Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Knowledge Recombination Across Technological Boundaries: Scientists vs. Engineers

Marc Gruber, Dietmar Harhoff, Karin Hoisl,

MANAGEMENT SCIENCE
Vol. 59, No. 4, April 2013, pp. 837–851
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)
http://dx.doi.org/10.1287/mnsc.1120.1572

Knowledge Recombination Across Technological Boundaries: Scientists vs. Engineers

Marc Gruber
College of Management of Technology, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland, marc.gruber@epfl.ch

Dietmar Harhoff
Institute for Innovation Research, Technology Management, and Entrepreneurship, Munich School of Management, Ludwig-Maximilian University Munich, D-80539 Munich, Germany; and Centre for Economic Policy Research, London EC1V 3PZ, United Kingdom, harhoff@bwl.uni-muenchen.de

Karin Hoisl
Institute for Innovation Research, Technology Management, and Entrepreneurship, Munich School of Management, Ludwig-Maximilian University Munich, D-80539 Munich, Germany, hoisl@bwl.uni-muenchen.de

Building on the seminal work of Thomas J. Allen, we contribute to the emerging microlevel theory of knowledge recombination by examining how individual-level characteristics of inventors affect the breadth of their technological recombinations. Our data set combines information from 30,550 European patents with matched survey data obtained from 1,880 inventors. The analysis supports the view that inventors with a scientific education are more likely to generate patents that span technological boundaries (in our case, 30 broad, top-level technological domains) than inventors with an engineering degree. A doctoral degree is associated with increased recombination breadth for all groups of inventors. The breadth of an inventor's technological recombinations diminishes with increasing temporal distance to his education, but the differences between scientists and engineers persist over time. Our findings provide several new insights for research on inventors, the literature on organizational learning and innovation, and strategy research.

Key words: inventors; scientists; engineers; recombinant search; technological breadth; patent classes; innovation
History: Received September 1, 2009; accepted February 7, 2012, by Lee Fleming, entrepreneurship and innovation. Published online in *Articles in Advance* July 18, 2012.

1. Introduction

The relationship between inventors and their output is a classical research domain for economists and management scholars (Schumpeter 1934, Griliches 1957, Gambardella 1995, Harhoff et al. 1999). The topics studied in this literature include the distribution of productivity among scientists and inventors (Lotka 1926, Narin and Breitzman 1995, Zucker and Darby 1996) and the role of collaborations and networks for inventive activities (Allen 1977, Rosenkopf and Nerkar 2001, Fleming et al. 2007).

In the present study, we focus on a key determinant of inventive outcomes that has largely been neglected in the prior literature, that is, systematic differences arising between the two main groups of inventors—scientists and engineers. These differences have been highlighted in the seminal work of Allen (1977). However, they have not been considered in contemporary theorizing about technological knowledge recombination processes and outcomes, nor have they been subject of a large-scale empirical test. Allen (1977) argues that scientists and engineers differ in terms of their knowledge endowments, preferences, and behaviors when engaging in inventive activities. He emphasizes that these differences not only lie in the kinds of people who are drawn to the two domains, but also in the different kinds of education and socialization they receive. We extend Allen's (1977) propositions by considering the level of education that the inventor has attained, and by taking into account the time dimension—as knowledge becomes obsolete and as effects of educational socialization get weaker over time (Bosworth 1978, Schott 1978, Pakes and Schankerman 1984), the differences between a scientific or engineering-based education may well diminish.

We relate Allen's (1977) work to the *breadth* of the inventor's technological recombination activities. This is arguably one of the most important outcome variables in technological innovation. The recombination of different technologies to generate radical innovations has been viewed as the "holy grail" of innovation research since the first half of the 20th century (Schumpeter 1934, Hargadon 1998). For instance, Hargadon (1998, p. 210) suggested the importance of "technological fusion," that is, "the combination of existing technologies from several industries, and the powerful market effects these combinations can create," and provided the example of the optoelectronics industry, which emerged from combining scientific discoveries in electronic, crystal, and optics technologies.

Our analysis is based on original survey data obtained from 1,880 inventors. These data were matched with comprehensive register information covering 30,550 patents filed at the European Patent Office (EPO). The combination of these two data sources is unique and enables us to uncover largely novel patterns of inventor characteristics and outcomes of invention processes. By combining survey data with longitudinal data on inventor careers, we also go considerably beyond most contributions that are based on correlational analyses of patent citations. Our results add key insights to emerging microlevel theories on technological recombination. In particular, they illuminate the systematic influence of the inventor's educational background on the breadth of his technological recombination and, thus, the type of inventive output that is generated. Furthermore, our study contributes to the innovation literature by introducing an improved index for capturing technological recombination breadth.

2. Theoretical Background and Hypothesis

Analysts of innovation and technological progress have long argued that the process of innovation critically relies on the recombination of existing ideas and artifacts. As Nelson and Winter (1982, p. 130) put it: "the creation of any sort of novelty in art, science, or practical life—consists to a substantial extent of a recombination of conceptual and physical materials that were previously in existence." This perspective on innovation rests on an implicit notion of boundaries between different technological domains, that is, certain components or technologies are seen as belonging together in a particular domain, as they address a common real-world problem, share some underlying pattern of socially constructed meaning, or draw on a distinct theory, method, language, code, etc. (von Tunzelmann 1998, Hargadon 2006).

Given that inventors are at the locus of the knowledge recombination activity, and that it is through their knowledge and skills that technological components get recombined and potentially path-breaking inventions created, one would expect that the literature holds rich insights on how individual-level differences among inventors affect knowledge recombination. Yet, most work examining knowledge recombination in technological innovation has been conducted at the *firm level*. For instance, studies have investigated the evolution and performance of firms pursuing different types of knowledge recombination (Helfat 1994, Rosenkopf and Nerkar 2001, Miller et al. 2007) and the mechanisms that firms use to acquire distant technological knowledge (Stuart and Podolny 1996, Capron et al. 1998, Rosenkopf and Almeida 2003). However, examinations at the *individual level* can lead to important new insights on knowledge recombination outcomes, as recent work by Fleming et al. (2007) indicates. Their results show how several individual-level factors (the inventor's network cohesion, breadth of experience, change of employers, and external ties) affect the creative generation of new *subclass* combinations in inventions.[1]

In the present study we continue this line of microlevel research on knowledge recombination by drawing on and extending the work of Allen (1977). Allen argues that one of the most important individual-level characteristics of inventors is their *educational background in science or in engineering*, because inventors with these distinct educational endowments draw on different knowledge bases, possess different skill sets, and are motivated by different factors when pursuing their research. Focusing on inventors with an education in either science or engineering also has empirical appeal, because these two groups typically dominate among inventors (e.g., they account for more than 90% of inventors in our data). Corresponding to Allen (1977), we define as inventors with a science degree those who obtained their education in the natural sciences, that is, "disciplines that deal only with natural events … using scientific methods" (Ledoux 2002, p. 34) such as astronomy, biology, chemistry, earth science, physics, atmospheric science, oceanography, and materials science. Conversely, we define as inventors with an engineering degree those who obtained an education in domains such as chemical engineering, civil engineering, electrical engineering, mechanical engineering, and computer engineering.[2]

[1] In the broader innovation literature, several inventor characteristics such as the inventor's personality traits, age, and mobility have been examined in relation to output measures such as the inventor's patent count (e.g., Zucker and Darby 1996, Simonton 1999, Jones 2005, Hoisl 2007).

[2] It is important to note that scientific and engineering studies are not fully mutually exclusive educational domains, as engineering curricula have advanced to include a larger portion of science content. In other words, any potential overlap between curricula will make it harder for us to detect differences between the inventive behaviors of these groups. Hence, our empirical tests tend to be conservative. Furthermore, we note that our distinction is based on the individual's field of education, which means that—even though educated in science—individuals may work later on in more applied domains and vice versa.

