientific ideas still are not integrated into technologies (OECD 2007). Further, the literature often assumes that technological concepts and commercialization are identical phenomena, measured by whether a scientific concept is paralleled with a patent (Murray 2002). There is, however, evidence that even though a scientific concept is translated into a technological possibility in the form of a patent it is not always commercialized (Mirowski and Sent 2002). We, thus, still lack an understanding of how and why some concepts move between knowledge spaces, but others fail to proliferate (Aldrich 1999).

There is disagreement within the literature about how to best conceptualize and measure knowledge mobility. At the interorganizational level much research has emphasized the role of strategic alliances in facilitating knowledge mobility (Mowery et al. 1996; Powell et al. 1996). Other research has shown that knowledge complexity and the social distance between organizations interacts in predicting the flow of knowledge between organizations (Sorenson et al. 2006).

It has proven more difficult to study knowledge mobility between different institutional environments. Recently a stream of research has captured the mobility of knowledge by measuring citation patterns among both scientific articles and patents (Gittelman and Kogut 2001; Katila and Ahuja 2002). While moving closer to a quantitative measure of knowledge citations fall short of measuring knowledge directly, because a citation can have multiple meanings and refer to a wide spectrum of concepts and ideas within the article. To develop a more specific notion of knowledge dynamics we turn our focus to the mobility of individual concepts.

Thus, the research question that we address in this paper is: What are the drivers of concepts mobility between science, technology and commercialization?
THEORY DEVELOPMENT

Cross-pollination

Studies of science and technology show that radical innovations spur the emergence of new fields (Basalla 1988). The development of genetic engineering, for example, enabled the emergence of the biotechnology industry (Powell et al. 1996; Zucker et al. 1998). Characteristic of most novel technological fields is that they proliferate at the periphery of existing technological fields cross-pollinating knowledge from separate technological areas (Hargadon and Sutton 1997; Hargadon 2003; von Hippel 1988). Cross-pollination is defined as the recombination of previously separate concepts. Biotechnology emerged at the intersection of biology and organic chemistry. The collaboration of people in multiple disciplines enabled the discovery of DNA as a double helix. Rosalind Franklin and her collaborator Maurice Wilkins were educated as chemists, James Watson had a degree in zoology, and Francis Crick had a degree in physics (Stokes 1982). The later development of molecular biology and its commercialization in the form of biotechnology drew attendance from people in chemistry, biology and physics. Digital sound, which emerged at the intersection between computer science and music, is another example of a new interdisciplinary field. It was only through the availability of persons with connections both within the computer science and the music department that the first digital synthesizer was developed. Digital sound is now a key element in the multibillion dollar music industry (Nelson 2005).

The existing research on cross-pollination has focused either on cross-pollination at the level of the individual, the team, or case studies of individual technologies (Fleming and Sorenson 2001; Hargadon and Sutton 1997; Nelson and Winter 1982; Schumpeter 1934). Schumpeter (1934) described the hallmark of entrepreneurship as recombining resources in novel ways. In Schumpeter’s account the locus of cross-pollination is with the individual entrepreneur. Fleming (2001) shows how entrepreneurs, who recombine new elements, in
general are less innovative, but the variance of cross-pollinated knowledge is much higher, which implies that individuals, who recombine concepts, create both influential breakthroughs and trivial inventions. Padgett (2001) also embodies this perspective as he views entrepreneurs as recombining logics in the environment in order to create new organizational forms. Another line of research has emphasized the group as the unit of analysis. Hargadon and Sutton demonstrate the condition under which teams are more likely to generate cross-pollinated results. Fleming (2004) shows that interdisciplinary teams produce more radical innovations than disciplinary based teams. Further, in the video-game industry Tschang (2007) finds that creativity in game design studios occurs through recombination of elements from prior games. Yet another line of research focuses on cross-pollination within individual technologies by examining the knowledge flows that facilitated their creation (Brusoni et al. 2001; Stankiewicz 2000). Common across these research streams is an implicit assumption that novel concepts are cross-pollinated. But the studies fail to examine cross-pollination at the level of the concept; instead they measure other proxies to assess whether cross-pollination has occurred.

Furthermore, the literature on cross-pollination has primarily been concerned with whether cross-pollination yields innovative outcomes (Fleming 2001; Fleming et al. 2007; Hargadon and Sutton 1997; Hargadon 2003; Nelson and Winter 1982). But for cross-pollinated ideas to impact technology and economic growth they need to move from their locus of first use to other institutional arenas. Otherwise the cross-pollinated concepts might be innovative, but they will never gain widespread acceptance. Schumpeter’s theory of entrepreneurship is, for example, based not only on the assumption that the entrepreneur recombines existing knowledge in the creation of novel concepts, but also that the novel concepts proliferate after cross-pollination has occurred (Schumpeter 1934). We thus hypothesize:
Hypothesis 1a(1b): Cross-pollination facilitates concept mobility from science (technology) to technology (commercialization)

As noted above research has shown that cross-pollinated concepts are often more innovative. Lacking from this research is an understanding of how concepts that are innovative behave after they move into a new institutional context (Evans 2004). There is, however, reason to believe that concepts that were cross-pollinated in science will provide a higher economic value if they become integrated into a technology than non cross-pollinated concepts (Fleming et al. 2007). We thus hypothesize:

Hypothesis 2: Patents containing cross-pollinated concepts have a higher quality than non cross-pollinated patents

Proximity

For concepts to move from one sphere to the other they need to be translated and integrated to fit the social structures and prescriptions characteristic of the receiving sphere (Bechky 2003). The process of translating concepts between spheres is made easier if individuals involved in the translation process possess knowledge from both spheres (Bonaccorsi and Thoma 2007).

It has become increasing common for scientists and industrial researchers alike to participate in both research and commercialization efforts. University scientists not only publish their work, but also write patents to claim the commercialization rights of their discovery. In most technical fields the origin of entrepreneurship can be traced to academic science (Klepper 2001; McKelvey 1996).

Furthermore, industrial scientists are no longer satisfied just commercializing their invention, but wish to publish their findings in academic journals (Colyvas and Powell 2006; Owen-Smith and Powell 2001). The multivocality of entrepreneurs facilitates the mobility of concepts between science and technology. In the process of science commercialization the
availability of individuals who are familiar with both science and commerce facilitates the mobility of concepts from science to technology. We thus hypothesize:

\[ H3a(3b): \text{Proximity to commerce facilitates concept mobility between science and technology (commercialization)} \]

The impact of commerce on science has been extensively debated in the literature (Dasgupta and David 1994). Some studies suggest that communication between industrial and academic scientists stimulates both scientific inventions and technological innovation. The argument draws on the observation that the nature of discovery is unpredictable and chaotic, and that interaction between institutions with different beliefs, goals and values can yield unexpected discoveries (Van de Ven et al. 1999). Other studies have questioned the benefits of collaborations between academia and industry. The first critique emphasizes that academic scientists will begin to substitute time and effort used for basic research with more applied activities (Azolulay et al. 2006; Breschi et al. 2007). \(^3\) The second argument highlights the long term risk that the norms of secretiveness and proprietary views of knowledge prevalent among industry scientists will take hold in academia. Thus, close relationships between academia and industry might lead to a diffusion of industry practices to scientific institutions (Dasgupta and David 1994; Etzkowitz 1998). Furthermore, gifts provided to academic institutions from industry might come with non-disclosure agreements, demands to provide knowledge that is relevant for industry, and pressure to not publish unflattering research results (Etzkowitz and Leydesdorff 1995).

A central question in the debate has been the extent to which industrial scientists produce knowledge that is radically novel. Science and commerce differ to the extent that they value the exploration versus the exploitation of knowledge (Nowotny et al. 2001). The culture of academia promotes and values the exploration of radically novel ideas and scientists are rewarded for perseverance in generating cross-pollinated ideas that depart from established

\(^3\) It is worth noticing that these studies did find strong evidence for their claim.
thought (Kuhn 1993 [1962]). In contrast industry scientists are employed to engage in work that will increase the company’s profitability in the near term. The exploitation of existing knowledge yield more sure bets and less risk for the firm than novel discoveries (Nowotny et al. 2003). Furthermore, it has been shown that access to industrial partners and industrial funding decreases the innovativeness of scientific research (Evans 2004). We thus hypothesize that the benefits of proximity will be outweighed by the difficulties of translating highly novel content between science and industry. There will thus be an interaction effect between proximity and cross-pollination, where industrial affiliation hinders the mobility of cross-pollinated concepts:

\[ Hp4a(4b): \text{Proximity to commerce hinders the mobility of cross-pollinated concepts between science and technology (commercialization)} \]

Impact of Technology Translation

Concept mobility from science to commercialization is often intersected by a presence in the technology space (David and Foray 1995). Many scientific concepts appear in patents before they are integrated into a product. Patents offer legal protection for the investment a firm makes in knowledge creation to prohibit that products can be reverse engineered and cheaply copied by a competitor (Scotchmer 2005). Moreover, they facilitate and incentivize technology transfer (Gans, Hsu, and Stern, 2005). In the last decades structured markets for technologies have emerged in important science based industries such as chemicals and pharma-biotechnology, computers and semiconductors, software, and other IT related services. The pace of technologies transferred is growing rapidly and incumbent firms, start-ups, public research organizations, and universities have deliberately elaborated patent based strategies and business models for the commercialization of their knowledge assets (Arora, Fosfuri, and Gambardella, 2001).
The first step in the commercialization process is often to make claims to intellectual property in a patent. Once detailed descriptions of how a scientific concept might be commercialized are outlined in a patent it is easier for the concept to subsequently be integrated into the commercial sphere. We thus hypothesize:

\[ H5: \text{The appearance of a concept in technology stimulates the concept's mobility into commercialization} \]

**DATA AND METHODS**

**Setting: Nanobiotechnology**

We test our hypotheses within the field of nanobiotechnology (nanobio). To choose our area of study we first conducted 11 interviews with nanoscientists at leading U.S. research institutions, to identify the area of nanotechnology that scientists thought were most revolutionary. The interviews identified nanobio as a field of increasing importance. The nanobio field is located at the intersection of two technological areas; biotechnology and nanotechnology, which provides a rich setting for studying the effects of cross-pollination. Nanotechnology emerged out of the intersection between material science, electrical engineering and physics in the beginning of the 1980s. The invention of new methods of inventing like the atomic force microscope (Darby and Zucker 2003) enabled novel research at the nano-scale. Biotechnology is a more established discipline that emerged at the intersection between biology and organic chemistry in the middle of the 1970s (Markel and Robin 1985).