Although inventors will generally face considerable difficulties when seeking to recombine knowledge across technological boundaries, we argue that inventors with a science education have superior abilities or skills in performing such recombination relative to inventors with an engineering degree. First, inventors with a science education will have an advantage in knowledge recombination across technological boundaries, because they have acquired a more abstract understanding of the technological problem-solving process and, because they have reflected on and honed their learning activities, are more likely to have developed an ability in metalearning that facilitates the assimilation of technological knowledge from other domains (Gibbons and Johnston 1974; Woolnough 1991, 1994). Because of the contextual embeddedness of distant technological knowledge (Nonaka 1994), an important part of this metalearning ability is the skill to characterize phenomena in general forms and to abstract from the complexity of particular objects or systems (Nersessian 2002).

Second, the relatively more refined scientific knowledge of inventors with a science education is likely to provide them with the equivalent of a map in technology searches, that is, a stylized presentation of the technology area that is being searched (Fleming and Sorenson 2004). Although the theoretical guidance that scientific knowledge can offer to searches of other technology areas may be incomplete and rough, theory may still provide inventors with some working knowledge of these areas and, hence, guide their search of scientific knowledge sources toward those contributions that will be more likely to provide useful answers (David et al. 1992, Gambardella 1995).

Third, inventors with a scientific education should be better able to assess the adequacy of knowledge components originating from another technological domain. In particular, their understanding of general scientific principles and of technological landscapes should help them in comprehending, analyzing, and assimilating far-flung technological knowledge (Cohen and Levinthal 1990). They will thus be able to make more informed choices about the distant technological knowledge that is available for recombination (Gambardella 1995).

Fourth, the creativity literature indicates that truly innovative outcomes are more likely to appear when people are willing to engage in the cognitive labor of exploring possible pathways to generate original though sensible combinations (Ward et al. 1997). In this vein, studies examining science and engineering students point to fundamental differences in their goals and motivations (e.g., Blade 1963, Allen 1977, Woolnough 1994), which lead us to believe that the former will, on average, execute the cognitive processes involved in technological recombination with more intensity and intellectual curiosity than people with an engineering education.

Taken together, these arguments suggest that inventors with a scientific education should have stronger skills, abilities, and curiosity for recombining knowledge across technological domains than inventors with an engineering background.

Hypothesis 1. *Inventors with a scientific education will be more likely to generate inventions that recombine knowledge across technological domains than inventors with an engineering education.*

Beyond the content of an individual's education, the intensity with which that individual has been exposed to that content should influence his ability to recombine technological knowledge from different domains. Several arguments suggest that the inventor's *level of education* affects the breadth of his recombinations.

First, the prior literature suggests that—irrespective of the field of education—inventors with higher educational attainments will possess a more abstract understanding of technological problem solving. As Gibbons and Johnston (1974, p. 239) explained: "It appears to be this second-order form of knowledge, the 'knowledge of knowledge,' whereby the problem-solver with a university education … knows where and how to go about seeking the kinds of information he needs, which constitutes an important difference between the university graduate and the worker with a part-time practically-oriented education." Notably, this difference seems to be associated less with the specific knowledge or techniques learned while undergoing formal education than with a more general ability in obtaining new knowledge, assessing its adequacy, and recombining it with other knowledge components.

Second, inventors are boundedly rational individuals whose recombination activities are constrained by their cognitive abilities (March and Simon 1958). For example, research on cognitive abilities indicates that the most complex set of interrelationships an individual can process in working memory is a three-way interaction (Halford et al. 1994). Thus, beyond some number of potential components and their combinations, inventors may become overwhelmed by the complexity of the combination possibilities that a distant technological domain offers and will not arrive at useful outcomes. Still, some inventors may have better cognitive abilities in dealing with such complexity in technological innovation than others, as they have developed knowledge structures (schemata) to cope with the many stimuli and uncertainties encountered in their inventive activities (Gagné and Glaser 1987, Walsh 1995). Because the attained level of formal education is reflective of an individual's cognitive ability and knowledge structures (Pelled 1996), inventors with higher educational attainment should have superior abilities in dealing with the complexity of knowledge recombination across technological domains than inventors with lower educational degrees.

Third, because technological domains isolate and constrain the use of knowledge, inventors require a certain level of openness to study other domains, to import the identified knowledge, and to combine it in new ways. In this vein, prior research indicates that individuals who are more educated tend to be more receptive toward innovation, and are more likely to engage in boundary-spanning activities (Hambrick and Mason 1984, Hargadon 2006).

These arguments all suggest that inventors who attained a higher level of education should have relatively better abilities and higher willingness to recombine knowledge across technological domains.[3]

Hypothesis 2. *The breadth of an inventor's technological recombinations will increase with the educational attainment of the inventor.*

The extant literature on scientists' productivity indicates that the *quantitative* output of inventors will change over their life cycle (e.g., Levin and Stephan 1991). Focusing on changes in the *qualitative* output over time, we argue that the effects of inventors' education on the breadth of their technological recombinations will diminish with increasing temporal distance to their education.

On one hand, it seems that with increasing distance to their education (and, thus, increasing years of work experience), inventors will have been for a longer time "at risk" of getting to know other technological domains (Fleming et al. 2007). As a result, these more experienced inventors will have amassed more technological knowledge that is available for recombination—including distant technological knowledge.

On the other hand, however, this effect is countered by the inventor's education knowledge becoming, at least to some extent, obsolete with the passage of time (Bosworth 1978, Schott 1978, Pakes and Schankerman 1984, Park et al. 2006). The knowledge acquired by an inventor during his education—the map that facilitates his technological search (Fleming and Sorenson 2004)—may have less and less relevance for his technological recombination activities, making it increasingly challenging and cumbersome to recombine technological knowledge across domains. On average, inventors should thus be less likely to perform distant recombinations than inventors possessing more recent cognitive maps.

In addition, inventors may seek to reap the salary benefits associated with being a specialist in a particular technological domain and thus engage in narrow recombination (Leahey 2006). Specialization is associated with an increasing quantitative research output, as individuals acquire in-depth knowledge of the domain and produce their output more efficiently (Birnbaum 1981, Leahey 2007).

In light of these arguments, we propose that the latter two effects will outweigh any broadening effects that may arise from increased exposure to technological domains over time. The specific nature of this relationship is ultimately an empirical question—one that we investigate with the following hypothesis.

Hypothesis 3. *The breadth of an inventor's technological recombinations will diminish with increasing temporal distance to his education.*

3. Research Design

3.1. Data

To examine our hypotheses, we require data on the education of inventors, their inventions, and a number of other factors that potentially influence the outcomes of invention processes. Because no public data set offers the information required for this study, we administered a survey to inventors and complemented the survey information with data from patent databases that provide detailed information on the underlying inventions, and about the inventors' activities across time (Harhoff et al. 1999). Patent data have been found useful to trace the recombination of technological knowledge components in a number of previous studies (Fleming 2001, Agrawal et al. 2006). Using survey data in addition to patent data provides an important advantage, because we do not have to infer a number of key variables in our study from the same source, which entails the danger of common method bias. We describe the combined data collection efforts in turn.