In the early days of nanotechnology, biotechnology was a marginal application area. During the 1990s the synergies between the biological sciences and nano-sciences emerged and the nanobio field has experienced accelerated growth ever since. A commercial nanobio field is in the making. Extraordinary scientific achievements have been accomplished and entrepreneurial firms are rapidly attempting to commercialize nanobio science (Darby and Zucker 2003). The core element that delineates the nanobio field from nanotechnology and
biotechnology is that it combines biological structures with inorganic molecules. Discoveries within nanobio address diagnostics, drug development and drug delivery. Many scientists and companies are working to create “lab-on-a-chip” to aid both drug discovery and delivery. In one of our interviews a material scientist for example explained that his research career was built on developing nanoparticles for use in disk-drives. Over the last couple of years he began working with molecular biologists to develop better sensors and diagnostic tools. In the collaboration they combined nanoparticles, normally used in disk-drives, with genes, proteins, and enzymes to develop new cancer diagnostics. This cross-pollination of knowledge led to high rates of improvement over the existing technologies. Other scientists are taking advantages of the novel properties of nanoparticles to develop methods for drug delivery, like encapsulating a drug within a nano molecule.

We chose to study the emergence of the nanobio field exclusively with the inventions filed in the US patent office because the United States dominates research in material and biological research. Moreover many important non-American inventions tend to be published and patented in the United States due to the importance of the American commercial and knowledge market (Hall and Trajtenberg 2004). This is particularly true for the nanotechnology field (Bonaccorsi and Thoma, 2007; OECD, 2007).

**Methodological Motivation**

Patents may be based not only on the prior art documented in other patents, but in part or fully on new scientific knowledge. Since published scientific research results can be used to illustrate the state of the art against which the application has to be evaluated, patent examiners will search for relevant references in the scientific literature. The logic of these references is to support the claims that are made in the application. Researchers have used patent citations to develop a taxonomy of industries (Grupp 1992; Heinze and Schmoch 2004; Tijssen 2004) and to document the networks of patents (Popp 2005; Verspagen 2005).
theoretical motivation for developing temporal patent networks is to grasp how knowledge
develops and evolves over time.

On the methodological side, there are several shortcomings of the existing measures of
non-patent literature. First, it is not clear to what extent non-patent literature citations are
assigned by inventors or by examiners. It is well known that inventors primarily introduce
references in the USPTO, while in the European system they are introduced by the examiners.
Breschi and Lissoni (2004) claimed that, at least in the US patent system, the variation in who
assigns non-patent literature citations creates severe distortions in the data. The full validity of
citation patterns has to be established, given that the motivations for a patent to cite other
patents are rather intricate and call upon legal and strategic considerations.

Second, non-patent citations do not convey any direct information on the degree to
which the scientific content was able to generate valuable innovation in future development of
the technology. Since we know that the distribution of patents by degree of usefulness is
extremely skewed, it is possible that patents with a high number of non-patent references are
among those that are never used, and so have limited economic value. One approach to
mitigate this limitation is given by a careful analysis of patent quality, using the indicators first
proposed by Trajtenberg (1990) and later developed by Jaffe et al. (1993). There is sufficient
evidence that the economic value of patents is associated with the number and quality of
citations received in other patents (Hall et al. 2005; Harhoff et al. 1999; Jaffe and Trajtenberg
2002). In addition several authors have suggested a complementary metrics, i.e. the existence
of patent litigation as a measure of value, because patents that assignees are willing to pay to
defend have a larger economic value (Agarwal and Bayus 2002; Harhoff et al. 2003; Lanjouw
and Schankerman 2001).

In this study we address the science and technology interaction using a novel approach.
We measure how scientific concepts move between three spheres: Science, technology and
commercialization. The proxy we use to measure concept mobility is the presence of
keywords in three document types: Scientific articles, patents, and press releases. We chose to focus on keywords, because a word is the most basic element of knowledge (Pierce 1931). The set of keywords within a given scientific field thus creates a representation of the knowledge within the field (Eco 1976). We chose to specifically focus on authors’ keywords, because these keywords are self revealed by the author (instead of being computer generated), that is these keywords were the ones that the author found to represent the most important and novel elements of the paper. Author generated keywords thus represent the core elements of the knowledge represented in the paper.

Science

We used the ISI database to locate nanotechnology and biotechnology keywords during the 14-year period between 1991 and 2005. Due to the difference in age between the biotechnology and the nanotechnology fields we used two different methods to isolate biotechnology and nanotechnology keywords.

Biotechnology keywords. To single out the nanobio science field we identified scientific publications that contained both biotechnological and nanotechnological search words. We selected author specified keywords from two specialized journals in the field of Biotechnology and Applied Microbiology and Cell Biology, Biotechnology and Bioengineering (BB) and Embo Journal (EMBO) respectively. Our criteria for selecting these journals were the following: First, we looked for journals that were widely read in the field: both BB and EMBO have been in the top quartile of the impact factor index distribution of their field since at least 1999 (ISI JCR, 2005). Secondly, we looked for journals founded before 1991 and consistently containing authors’ keywords: ISI began collecting authors’ keywords regularly after January 1991. Finally, we looked for journals that targeted broad

---

4 Even though author revealed keywords represent the most robust representation of the ideas present in an article then author revealed keywords are a possible source of strategic behaviour. For example we do not know whether authors’ participation in a social group, like an invisible college or academic department, might create an unobserved bias in their choice of keywords.
topics within the field and that published many articles in absolute terms. We isolated all keywords used in BB and EMBO in the period 1991-2005 obtaining a combined list of 28,194 biotechnology keywords.

*Nanotechnology keywords.* Because there are no established nanotechnology journals that have been around for a long time we had to use a different search strategy to isolate nanotechnology articles. To identify nanotechnology publications we used the ISI Fraunhofer Institute word list to search titles, keywords and abstracts (Fraunhofer-ISI 2002). This search strategy retrieved more than 240,000 publications from ISI during the period 1991-2004. We retrieved all the keywords from this set of articles, which generated a basic pool of 146,484 nanotechnology keywords.

*Nanobio publications.* To isolate nanobio keywords we looked at the overlap between the biotechnology and the nanotechnology keywords, which generated a list of 7,715 nanobio keywords.

**Technology**

As mentioned, in the following analysis we selected data from USPTO. Due to endogeneity concerns we could not use the same search words to isolate nanobio patents that we used to identify the nanobio articles. To delineate a nanobiotechnology field we followed two search strategies according to two different knowledge constructs within the field. In the first search we isolated patents through a static process. We used the nanobio search words identified by Fraunhofer-ISI (2002). We searched for patents that had any of the search words in either the titles or abstracts during the period 1971-2004. We obtained a dataset of 1,491 patents in that period. Characteristic of these patents is that they involved a specific technique or compound that is unique to nanobio and is found neither within nanotechnology or biotechnology.

We also employed a second search strategy to isolate patents that contained cross-pollination of knowledge from the biotechnology and the nanotechnology field. To isolate
these patents we looked at the overlap between nanotechnology and biotechnology patents.

The US Patent and trademark office have for many years had specific patent classifications for biotechnology innovations. We use the IPC based strategy used by Schmoch (2003) to identify biotechnology patents, and search the USPTO database in the period 1971-2004. This search generated a dataset of 43,310 patents. Figure 2 depicts the exponential growth in the patenting activity within the biotechnology field.

The search strategy for nanotechnology patents had to be mainly based on keywords, since the specific IPC-subclass B82B for this field was introduced in the year 2004 (Commerce 2004) and does not cover former years. We used a keyword search strategy suggested by Fraunhofer ISI Institute in Karlsruhe, which we found to be the most complete and validated by experts among the static keywords methodologies. Articles and reports have already been published using this search methodology (Bonaccorsi and Thoma 2007; Fraunhofer-ISI 2002).

We performed the search in the titles and the abstracts of the patents, and obtained a sample of 4,828 patents granted before May 2004. The nanotechnology patents, like the biotechnology patents, grew exponentially, especially in the last years (1996-2002). The USPTO has patented several thousands of inventions in nanotechnology, with around 4,500 patents filed in 2003.

To isolate the nanobio patents corresponding to the second knowledge combination we identified the overlap between the datasets of nanotechnology and biotechnology patents. This resulted in a sample of 406 patents over the period. We then combined the two datasets that we had obtained using the different search methodologies to obtain a complete sample of the nanobio space. This yielded a total of 1,573 patents. The first patent in the field was granted in the 1975, but only during the 1990s did the growth in nanobio patenting begin to accelerate.
**Commercialization**

The commercialization of a scientific and technological concept involves creating new products. We tracked the commercialization of concepts by retrieving company press releases - newswires - in the Lexis-Nexis database over the period 1980-2005. Our search strategy was two fold. The first was based on the same nanobio keywords that we used for patents, obtaining a sample of around 2,307 news events. Second, we considered the events that the announcing firms classified as pertaining to the biotechnology and nanotechnology industries. The second search strategy yielded an output of 730 press releases. We combined the two search strategies, obtaining 2,837 press releases.

**OPERATIONALIZATION AND MEASURES**

The unit of analysis used in this paper is *concepts* measured as keywords in a given year. We analyze how scientific concepts move between three institutional areas: Science, technology and commercialization represented by scientific articles, patents, and press releases.

To increase the validity of our measures we used only *authors’ keywords*, which are self-revealed by authors in scientific publications. These keywords are stated by authors and not computer generated through an algorithm.

ISI began to systematically collect authors’ keywords in 1991, which is the year that we begin our analysis. In the sample we excluded the keywords that are composed by four letters or less, thus reducing the probability of having homonym matched keywords. We isolated 133,128 nanotechnological and biotechnological keywords over the period 1991-

---

5 Eliminating keywords that are shorter than four letters changed the ration between nanobiotechnological and nanotechnological words with 0.3%. Examples of eliminated keywords include: “CN”, “201” and “SBTN”.

2004, as we defined before. We tracked the occurrence of these keywords in scientific publications over the period 1991-2004, having a non balanced panel of 243,022 observations.

**Dependent variables.** To test out hypothesis we develop the following three variables:

Technology: Measures the extent to which a scientific concept is integrated as part of a technology. *Technology* is a count variable measuring the occurrences of a specific keyword in a nanobio patent (title, abstract, or claims) the year after the keyword occurred in a scientific publication.

Commercialization: Measures the extent to which a concept has been commercialized. *Commercialization* counts the occurrences of a keyword in the title or body of a nanobio newswire in the year after the keyword occurred in a scientific publication.

Patent Quality Index: Measures the quality of the patents by relying on a factor index combining three patent indicators: forward citations which measure the technological importance of a patent; family size that correlates with the expected profitability of an invention; number of technological classes which is indicative of the patent scope. For further details see Hall et al. (2007).

**Independent variables:** They are operationalized in two dimensions:

Cross-pollination: Measures the cross-pollination effect. *Cross-pollination* is a binary variable, which takes on a positive value if in a given year a keyword has occurred in an article with both nanotechnological and biotechnological keywords.