3.1.1. Survey Data. We collected primary data with a self-administered survey of inventors. After conducting an extensive pilot study (including interviews with 10 inventors), we developed the survey instrument and pretested it with 60 inventors. The survey instrument was sent out together with a

[3] The curriculum that individuals with an advanced degree have been exposed to should also have an effect on the breadth of their technological recombinations. Yet, these considerations are less clear than the preceding arguments: on the one hand, an advanced degree often requires the person to be exposed to more streams of research, taking courses in more fields than in an undergraduate or graduate curriculum. On the other hand, however, people may specialize in a certain technology domain during their advanced education and, thus, are exposed to narrower technological knowledge.

free-franked return envelope to our sample of inventors listed on 10,500 European (EP) patents. Based on a list of all granted EP patents with priority dates between 1993 and 1997 (total of 15,595 EP patents), 10,500 EP patents listing inventors residing in Germany were chosen by stratified random sampling. (A stratified random sample was used to oversample potentially important patents.[4]) The survey instrument was sent to the first inventor listed on the patent document, as interviews conducted with patent attorneys and patent examiners indicated that the order of the names on the patent document does not correspond to any pattern. In total, answers were received from 3,049 inventors, resulting in a response rate of 32%.

3.1.2. Patent Data. In the present study we use patent data obtained from the EPOLINE and the PATSTAT databases made available by the EPO. Patents at the EPO are typically associated with at least one publication that discloses the application. This publication also contains the content of the examiner's search report with backward citations to relevant prior art. If the patent passes examination successfully, there will also be a later publication of the granted patent.

Our information taken from EPO patent publications includes the names and the addresses of the inventors, the identity of the applicant company, the content of the search reports (i.e., backward citations), the type of invention, and the technology class(es) the patented invention is associated with. It is this latter information that is of focal interest to our research. At the EPO, patents are classified according to the "International Patent Classification" (IPC). The IPC is applied in 54 countries and has been in use since 1975. Overall, the IPC contains about 70,000 entries represented by seven-digit alphanumeric classification symbols. Each patent application is assigned by the patent office to one or more classification symbols corresponding to the invention (for instance, F03D 1/02 denotes wind motors with a plurality of rotors, typically used in the production of electrical power from wind). To ensure comparability and consistency, examiners have to follow precise guidelines on how to classify patent applications. Because the IPC serves as the basis for assigning the patent application to the examiners at the EPO (the latter drafts a report listing prior art that might impede patentability, that is, violate the novelty requirement), the office has a strong incentive to classify patent applications carefully.

Technological classifications employing the IPC system are typically based on the information contained in the description of the technological invention as well as the examples, drawings, and claims provided in the application document. This is a key difference between the IPC system and the U.S. Patent Office Classification (USPOC). Whereas the USPOC classifies patents according to the claims stated within the application document (i.e., the scope of protection), the IPC system considers the complete technological information contained in the application document, and thus classifies patents with respect to the technologies associated with the invention (OECD 1994). For a study interested in the recombination of technological knowledge components across technological areas, the IPC thus provides a more suitable classification system than the USPOC. Because the EPO is an independent classification authority, the assigned technology classes are determined objectively by the examiner, and not by the inventors themselves.

Several proposals exist that convert the over 70,000 symbols into a technological nomenclature suitable for statistical analyses. In this study, we use a nomenclature proposed by the German Fraunhofer Institute for Systems and Innovation Research and the French Intellectual Property Institute to form largely homogeneous technology groups. This classification aggregates the IPC classes to 30 technological classes (OECD 1994) such as telecommunications, optics, and biotechnology.[5] Depending on the type of invention, patents can fall either within or across these 30 classes.

Our analysis will make use of these 30 technological classes, which represent an extremely broad form of technological recombination and thus allow for a fairly conservative empirical test of our hypotheses. In proxying technological components (areas) with patent classes, we follow earlier studies on technological recombination (e.g., Fleming 2001). It is important to note, however, that unlike most extant research,

[4] The sample includes all opposed patents (1,048) and patents that were not opposed and whose citation rates were in the top decile (5,333). A random sample of 4,119 patents was drawn from the remaining 9,212 patents. The results described below do not change qualitatively when we apply population sampling weights. We oversampled important patents because one of the purposes of the broader research project was to analyze the determinants of patent value. Oversampling led to a sufficiently large number of valuable patents in the final sample.

[5] The 30 technological areas comprise electrical devices/electrical engineering, audiovisual technology, telecommunications, information technology, semiconductors, optics, analysis/measurement/control, medical engineering, organic fine chemistry, macromolecular chemistry/polymers, pharmaceuticals/cosmetics, biotechnology, materials/metallurgy, agriculture/food, general technological processes, surfaces/coating, material processing, thermal processes and apparatus, chemical industry and petrol industry/basic materials chemistry, environment/pollution, machine tools, engines/pumps/turbines, mechanical elements, handling/printing, agricultural and food machinery and apparatus, transportation, nuclear engineering, space technology/weapons, consumer goods and equipment, and civil engineering/building/mining.

we do not use the technological classification of the resulting inventive output for our analysis, but utilize the technological classification of the references in the search report (the prior art that was regarded as relevant for the patentability of the underlying invention). Hence, our measures should provide a relatively more precise proxy for the process of recombinant technological search than those proxies relying on the technological classification of the focal patent itself.

3.1.3. Matching of Survey and Patent Data. The survey data was merged with bibliographic and procedural information from the EPOLINE and PATSTAT databases. We carefully identified, screened, and assembled the complete patenting history of the surveyed inventors. To trace the patent applications of each inventor over time, the EPOLINE database was used to search for all patent applications belonging to the 3,049 inventors with priority dates between 1977 and 2003. The search procedure resulted in a total of 39,410 EP patent applications.[6] Because it is not possible to reliably assess technological breadth for inventors who are only responsible for a very small number of patents (as an increasing number of patents increases the precision of our measure), we excluded inventors from the sample who are listed on less than three patent applications between 1977 and 2003. As a robustness check, we also excluded inventors with (a) less than 5 and (b) less than 10 patents. Results did not change in terms of coefficient signs or significance.

Our inventor survey data includes information on 3,049 individuals. Because we exclude inventors with fewer than three patents (579 cases), inventors with a nonacademic background (255 cases), and patents owned by individuals or public organizations (158 cases), and because we encounter missing variables in some of the independent variables (177 cases), the final data set includes information on 1,880 inventors in our regression analysis. These inventors were named on 30,550 EPO patents; for these we have obtained all relevant bibliographic data.

3.2. Definition and Measurement of Variables

3.2.1. Dependent Variable. *Technological Recombination Breadth.* We measure the breadth of the inventor's activities from information contained in the search reports associated with these patents. Note that each backward citation (i.e., a patent referenced in the search report) can itself be associated with multiple technological areas. Let referenced patent j—where j belongs to the set $(1, \ldots, J)$ of references (backward citations) for a given patent—be distributed over K discrete technological domains. Then s_{jk} is patent j's share of technology classifications associated with technological area k. Shares s_{jk} aggregated over all K areas sum to unity for each of the J referenced patents.

The extant literature proposes several measures that could be used to assess the breadth of technological recombinations. As we will discuss in some detail below, however, none of these measures—the Herfindahl index (e.g., von Tunzelmann 1998), the index used by Lerner (1994), and the index proposed by Leahey (2006)—captures the breadth of an inventor's technological recombinations in a fully satisfactory manner.

Perhaps the most widely used measure for assessing the breadth of activities is derived from the Herfindahl index. For a patent with J backward citations that are associated with technical areas $k = 1, \ldots, K$, it can be defined as

$$B_{\mathrm{HERF}} = 1 - \sum_{k=1}^{K} \left(\frac{\sum_{j=1}^{J} s_{jk}}{J} \right)^2. \qquad (1)$$

In the context of the present study, a particularly critical shortcoming of this measure is that it is invariant if the distribution of two patents were broadened simultaneously. Consider two patents mentioned in a search report of a patent A with distribution $(1, 0, 0)$ and $(0, 1, 0)$ over three technological domains. The breadth measure for patent A would compute to 0.5. Consider now two patents listed in the search report of patent B with distribution $(1, 1, 0)$ and $(1, 1, 0)$. Patent B yields the same Herfindahl measure of breadth as patent A. However, in the first case (patent A), the references consist of two maximally specialized patents, whereas in the second case (patent B), the search report contains two nonspecialized patents.