Proximity: *Proximity* is a count variable, which measures the number of keyword occurrences in an article, where at least one author is affiliated with a commercial firm.

**Control variables.** In the model which explains the mobility of the keyword we include the following factors:

Length: Measures the length of a keyword. *Length* is a count time invariant variable, calculated by the number of letters in the keyword.
Absolute use: Measures the diffusion of a keyword in science. *Absolute use* is a count variable, calculated by the number of occurrences of a keyword in sciences in a given year.

Interdisciplinary: Measures the interdisciplinary effect and controls for journal effects. *Interdisciplinary* is a count variable, constituted by the mean number of fields in which a keyword appears according to the ISI journal citation report on subject categories.

Lagged variables: To take into consideration any reinforcing effect across the technological and commercial contexts, we add a one year lag of the variables “Technology” and “Commercialization” respectively when estimating concept mobility in the pooled panel regression (test of hypothesis 5).

In the equation estimating patent quality we rely on Bonaccorsi and Thoma’s (2007) previous analysis to construct controls. First, we cluster patents by controlling for characteristic at the level of the inventor according to whether she published in science: i) If all the inventors have scientific publications (labeled only-authors); ii) If all the inventors do not have publications in science (labeled “only-inventors”); iii) if none of the above two extreme cases is true, they fall into a residual group called “authors–inventors”.

Second, we control for whether the patentee is a firm, a public research organization, or an elite university, measured by whether the university was included in the top 50 of the Shanghai Index of World Research Universities in 2003.⁶

The descriptive statistics of the variable are reported in Table 1. We implemented a cleaning procedure for non ASCII standard and alphanumeric characters, removing and condensing the blank spaces, tabulations, commas and other word separators. Then we collapsed the occurrences of the cleaned keywords. As we can notice from Table 1, the average values of keyword occurrences are very low: the average diffusion in Science – *Absolute use* - is 2.11 and the median 1. We observe a similar scaling factor for technology and commercialization. The standard deviation is large compared to the mean value,

⁶ See [http://www.webometrics.info/](http://www.webometrics.info/), visited in May 2008. We chose the year 2003, because it was the first year that the Shanghai index was available.
suggesting a highly skewed distribution with some keywords frequently occurring in
publications, patents and newswires and a large majority of keywords that are rarely used. The
skewed distribution is confirmed by the maximum value statistics. Examples of frequent
keywords are: Nanoparticle, Biosensor, Nanoarray, Celladhesion, and Biodegradablepolymer.

EMPIRICAL RESULTS

Growth of the Field

*Scientific developments.* Since the early 1990s nanotechnology has undergone a dramatic
development. Figure 1 compares the growth in the general nanotechnology field to the growth
of nanobio (base year 1991). Nanotechnology has undergone an exponential growth in the
number of nanotechnology publications, but the growth rate for the nanobio subfield has been
even higher. Figure 1 shows that during the period 1991 to 2004 the nanobio field grew from
constituting 14 percent to 28 percent of the overall nanoscience field. The growth rate of
nanobio was thus significantly higher than the overall growth in nanotechnology.

*Technology developments.* The growth of nanobio technology parallels the growth of
nanobio science. Figure 2 depicts the dynamics of biotechnology, nanotechnology and
nanobio over time. Both nanotechnology and biotechnology have experienced an exponential
growth in the production of patents during the 1990s and early 2000s. The growth rate within
the nanobio field was, however, higher than in the two parent fields. During the 1990 to 2004
period the patent stock rose 9 times in biotechnology, 15 times in nanotechnology, but an
extraordinary 54 times in nanobiotechnology.
Commercialization development. Figure 2 illustrates the growth in press releases containing nanobio concepts from 1991 to 2004. During the 1980s nanobio commercialization grew slowly, but began escalating after 1990. Another dramatic increase in the amount of nanobio press releases happened around year 2000 with the subsequent years producing hundreds of nanobio announcements.

Determinants of Growth
We used a negative binomial count regression to model keyword mobility, because our dependent variable is a nonnegative integer that is not upper bounded. The use of the Poisson regression count data model would require more restrictive assumption on the shape of the distribution dependent variable’s variance (Wooldridge, 2002).

Moreover we adopted two estimation approaches. First we estimated a pooled negative binomial estimator with time controls. While this technique allows us to use a large number of observations in the estimation, it does not rule out potential endogeneity of the independent variables - that is the correlation of explanatory variables with the disturbance. If this assumption is violated then the consistency of estimates cannot be obtained with a pooled regression approach.

One important source of potential endogeneity is unobserved and time invariant specific effects at the level of the keyword, which could be correlated with the disturbances. It can be argued that a keyword is imprinted by a different level of novelty and growth potential since its inception and hence it is intrinsically different for all the others. For example when the “carbon tube” was discovered many scholars thought that this new idea could fuel important advancements in many scientific and technological contexts.

In the case of specific effects at the baseline unit, the econometric literature has proposed the panel fixed effect estimator that can generate consistent results when the time
dimension is sufficiently large.\(^7\) Hence the second estimation approach we used is a fixed
effect negative binomial estimator. In particular we followed a conditional maximum likehood
approach based on the annual counts of a keyword conditional on all counts for all years (see
Hausman, Hall and Griliches (1984) for more details regarding this approach). The fixed effect
estimator reduced the dataset to the keywords that appear in more than one period but
controlling for a source of potential correlation across the explanatory variables and the
disturbance. As Hausman et al. (1984) has shown the conditional negative binomial fixed
effect estimator has robust and consistent properties.