This problematic property is shared by the breadth measures proposed by Lerner (1994) and Leahey (2006). Specifically, Lerner (1994) proposed a four-digit count measure, that is, the cumulative number of unique patent classification codes. Leahey (2006) argued that a breadth measure of scientists should also take the number of works produced into account—for scientists with a relatively large number of works, breadth is easier to produce than for scientists with fewer works—and proposed to measure breadth as the ratio of the cumulative number of unique classification codes and the cumulative number of patents. Using the example given above, these measures yield breadth measures of 2 (Lerner 1994) and 1 (Leahey 2006) for *both* patents A and B—in

[6] Following an indicator-based approach comparable to that of Trajtenberg et al. (2006), inventor- and patent-related information (inventor names, addresses, applicant names, and IPC classes) was used to search for patents belonging to the same inventors. Results were checked manually to remove false matches. A detailed description of the matching procedure can be found in Hoisl (2007).

Table 1 Technological Recombination Breadth: Comparison of Alternative Indexes

Patents listed in the search report	Area 1	Area 2	Area 3	Area 4	Area 5	(1) Traditional Herfindahl-related index $B_{JK}=1-\sum_{k=1}^{K}\left(\frac{\sum_{j=1}^{J}s_{jk}}{J}\right)^2$	(2) Index proposed by Lerner (1994) $B_{\text{Lerner}}=\#CCs$	(3) Index proposed by Leahey (2006) $B_{\text{Leahey}}=\frac{\#CCs}{\#PATs}$	(4) Adapted Herfindahl-related index $\bar{B}_{JK}=1-\sum_{k=1}^{K}\left(\frac{\sum_{j=1}^{J}s_{jk}}{S}\right)^2$
Inventor 1									
Patent 1	x					$1-HI_{I1}=1-\left(\frac{3}{3}\right)^2=0$	1	$L_{I1}=\left(\frac{1}{3}\right)=0.33$	$1-\overline{HI}_{I1}=1-\left(\frac{3}{3}\right)^2=0$
Patent 2	x								
Patent 3	x								
Inventor 2									
Patent 1	x	x				$1-HI_{I2}=1-\left(\frac{0.5+1}{3}\right)^2$	2	0.67	$1-\overline{HI}_{I2}=1-\left(\frac{0.5+1}{4}\right)^2$
Patent 2	x					$+\left(\frac{0.5+1}{3}\right)^2=0.5$			$+\left(\frac{0.5+1}{4}\right)^2=0.72$
Patent 3		x							
Inventor 3									
Patent 1	x	x							
Patent 2	x	x				0.50	2	0.67	0.87
Patent 3	x	x							
Inventor 4									
Patent 1	x	x							
Patent 2		x	x			0.61	3	1	0.90
Patent 3		x	x						
Inventor 5									
Patent 1	x								
Patent 2		x				0.67	3	1	0.67
Patent 3				x					
Inventor 6									
Patent 1	x	x							
Patent 2		x	x			0.78	5	1.67	0.94
Patent 3				x	x				

Note. $\#CCs$, cumulative number of unique classification codes; $\#PATs$, cumulative number of patents; s_{jk}, share of patent j associated with technological field k.

other words, these measures are unable to capture the fact that the patents mentioned in the second search report show greater technological recombination breadth at the patent level.

Given these problematic properties, we adapt the first breadth measure based on the Herfindahl index so that it rewards technological recombination *at the referenced patent level*. Our adapted breadth index is defined as

$$\bar{B}_{JK}=1-\sum_{k=1}^{K}\left(\frac{\sum_{j=1}^{J}s_{jk}}{S}\right)^2, \qquad (2)$$

where S is the cumulative number of unique classifications attributed to patents in the search report (i.e., the cumulative number of distinct technological domains). For patent A, described above, S is equal to 2, and thus breadth computes to 0.5. In the second case (patent B), S is equal to 4, and breadth is 0.875.

To obtain a more intuitive understanding of this adapted breadth measure and its advantages over existing measures, Table 1 shows the calculation of index values for different types of inventors. As an illustration, consider the case of inventor 5, that is, someone whose patent does not draw on earlier patents mentioned in the search report that recombined technological knowledge within each patent, even though the references are assigned to several technological domains. Comparing the existing indexes with our adapted measure, we see that the former rank this inventor as second highest, whereas the same inventor is, more correctly, ranked second lowest in terms of technological recombination breadth using our adapted Herfindahl index.[7]

[7] Extant research has also relied on other measures of patent breadth, yet as explanatory variables. For instance, Trajtenberg et al. (1997) proposed two measures of patent breadth relying on patent citations, that is, the generality index and the originality index. Whereas these measures suffer from the problems discussed in relation to the traditional Herfindahl index, they are also problematic in the context of European patent data, as, for instance, the generality index can only be calculated for patents with at least one citation (otherwise the index values are assumed to be zero). Yet, as EP patents receive fewer citations than U.S. patents (average citation counts: EP, 4.37; U.S., 12.96; see Michel and Bettels 2001), with a large proportion of EP patents receiving no citation at all, the index values would be zero for 48% of patents in our sample.

Furthermore, studies of knowledge recombination typically use the breadth of the patent application (knowledge output) as a proxy for the degree of knowledge recombination during research and development (R&D) (knowledge input). The fact that we use the technology classification of the patent references mentioned in the examiner's search report rather than the technology classes assigned to the patent application helps to overcome this limitation. On average, search reports of EP patents in our sample contain 4.6 patent references. These references are on average associated with six IPC classes.

3.2.2. Independent Variables. *Field of Study.* We created a set of dummy variables to distinguish between three fields of study: science (e.g., physics, chemistry), engineering (e.g., mechanical engineering, process engineering), and other fields (e.g., mathematics, economics). The complete coding routines are available upon request. To get deeper insights into the role of science with respect to knowledge recombination breadth, we split the science dummy into three dummies: "physics," "chemistry," and other "scientific fields."

Doctoral Degree. We employ a dummy variable to flag inventors who attained a doctoral degree (1) and those with lower educational degrees (0).

Years Since Final Educational Degree. This variable captures the number of years that have passed since the inventor completed his education. This variable is calculated at the level of the inventor's patents, that is, for each patent, the year in which the inventor obtained his highest degree was subtracted from the priority year of the patent (year of first application).

3.2.3. Controls. *Mobility.* Mobile inventors may have been exposed to a larger number of technological domains. We thus created a dummy variable indicating whether the inventor has changed his employer. Following prior research, we defined a move of an inventor as when successive patents have different assignees (Hoisl 2007, Marx et al. 2009). The dummy was derived manually on the basis of the applicants listed on the EP documents and information drawn from our inventor survey.

Average Technological Recombination Breadth Firm. Some firms operate using a broader range of knowledge than others (Pavitt et al. 1989). To calculate a technological breadth index for each employing firm, we searched for the patent portfolio of each applicant. Then, we calculated a special breadth index for each firm's patent portfolio, following the procedure described for the dependent variable. This index takes into account that some inventors have changed their employers, and that the recombination breadth of firms has changed over time.

Average Inventor Group Technological Recombination Breadth. A detailed four-step procedure was used to develop a proxy that captures the average composition of the inventor's group of collaborators. First, we searched for all coinventors, i.e., all inventors listed on at least one patent of the focal inventor. Second, we identified all patents of these coinventors between 1977 and 2003. The search procedure resulted in 25,393 coinventors, involved in a total of 240,886 EP patent applications. Third, for each coinventor's patent portfolio, we calculated our technological breadth measure following the procedure described for the dependent variable. Fourth, we derived the average group composition variable by calculating the mean technological breadth of all of the focal inventor's collaborators.