In Table 2 Model 1 and 2 report the results of the pooled negative binomial estimator,
whereas Model 3 and 4 show the results of the fixed effect panel estimation of the negative
binomial count model to control for unobserved heterogeneity at the level of the keyword.
Table 3 depicts the impact of cross-pollination on patent quality. The fixed effect panel
estimation is restricted to the keywords that occurred in at least five different years, which
reduced the sample from 133,128 to 10,109 keywords.\(^8\) This strong skewness in keyword use is
coherent with the general distribution of invention: Most innovations fail, but few innovations
are extremely successful.

\[\text{Table 2, and 3 about here}\]

\[\text{Table 2, and 3 about here}\]

\[\text{Hypothesis 1a and 1b: We find support for hypothesis 1a and 1b. If a concept}\]
\[\text{appears together with keywords pertaining to both nanotechnology and biotechnology then the}\]
\[\text{concept has a higher likelihood to both subsequently be integrated into a technology and to be}\]
\[\text{commercialized. This result supports our hypothesis that cross-pollination between concepts}\]
\[\text{from different disciplines creates ideas that are more likely to proliferate. Interestingly, this}\]

\(^7\) Other methods to analyze for endogeneity problem due to specific effects at the level of the keyword could
include random effects panel data estimation for controlling for random keywords characteristics.

\(^8\) It is worth to take into the account that this reduction could be associated with potential sample selection at the
level of the keyword.
positive effect of cross-pollination occurs in addition to the rough measure of whether a
color concept was published in a journal that spans multiple disciplines. Publication of the concept
in an interdisciplinary journal actually positively impacts the possibility that the concept will
later be commercialized, but the effect is smaller. In terms of elasticities, the panel estimation
suggests that a standard deviation increase in the cross-pollination effect has a positive impact
of about 13.2% on the probability that a keyword will be incorporated in a technology and
4.7% that it will later be commercialized.

\textit{Hypothesis 2:} We find support for hypothesis 2 at the 5% level. If a cross-pollinated
keyword appears in a patent it will increase the value of the patent. This is true even when we
control for whether the inventors were publishing in science, and whether the patent is held by
an elite research university, a public research organization or a firm.

\textit{Hypothesis 3a and 3b:} We find support for hypothesis 3a and 3b at the 1%
significance level. If a person affiliated with a private company presents a concept in a
scientific article then the concept has a higher likelihood of subsequently being incorporated
into a technology and of being commercialized.

The effect of proximity to market is much smaller than the effect of cross-pollination
both in the pooled and fixed effect panel regressions. In the latter case, if a keyword occurs in
an article that contains nanotechnology and biotechnology concepts, then the likelihood that it
will be incorporated in a patent is 13.2% higher than if no cross-pollination occurs. If an
author is affiliated with a private company the likelihood that the concept will be translated
into a technology is only 4.5% higher than if all the authors are scientists. Similarly, if a
concept is published in an article that contains both nanotechnology and biotechnology
concepts then the likelihood that it will later be commercialized is about 4.7% higher than if
no cross-pollination occurs, whereas the effect of the industrial affiliation of one of the authors
is only 1.7%.
**Hypothesis 4a and 4b**: We find support for both hypothesis 4a and 4b at the 1% significance level. The interaction effects have negative coefficients, which indicate that the positive effect of cross-pollination between nanotechnology and biotechnology on the probability that the concept will be commercialized only holds true if the authors are scientists. If the authors instead are affiliated with a company, cross-pollination actually has a negative effect on both the probability that it will be incorporated into a technology and that it will later be commercialized.

**Hypothesis 5**: We find support for hypothesis 5 at the 1% significance level. If a concept has already been incorporated into a technology it is 9% more likely that it will subsequently be commercialized. In the pooled regressions this effect is the second most powerful predictor of whether a scientific concept will be commercialized, although it is small in panel estimation.

**Overall goodness of the model**: Overall the model has statistically significant explanatory power, especially considering the limited number of variables included in the model. The model explains 17% of the variance with regards to whether a concept will be incorporated into a technology, and 11% of the variance with regards to whether a concept will subsequently be commercialized. These results show the strong predictive value of the mobility of concepts between science, technology and commercialization.

**Robustness checks**: To validate the robustness of the results we advanced a cross-sectional regression analysis for the first and the median year each keyword occurred in a scientific publication. Given that large part of the technological and commercial development of nanobiotechnology did not take place until the late nineties (see Figure 2) we had to extend the lag time for observing concept mobility from a 1 year to a 5 year window. The estimation of the negative binomial model with only a 1 year window is not possible, because there are

---

9 Note that one problem with the estimation of first occurrence is that the data has a left censoring bias. Due to the short life span of most keywords the left censoring bias is particularly problematic in the early years of the time series.
too few positive outcomes in the dependent variables.\textsuperscript{10} The results of the robustness check are reported in Table 2 Model 5-8.

The cross-sectional regression analysis based on the median year is consistent with the previously reported findings, whereas the regression based on the first year only supports the cross-pollination hypothesis. These findings indicate that in the beginning when a new concept is originated only the quality of the concept matters for its mobility, i.e. good ideas will travel no matter who produces them. For the first occurrence it does not matter whether a keyword is published by academic or industrial scientists. The mobility of subsequent discoveries based on the same concepts is, however, dependent upon their locus of inception. After the newness wears off a keyword will be more mobile if it is presented in science by an industrial scientist. Interestingly, we find that industrial scientists produce fewer of the new occurrences than of keywords in general: Whereas industrial scientists employ 16\% of the total keyword in the sample they introduce only 12\% of the first occurrences. This finding indicates that industrial scientists produce less highly novel knowledge; instead they tend to expand on existing knowledge in their scientific production.