Number of References per Patent. To control for the fact that recombination breadth as well as the precision of the measure increases with the number of references used to calculate the recombination breadth index, we controlled for the number of references listed in the search reports.

Classification Authority. We created three dummy variables to distinguish between patents classified by the EPO (87% of all cases), the Japanese Patent Office (JPO) (9% of all cases), or any other classification authority. Before 1988, the JPO required patent applications to be limited to a single claim, resulting in an inflation of narrow patent applications. Although multiple claim applications have been allowed since 1988, Japanese patent officers favor narrow patent applications containing only a small number of claims (Kotabe 1992).

Average Number of Claims. Patent claims define the scope of an invention for which patent protection is requested. This variable measures the average number of claims an inventor is requesting for his patents. A larger number of claims may result in multiple classifications. Because the size of this effect may be decreasing at the margin, we also include the square of the number of claims as a control variable.

Frequency of Technology Combination. Because particular technological elements belong together in technological fields, recombination between these elements is more likely to occur. Following Fleming (2001), we thus control for the frequency with which different technological areas have previously been combined, that is, listed jointly on EP patent applications. First, we calculated the cumulated yearly frequency of all possible combinations of the 30 technological areas listed on EP patent applications for the total number EP patent applications, starting with foundation of the EPO in 1978. Second, we assigned the resulting frequencies to the inventors' patents on the basis of the priority year.

Firm Size. Studies indicate systematic differences in research conducted by small and large firms

(Nelson 1959). To control for this type of variation in our data, we follow prior studies (e.g., Rosenkopf and Almeida 2003) and include a variable indicating the size of the employers. Information on firm size—operationalized as the number of employees—was obtained through extensive internet research.

Technological Areas. Because the underlying state of technology in part determines the potential for technological innovation (Klevorick et al. 1995) and because different technological areas show different propensities to patent inventions, we control for this variation in our data by including dummies for the different technological areas the patents have been assigned to.

Time Trend. To control for a potential time trend we add dummy variables capturing the priority years of the inventors' patent applications.

3.3. Methods

Our dependent variable can vary between 0 and 1, yet, empirically, the upper limit is not reached (max = 0.90). We use a random effects Tobit panel regression model with censoring at zero to estimate the determinants of the technological breadth of the inventors. The model is estimated at the level of inventor–priority year groups, which leaves 13,110 observations for 1,880 inventors. In our estimation function, we specify the latent (uncensored) variable y^*_{it} to depend on regressors, an idiosyncratic error ε_{it}, and an inventor-specific error α_i:

$$y^*_{it} = x'_{it}\beta + \alpha_i + \varepsilon_{it}, \tag{3}$$

where $\alpha_i \sim N(0, \sigma^2_\alpha)$, and $\varepsilon_{it}, \sim N(0, \sigma^2_\varepsilon)$. Because of left censoring, we observe

$$y_{it} = \begin{cases} y^*_{it} & \text{if } y^*_{it} > 0, \\ 0 & \text{if } y^*_{it} \le 0; \end{cases} \tag{4}$$

y_{it} refers to recombination breadth as revealed in an inventor's patents. The regressors x_{it} comprise the inventors' field and level of education and temporal distance to the final degree as well as a number of control variables related to the work environment of the inventor, e.g., the recombination breadth of coinventors and the recombination breadth pursued by the employer as well as for variables capturing the invention and the patent system (e.g., number of claims and classification authority).

4. Results

4.1. Descriptive Statistics

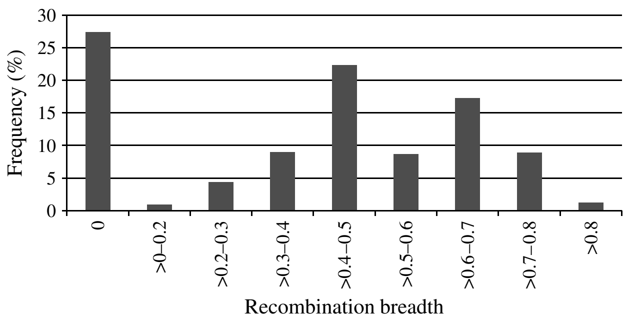
Descriptive statistics and the correlation matrix are reported in Table 2. Correlations between independent variables are relatively low, indicating that

Table 2 Descriptive Statistics and Correlation Matrix

Variables	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1 *Recombination Breadth Patent*	0.39	0.27	1																		
2 *Field of Study—Engineering*	0.44		−0.18*	1																	
3 *Field of Study—Science*	0.53		0.19*	−0.93*	1																
4 *Field of Study—Physics*	0.06		0.02*	−0.22*	0.24*	1															
5 *Field of Study—Chemistry*	0.44		0.18*	−0.78*	0.84*	−0.22*	1														
6 *Field of Study—Other Science*	0.03		0.02*	−0.15*	0.16*	−0.04*	−0.15*	1													
7 *Field of Study—Other Fields*	0.04		−0.04*	−0.17*	−0.20*	−0.05*	−0.17*	−0.03*	1												
8 *Years Since Terminal Degree*	16.77	9.54	−0.04*	0.18*	−0.16*	−0.03*	−0.15*	0.02*	−0.04*	1											
9 *Doctoral Degree*	0.58		0.17*	−0.65*	0.64*	0.06*	0.61*	0.04*	−0.002	−0.27*	1										
10 *Mobility*	0.38		0.04*	0.002	−0.01	0.06*	−0.06*	0.06*	0.03*	−0.001	−0.002	1									
11 *Recomb. Breadth Firm*	0.83	0.17	0.14*	−0.18*	0.24*	0.04*	0.22*	0.01	−0.16*	−0.05*	0.23*	−0.07*	1								
12 *Recomb. Breadth Coinventors*	0.35	0.26	0.30*	−0.33*	0.34*	−0.02*	0.35*	0.01	−0.05*	−0.09*	0.29*	0.01	0.21*	1							
13 *Number of References per Patent*	4.53	2.79	0.21*	0.08*	−0.07*	0.02*	−0.08*	0.002	−0.01	0.04*	−0.07*	−0.004	−0.01	0.02*	1						
14 *Class. Authority EPO (share)*	0.87		−0.04*	0.09*	−0.09*	0.01	−0.09*	0.001	0.0001	0.04*	−0.07*	−0.005	−0.03*	−0.05*	0.004	1					
15 *Class. Authority JPO (share)*	0.09		0.04*	−0.08*	0.08*	−0.004	0.08*	0.001	0.0001	−0.03*	0.07*	−0.003	0.03*	0.059*	−0.003	−0.81*	1				
16 *Class. Authority Other (share)*	0.04		−0.003	−0.03*	0.03*	−0.01	0.04*	−0.003	−0.0003	−0.03*	0.02*	0.01	0.02*	0.01*	−0.003	−0.53*	−0.06*	1			
17 *Number of Claims*	10.71	7.36	0.06*	0.06*	−0.07*	0.06*	−0.11*	0.02*	0.04*	0.06*	−0.06*	0.07*	−0.07*	0.003	0.17*	0.03*	−0.02*	−0.01	1		
18 *Frequ. Technology Combination*	14.25	13.02	−0.22*	0.20*	−0.22*	0.07*	−0.25*	−0.03*	0.06*	0.12*	−0.20*	−0.05*	−0.10*	−0.25*	0.02*	0.07*	−0.05*	−0.04*	0.13*	1	
19 *Firm Size*	60,715	89,161	−0.004	−0.05*	0.08*	0.14*	0.01	0.01	−0.08*	−0.06*	0.10*	−0.04*	0.28*	0.05*	−0.05*	−0.02*	0.02*	0.01	−0.08*	0.02*	1

Note. Pearson correlation coefficients (for two continuous variables), point biserial coefficients (for one continuous variable and one dummy variable), and phi coefficients (for two dummy variables) ($N = 30{,}550$) are shown.
*Significant at the 1% level.