Other robustness checks could dig further into the problem of the endogeneity of cross-pollination and keyword mobility. First, the fixed effects estimator could be compared with other panel data models - such as random effects - to account for potential endogeneity propelled by random keywords characteristics. Second, it could be argued that the cross-pollination variable is influenced by a selection problem at the level of the individual researcher: as we stated previously, the author’s participation in a social group - like an invisible college or academic department - might create an unobserved bias in their choice of keywords. In this case the robustness analysis could advance with a two stage selection model.

\textsuperscript{10} Another possibility to overcome this limitation could have been to estimate a zero-inflated negative binomial model, while continuing to maintain a one year window. However, that would require an extensive discussion of the more demanding assumption of the zero-inflated negative binomial model, which we do not think is warranted for a robustness check.
where in the first stage the determinants of cross-pollination are analyzed whereas in the second stage the keyword’s mobility equation is estimated.

**DISCUSSION**

The growth of new industries and commercial fields is central to the sustainability of economic growth within a modern society (Arora et al. 1998; Rosenberg 1998). Chemical engineering, for example, emerged from the oil and petroleum refining and dyes industries during the late 19th century. Indeed, the benefits to overall economic growth from discoveries made in chemical engineering were substantial and unfolded for decades.

In this paper we show that a new field, which is growing more rapidly than its two parent fields, is emerging at the intersection between nanotechnology and biotechnology. Our results indicate that the success of the field is partly driven by a cross-pollination of knowledge between nanotechnology and biotechnology. We base our analysis on robust estimation techniques (e.g. fixed panel estimation), which controls for unobserved heterogeneity at the level of the single keyword. We conducted robustness checks based on cross-sectional analyses, which confirmed the cross-pollination effect, but found that for the first occurrence of a keyword the locus of inception is not important, whereas it becomes important later on. Finally, we document that concepts that appear in scientific articles, which contain both nanotechnology and biotechnology concepts, have a higher likelihood of later being incorporated into technology and subsequently be commercialized. Studies have documented that the cross-pollination of knowledge generates more creative ideas and concepts (Hargadon and Sutton 1997; Hargadon 2003). We, however, demonstrate that the cross-pollination of knowledge contributes to knowledge dynamics by facilitating concept mobility.

We further show that the mobility of scientific concepts into technology is aided when one or more of the authors are affiliated with industry. It has been debated which role scientists with industrial affiliation play in the translation of knowledge between science and
Some researchers have claimed that industrial scientists only publish their findings in scientific journals if the knowledge does not have commercial value (Bird et al 1993). The argument behind this claim is that companies are reluctant to share any information that might provide their competitors with increased insight. Companies might thus choose to only publish information that is basic research, and thus far away from commercial possibilities. Our results counter this hypothesis. The concepts presented by industrial affiliates have a larger chance of appearing in a patent. Companies thus present concepts within scientific articles that contain commercial value.

Studies have found research conducted by or in collaboration with industrial partners is less innovative than research done purely for the sake of science. Evans (2004) shows that industrial partners and industrial funding decreases the innovativeness of plant biotechnology research. Within the nanobio field industrial affiliates also display conservatism in their publishing efforts. First industrial affiliates have a higher tendency than university scientists to include concepts in their publications that are common; in particular they introduce fewer completely novel concepts than their academic counterparts. Second the strong positive effect on commercialization of cross-pollinating concepts from nanotechnology and biotechnology is reversed for industrial affiliates. If a concept is published together with a person that works in a company cross-pollination diminishes the probability that the concept will be commercialized. The straight forward interpretation of this result is that the cross-pollinated concepts produced by industrial scientists have less commercial potential. It can, however, also be the case that firms do not disclose inventions if they are valuable, i.e. industrial scientists are only allowed to publish their scientific results as long as their results have little or no commercial value.

Future research might address the effect that the integration of a concept into a technology and the commercialization of a concept have on scientific development. In particular, the direction might be that of disentangling the existence and the intensity of the
feedback reinforcing processes of technological and industrial developments onto scientific production.

Many scholars have criticized the linear model of innovation, which only describes a movement of concepts from science to technology and subsequently to commercialization, but not the reverse knowledge flow (Kline and Rosenberg 1986; Mowery and Sampat 2005). Rosenberg (1982) has provided in-depth historical accounts of how industrial development aids the growth of science, by both providing scientists with results unexplainable by existing scientific theories, and by developing tools that facilitates data collection. This important dynamic relationship between science and technology has, however, not been tested on a large empirical dataset. Future research might thus explore the role of cross-pollination in stimulating the mobility of concepts from commercialization to science.
FIGURES AND TABLES

TABLE 1 Descriptive Statistics
(243,022 observations, 133,128 distinct keywords)