Figure 1 **Histogram: Technological Recombination Breadth** ($N = 30{,}550$)



collinearity of covariates should not be a concern. We also computed variance inflation measures for all variables, but found only relatively small variance inflation factor values. Figure 1 provides a histogram of our dependent variable, showing that the technological breadth measure displays considerable variation. We can also see that 27.4% of the inventors are highly specialized, as their technological breadth measure is at its minimum (zero). Engineers account for 43.6% of the overall number of inventions, and scientists for 52.8%.[8]

4.2. Multivariate Results

Results of the random effects Tobit regression models predicting an inventor's recombination breadth are reported in Table 3. The results in column (1) employ the control variables only; the models in columns (2) and (3) add our main independent variables. Finally, columns (4) and (5) include interactions that examine potential heterogeneity between educational fields. The Tobit model explicitly takes the limited nature of the dependent variable into account. The estimated coefficients reflect the marginal effect of the independent variables on the (unobserved) uncensored variable. Our inference and model testing is performed on the basis of these coefficients; the marginal effects with respect to the observed outcomes will be discussed later.[9]

We first consider the results of the baseline model in column (2). In support of our first hypothesis, we find that inventors with a degree in science generate inventions with significantly greater technological breadth than inventors with an engineering degree (reference group). The dummy coefficient is estimated very precisely ($p < 0.01$). Inventors with other disciplinary backgrounds do not display any significant differences in recombination breadth when compared to the reference group. As predicted by our second hypothesis, inventors with a doctoral degree also display larger recombination breadth than inventors without doctoral training—the effect is approximately one-third of the impact of the science dummy variable and significant at the 5% level. Finally, this regression also yields evidence in favor of our third hypothesis: with time since completion of the terminal degree passing, recombination breath decreases. The respective coefficient is again significant at the 5% level.

The extended model in column (3) splits the science indicator into three subfield dummies (physics, chemistry, and other science) while maintaining engineering background as the reference group.[10] The coefficients of all scientific fields exhibit positive signs and are significantly different from zero. The size[11] of the coefficient for "chemistry" is approximately 50% larger than the effects of the other fields of science. However, the differences between the coefficients for different science fields are not significant, and the overall model fit does not improve significantly in comparison to model 2 ($\chi^2(2) = 2.26$, $p = 0.324$).

In column (4), we extend the main effects model in column (2) by including interaction effects between the science dummy and doctoral degree as well as between the science dummy and years since terminal degree. In a Wald test, the respective coefficients turn out to be insignificant. ($\chi^2(4) = 2.79$, $p = 0.593$). In column (5) we extend the model in column (3) by including a full set of interactions between the science subfield dummies and doctoral degree, and between science subfields and years since terminal degree. While some interaction terms become marginally significant, the overall explanatory contribution of the interaction effects remains very small and statistically insignificant (likelihood ratio test: $\chi^2(8) = 7.63$, $p = 0.470$). Hence, we maintain the relatively simple model in column (2) as the preferred specification and turn now to a discussion of the marginal effects. Our Tobit estimates do not reveal the marginal effects of our key variables on the observed outcome, because the underlying specification is nonlinear in nature. Figure 2 contains the average predictions of recombination breadth (conditional on censoring) as

[8] There is considerable variation in educational backgrounds over the technological areas of the focal patents. In some fields (telecommunication, handling/printing, machine tools, mechanical elements, and transportation) the share of engineers among inventors is above 90%. Inventors with science training are the dominant group of inventors in information technology, semiconductors, the chemistry-related areas, surface technology, materials, and space technology/weapons.

[9] The relative effect sizes can be readily interpreted in the Tobit model, because the marginal effects on the expected value of the uncensored outcome are proportional to the marginal effects of the expected value of the censored variable. The scaling factor is the probability of observing an uncensored outcome.

[10] Our data do not allow us to distinguish between different types of engineering training.

[11] We argue below that this is a decent approximation of the effect size, because censoring is not particularly pronounced in our data.

Table 3 Random Effects Tobit Models of Technological Recombination Breadth

Variables	(1) Coeff. [SE]	(2) Coeff. [SE]	(3) Coeff. [SE]	(4) Coeff. [SE]	(5) Coeff. [SE]
Field of Education (Reference Group: Engineering)					
Science	#	0.056*** [0.010]	#	0.075*** [0.019]	#
Physics	#	#	0.046*** [0.015]	#	0.047 [0.030]
Chemistry	#	#	0.068*** [0.013]	#	0.085*** [0.026]
Other Fields of Science	#	#	0.045** [0.020]	#	0.111** [0.046]
Other Fields	#	−0.028 [0.021]	−0.026 [0.021]	−0.007 [0.041]	−0.006 [0.041]
Doctoral Degree (*dummy*)	#	0.020** [0.010]	0.018* [0.010]	0.022* [0.013]	0.023* [0.013]
Science × *Doctoral Degree* (*dummy*)	#	#	#	−0.005 [0.019]	#
Physics × *Doctoral Degree* (*dummy*)	#	#	#	#	−0.028 [0.030]
Chemistry × *Doctoral Degree* (*dummy*)	#	#	#	#	−0.003 [0.025]
Other Science × *Doctoral Degree* (*dummy*)	#	#	#	#	−0.023 [0.041]
Other Fields × *Doctoral Degree* (*dummy*)	#	#	#	−0.015 [0.041]	−0.013 [0.041]
χ^2 test				$\chi^2(2) = 0.18$ $p = 0.914$	$\chi^2(4) = 1.12$ $p = 0.891$
Years Since Terminal Degree (*10 years*)		−0.009** [0.004]	−0.009** [0.004]	−0.005 [0.004]	−0.004 [0.004]
Science × *Years Since Terminal Degree* (*10 years*)	#	#	#	−0.010* [0.006]	#
Physics × *Years Since Terminal Degree* (*10 years*)	#	#	#	#	0.011 [0.013]
Chemistry × *Years Since Terminal Degree* (*10 years*)	#	#	#	#	−0.011* [0.007]
Other Science × *Years Since Terminal Degree* (*10 years*)	#	#	#	#	−0.030* [0.017]
Other Fields × *Years Since Terminal Degree* (*10 years*)	#	#	#	−0.009 [0.018]	−0.009 [0.018]
χ^2 test				$\chi^2(2) = 2.83$ $p = 0.243$	$\chi^2(4) = 6.86$ $p = 0.144$
Mobility (*dummy*)	0.014* [0.007]	0.012* [0.007]	0.012* [0.007]	0.012* [0.007]	0.012* [0.007]
Recombination Breadth Employer	0.114*** [0.017]	0.108*** [0.017]	0.108*** [0.017]	0.108*** [0.017]	0.107*** [0.017]
Average Recombination Breadth Coinventors	0.179*** [0.011]	0.174*** [0.011]	0.173*** [0.011]	0.173*** [0.011]	0.172*** [0.011]
Number of References per Patent	0.023*** [0.001]	0.023*** [0.001]	0.023*** [0.001]	0.023*** [0.001]	0.023*** [0.001]
Classification Authority EPO (*dummy*)	0.021 [0.013]	0.021 [0.013]	0.021 [0.013]	0.021 [0.013]	0.021 [0.013]
Classification Authority JPO (*dummy*)	0.039** [0.016]	0.037** [0.016]	0.037** [0.016]	0.037** [0.016]	0.037** [0.016]
Number of Claims	0.001* [0.000]	0.001* [0.000]	0.001* [0.000]	0.001* [0.000]	0.001* [0.000]
Frequency of Technology Combination (*1,000*)	−0.006*** [0.000]	−0.006*** [0.000]	−0.006*** [0.000]	−0.006*** [0.000]	−0.006*** [0.000]