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>Std</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology</td>
<td>0</td>
<td>202</td>
<td>0.13</td>
<td>1.72</td>
<td>0</td>
</tr>
<tr>
<td>Commercialization</td>
<td>0</td>
<td>680</td>
<td>0.68</td>
<td>66.45</td>
<td>0</td>
</tr>
<tr>
<td>Cross-pollination</td>
<td>0</td>
<td>1</td>
<td>0.03</td>
<td>0.03</td>
<td>0</td>
</tr>
<tr>
<td>Interdisciplinary</td>
<td>1</td>
<td>15</td>
<td>1.68</td>
<td>0.69</td>
<td>2</td>
</tr>
<tr>
<td>Proximity</td>
<td>0</td>
<td>57</td>
<td>0.24</td>
<td>0.85</td>
<td>0</td>
</tr>
<tr>
<td>Length</td>
<td>5</td>
<td>133</td>
<td>16.70</td>
<td>60.86</td>
<td>16</td>
</tr>
<tr>
<td>Absolute use</td>
<td>1</td>
<td>499</td>
<td>2.11</td>
<td>47.38</td>
<td>1</td>
</tr>
</tbody>
</table>
FIGURE 1: Nano and Nanobio Publications (base year 1991)

FIGURE 2: Nano, Bio and Nanobio patents and Nanobio newswires over time (base year 1991)
TABLE 2: Regression Analysis of the determinants of keyword mobility  
(pooled panel regression: 243,022 observations, 133,128 distinct keywords)  
(negative binomial fixed effect regression: 78,830 observations, 10,109 keywords occurring in at least 5 year\(^1\))

<table>
<thead>
<tr>
<th></th>
<th>Main Results</th>
<th>Robustness Checks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pooled panel regression</td>
<td>negative binomial fixed effect regression</td>
</tr>
<tr>
<td><strong>MODEL 1</strong></td>
<td>Patents</td>
<td>Newswires</td>
</tr>
<tr>
<td>Cross-pollination</td>
<td>0.97***</td>
<td>1.11***</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>Proximity</td>
<td>0.08***</td>
<td>0.22***</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Commercialization</td>
<td>0.11***</td>
<td>0.00***</td>
</tr>
<tr>
<td>One Year Lag</td>
<td></td>
<td>(0.00)</td>
</tr>
<tr>
<td>Technology</td>
<td>0.88***</td>
<td>0.01***</td>
</tr>
<tr>
<td>One year Lag</td>
<td></td>
<td>(0.02)</td>
</tr>
<tr>
<td>Cross-pollination*</td>
<td>-0.18***</td>
<td>-0.03***</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Proximity</td>
<td>0.11**</td>
<td>0.06**</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Commercialization</td>
<td></td>
<td></td>
</tr>
<tr>
<td>One Year Lag</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technology</td>
<td>0.18***</td>
<td>-0.18***</td>
</tr>
<tr>
<td>Length</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Absolute use</td>
<td>0.06***</td>
<td>0.04***</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Year dummies</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Constant</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>R squared</td>
<td>17%</td>
<td>11%</td>
</tr>
</tbody>
</table>

### Elasticities at the mean value

<table>
<thead>
<tr>
<th></th>
<th>Elasticities at the mean value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross-pollination</td>
<td>2.2% ***</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
</tr>
<tr>
<td>Proximity</td>
<td>1.0% ***</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
</tr>
<tr>
<td>Commercialization</td>
<td>0.2% ***</td>
</tr>
<tr>
<td>One Year Lag</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Technological Lag</td>
<td>8.9% ***</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
</tr>
<tr>
<td>Cross-pollination*</td>
<td>-0.4% ***</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
</tr>
</tbody>
</table>

Notes: *** 1% level significance; ** 5% level significance; * 10% level significance
\(^1\)The fixed effect negative binomial regression is computed only for observations appearing in at least five years, because it exploits the time dimension variability to estimate the effects of the explanatory variables over dependent ones.
### TABLE 3: Linear Regression Analysis of Patent Quality
(1,022 patents applied over the period 1990-2000)

<table>
<thead>
<tr>
<th>Variable</th>
<th>MODEL 1</th>
<th>MODEL 2</th>
<th>MODEL 3</th>
<th>MODEL 4</th>
<th>MODEL 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross-Polination</td>
<td>0.014** (0.006)</td>
<td>0.013** (0.006)</td>
<td>0.014** (0.006)</td>
<td>0.014** (0.006)</td>
<td>0.013** (0.006)</td>
</tr>
<tr>
<td>Only-Authors</td>
<td>-0.162** (0.075)</td>
<td>-0.138* (0.075)</td>
<td>-0.162** (0.075)</td>
<td>-0.138* (0.075)</td>
<td></td>
</tr>
<tr>
<td>Only-Inventors</td>
<td>-0.192*** (0.062)</td>
<td>-0.192*** (0.062)</td>
<td>-0.192*** (0.062)</td>
<td>-0.187*** (0.062)</td>
<td></td>
</tr>
<tr>
<td>PROs Assignees</td>
<td></td>
<td>-0.063 (0.060)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Business Assignees</td>
<td></td>
<td></td>
<td>0.159** (0.071)</td>
<td>0.161 (0.072)</td>
<td></td>
</tr>
<tr>
<td>Elite Universities</td>
<td></td>
<td></td>
<td>0.267*** (0.103)</td>
<td>0.235*** (0.104)</td>
<td></td>
</tr>
<tr>
<td>Time Dummies</td>
<td>Yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Costant</td>
<td>Yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>R-squared</td>
<td>2.0%</td>
<td>2.8%</td>
<td>2.0%</td>
<td>2.5%</td>
<td>3.2%</td>
</tr>
</tbody>
</table>
REFERENCES


Gittelman, M., and Bruce Kogut. 2001. "Does Good Science Lead To Valuable Knowledge? - Biotechnology Firms and the Evolutionary Logic of Citation Patterns." *Workingpaper*.