Table 3 (Continued)

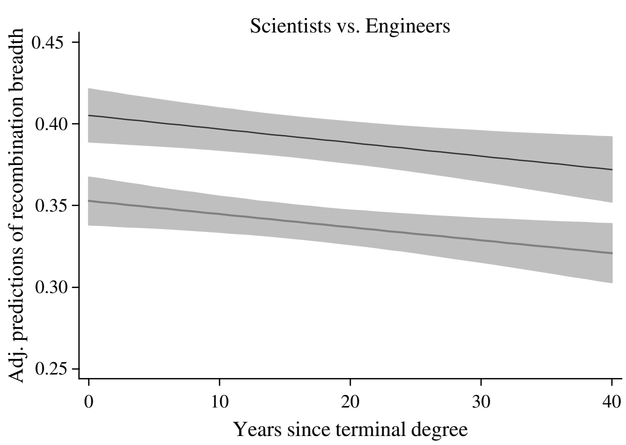
Variables	(1) Coeff. [SE]	(2) Coeff. [SE]	(3) Coeff. [SE]	(4) Coeff. [SE]	(5) Coeff. [SE]
ln(*Firm Size*)	−0.004**	−0.006***	−0.006***	−0.006***	−0.006***
	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
30 *Technological Areas* (dummies)	Included	Included	Included	Included	Included
Year Dummies	Included	Included	Included	Included	Included
Constant	0.021	0.035	0.035	0.022	0.020
	[0.043]	[0.043]	[0.043]	[0.044]	[0.044]
Rho	0.210	0.199	0.199	0.200	0.199
Observations	13,110	13,110	13,110	13,110	13,110
Number of inventors	1,880	1,880	1,880	1,880	1,880
Wald χ^2	2,336.97	2,457.96	2,463.20	2,460.38	2,473.75
p-value	0.000	0.000	0.000	0.000	0.000
Log-likelihood	−2,529.76	−2,496.87	−2,495.74	−2,495.42	−2,491.86

Note. Standard errors are in brackets.
$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

a function of time since terminal degree and our most important group variables. As is evident from this graph, the marginal effect of a science background is quite large and amounts to roughly one quarter of a standard deviation of the dependent variable. It declines as time passes after the inventor has graduated, but the decline is very gradual and there is no evidence in the data of any convergence.

It is important to note that these systematic differences between inventors are evident even after controlling for several key team-level and organizational-level characteristics. Specifically, we point to the significance of two main control variables, that is, the breadth of an inventor's group of collaborators and the breadth of the employer organization. As one can expect, both have a positive relationship with the inventor's technological recombination breadth.

Figure 2 Average Predicted Values of Technological Recombination Breadth



Notes. Results are based on Table 3, column (2). All covariates are fixed at sample means for the two groups. Top line, inventors with science degrees; bottom line, inventors with engineering degrees (shading: 95% confidence intervals).

The remaining controls also display the expected signs.[12]

5. Discussion

In this section we conclude by discussing implications and limitations of our research. We have employed a unique and novel data set based on matched patent register and survey data to study the relationship between inventive behavior and educational background. Our results also indicate that inventors with a science background and inventors with higher educational attainment (in this case a doctoral degree) are more likely to combine knowledge across technological boundaries. Moreover, our study delivers first results regarding the breadth of technological recombination over time—we find clear evidence that both for engineer and scientist inventors, specialization becomes stronger over time, albeit at a relatively slow rate. By identifying and documenting these systematic microlevel differences in inventors' abilities to recombine technological knowledge, our results offer several new insights for research on inventors, on organizational learning and innovation, and for strategy research. From a methodological perspective, we advance innovation research by introducing a novel index that has superior properties for capturing the breadth of an inventor's technological recombination activities relative to existing measures such as the traditional Herfindahl index.

[12] We have undertaken numerous robustness tests of these results. Exploring alternative functional forms (especially for the time variable, e.g., in logarithmic or hyperbolic transformations) shows that the simple linear model is preferred. We also estimated the model via Tobit on the full 30,550 observations at the inventor-patent level, using standard errors clustered at the inventor level. The coefficient estimates are almost identical to those reported here. We conclude that the results from the estimations presented here are remarkably stable and robust. The results are available upon request.

5.1. Implications

5.1.1. Implications for Research on Inventors. To date, Allen's (1977) observations have neither been the subject of a large-scale empirical test nor been considered in contemporary theorizing about technological knowledge recombination processes and outcomes. Our results thus provide an important empirical validation of these hypotheses using data that span the inventive histories of a large set of inventors. Moreover, we find that the differences between scientists and engineers are so fundamental that they cannot even be compensated by obtaining a relatively higher educational degree in engineering. Over time, specialization increases for all inventor groups, but there is no evidence in favor of convergence. Given that these differences between scientists and engineers appear to be highly robust, future studies examining the recombinant activity of inventors may profit from distinguishing between these two groups.

5.1.2. Implications for Research on Organizational Learning and Innovation. The literature on learning and innovation has a core interest in understanding how firms may tap into new technological domains to achieve path-breaking innovation. Whereas most research in this vein has been conducted at the firm level (Almeida and Kogut 1999, Wadhwa and Kotha 2006, Miller et al. 2007), we contribute to this line of inquiry by shedding light on the role of individual-level factors in technological recombination, and thus can provide a more complete picture of the factors that may constrain or facilitate organizational learning (Hargadon and Sutton 1997, Hargadon 2006). Our findings contribute to a body of work examining the tension between the breadth and depth of research activities (Kuhn 1962, Leahey 2007, Leahey and Reikowsky 2008, Jones 2009). Although both narrow and broad technological recombination can lead to innovation, these two types of recombination entail fairly different challenges and require different mindsets, abilities, and routines from inventors (Rosenkopf and Nerkar 2001, Gupta et al. 2006). Our results suggest that scientific education enhances the likelihood of knowledge recombination across technological boundaries. Hence, although inventors face an increasing stock of knowledge that is generated with technological progress (Jones 2009), a thorough understanding of science may help them in identifying and assimilating knowledge elements in more distant technology domains. Thus, to stand "on the shoulders of giants," they do not necessarily have to make the investment to "climb up their backs" (Jones 2009, p. 284), yet may take an important "shortcut" by understanding how the advanced domain-specific knowledge relates to basic scientific insights.

5.1.3. Implications for Research on Strategic Management. The ability to engage in distant technological search is of particular importance for sustaining the firm's competitiveness (March 1991). By examining the microlevel of technological recombination, the findings of this study also offer new insights to strategy research investigating the capabilities required for recombination across technological boundaries. For instance, Rosenkopf and Nerkar (2001) suggested that firms have to establish "second-order competence," that is, the ability to create new knowledge across technological boundaries, and Kogut and Zander (1992) argued that firms need to build "combinative capability" to be able to synthesize existing and newly acquired knowledge. By showing which types of inventors tend to produce boundary-spanning inventions, the results of the present study provide insights that may facilitate the development of such competences (capabilities) in organizations.

5.2. Limitations

In interpreting the results of this study, certain limitations must be kept in mind. First, as many other studies in this field, we cannot employ experimental (or at least quasi-experimental) data. Various selection effects may be at work. In particular, employers may very well be aware of the differences between scientists and engineers, of the impact of doctoral training, and of the "mellowing" effect as inventors become older. In this case, our research simply reveals a difference between the two types of inventors that has gone unnoticed in the literature, but we cannot claim that the results can be interpreted in a causal manner. Second, our study is partly based on patent data, because the archival nature of this type of data allows us to trace the inventive history of individuals. This type of data also allows us to trace an individual's technological recombination activities as he or she moves across organizations (Fleming et al. 2007). However, we also note that patent data provide an incomplete coverage of innovative activity, because not all outcomes of R&D processes are patented or patentable (Cohen et al. 2000). Third, as pointed out before, employing technology classifications brings along certain hazards. We employ what we deem to be the best measure available. The technological classification of the EPO provides a better measure of technological breadth than the technological classification of the U.S. patent system, because the IPC classes listed on EP patents reflect the technology described in the complete patent document, whereas the U.S. classification is based on information given in the claims only. But technological recombination may be easier with some patent classes than with others, as different pairs of IPC classes are usually not separated by the same technological distance; for instance, IPC classes in the fields of chemistry and biology are closer than those in chemistry and mechanics. We have used various approaches to limit distortions, but we cannot fully exclude the possibility of measurement errors.[13]

5.3. Conclusion

We have built on some classical research results from the 1970s to motivate our main research question: what is the impact of different types of inventors—engineers versus scientists, doctorates versus nondoctorates—on the recombination of knowledge in the course of the invention process? Together with the studies discussed here, our findings form a body of research that shifts the attention toward the role of the microlevel in explaining innovation performance and suggests that the conventional predictors of innovation (e.g., the firm's R&D expenditures) need to be complemented by careful examination into how differences on the microlevel may lead to performance heterogeneity in R&D (Laursen and Salter 2006). In future studies, researchers may want to develop a better understanding of how the commercialization of patents that recombine knowledge across technological boundaries differs from the commercialization of narrower patents. For instance, research on technology entrepreneurship and patenting behavior indicates that patent characteristics are an important determinant of start-up creation (Shane 2001). Ultimately, a better understanding of the kind of searches underlying inventive activity may also support the development of improved patent valuation techniques (Gittelman 2008).

Acknowledgments

The authors thank the European Commission [Contract N. HPV2-CT-2001-00013] for supporting the creation of the data set. They appreciate feedback on drafts from Joern Block, Georg von Graevenitz, Joachim Henkel, Fabio Montobbio, Marie Thursby, Jerry Thursby, Christopher Tucci, Anu Wadhwa, and John Walsh, as well as seminar participants at the 2008 EPIP Conference, INNO-tec (LMU Munich), the College of Management at the Georgia Institute of Technology (Atlanta, Georgia), the 2009 Academy of Management Conference, the 2010 DRUID Conference, the 2010 Monte Verita Conference, the Social Science and Technology Seminar at Stanford University, the 2009 TIM Conference (Aachen, Germany), and the University of Lausanne. The comments of the editors and three anonymous reviewers are gratefully acknowledged. The authors contributed equally, and their names are listed alphabetically.

[13] For example, we decided for a conservative measure of technological breadth at the level of 30 broad technological fields, and against finer-grained levels such as the seven-digit level of the IPC. Because our results may be affected by a strong effect of chemicals and pharmaceuticals, as IPC classes and subclasses are most differentiated in these technological fields, we also control for the technological field(s) in which the inventors are active and for the frequency at which combinations of classes occur in the patent system.

References

Agrawal AK, Cockburn IM, McHale J (2006) Gone but not forgotten: Labor flows, knowledge spillovers, and enduring social capital. *J. Econom. Geography* 6:571–591.

Allen TJ (1977) *Managing the Flow of Technology* (MIT Press, Cambridge, MA).

Almeida P, Kogut B (1999) Localization of knowledge and the mobility of engineers in regional networks. *Management Sci.* 45(7):905–917.

Birnbaum PH (1981) Integration and specialization in academic research. *Acad. Management J.* 24(3):487–503.

Blade E (1963) Creative science. Coler MA, ed. *Essays of Creativity in the Sciences* (New York University Press, New York), 183–206.

Bosworth D (1978) The rate of obsolescence of technical knowledge—A note. *J. Indust. Econom.* 26(3):273–279.

Capron L, Dussauge P, Mitchell W (1998) Resource redeployment following horizontal acquisitions in Europe and North America, 1988–1992. *Strategic Management J.* 19(7):631–661.

Cohen WM, Levinthal DA (1990) Absorptive capacity: A new perspective on learning and innovation. *Admin. Sci. Quart.* 35(1):128–52.

Cohen WM, Nelson RR, Walsh JP (2000) Protecting their intellectual assets: Appropriability conditions and why U.S. manufacturing firms patent (or not). NBER Working Paper 7552, National Bureau of Economic Research, Cambridge, MA.

David P, Mowery D, Steinmueller E (1992) Analyzing the economic payoffs from basic research. *Econom. Innovation New Technologies* 2(1):73–90.

Fleming L (2001) Recombinant uncertainty in technological search. *Management Sci.* 47(1):117–132.

Fleming L, Sorenson O (2004) Science as a map in technological search. *Strategic Management J.* 25(8–9):909–928.

Fleming L, Mingo S, Chen D (2007) Collaborative brokerage, generative creativity, and creative success. *Admin. Sci. Quart.* 52(3):443–475.

Gagné RM, Glaser R (1987) Foundations in learning research. Gagné RM, ed. *Instructional Technology: Foundations* (Lawrence Erlbaum, Hillsdale, NJ), 49–83.

Gambardella A (1995) *Science and Innovation: The US Pharmaceutical Industry During the* 1980*s* (Cambridge University Press, Cambridge, UK).

Gibbons M, Johnston R (1974) The roles of science in technological innovation. *Res. Policy* 3(3):220–242.

Gittelman M (2008) The value of European patents (comment). *Eur. Management Rev.* 5(2):85–89.

Griliches Z (1957) Hybrid corn: An exploration in the economics of technological change. *Econometrica* 25(4):501–522.

Gupta AK, Smith KG, Shalley CE (2006) The interplay between exploration and exploitation. *Acad. Management J.* 49(4): 693–706.

Halford G, Wilson W, Guo J, et al. (1994) Connectionist implications for processing capacity limitations in analogies. Gayler RW, Wiles J, Stewart JEM, eds. *Advances in Connectionist and Neural Computation Theory*, Vol. 2. (Ablex, Norwood, NJ), 363–415.

Hambrick DC, Mason P (1984) Upper echelons: The organization as a reflection of its top managers. *Acad. Management Rev.* 9(2):193–206.

Hargadon AB (1998) Firms as knowledge brokers: Lessons in pursuing continuous innovation. *California Management Rev.* 40(3): 209–227.

Hargadon AB (2006) Bridging and building: Towards a microsociology of creativity. Thompson L, ed. *Creativity and Innovation in Groups and Teams* (Lawrence Erlbaum, Mahwah, NJ), 199–216.

Hargadon AB, Sutton RI (1997) Technology brokering and innovation in a product development firm. *Admin. Sci. Quart.* 42(4):716–749.

Harhoff D, Narin F, Scherer FM, Vopel K (1999) Citation frequency and the value of patented inventions. *Rev. Econom. Statist.* 81(3):511–515.

Helfat CE (1994) Evolutionary trajectories in petroleum firm R&D. *Management Sci.* 40(12):1720–1747.

Hoisl K (2007) Tracing mobile inventors: The causality between inventor mobility and inventor productivity. *Res. Policy* 36(5):619–636.

Jones BF (2005) Age and great invention. NBER Working Paper 11359, National Bureau of Economic Research, Cambridge, MA.

Jones BF (2009) The burden of knowledge and the "death of the renaissance man": Is innovation getting harder? *Rev. Econom. Stud.* 76(1):283–317.

Klevorick AK, Levin RC, Nelson RR, Winter SG (1995) On the sources and significance of interindustry differences in technological opportunities. *Res. Policy* 24(2):185–205.

Kogut B, Zander U (1992) Knowledge of the firm, combinative capabilities, and the replication of technology. *Organ. Sci.* 3(3):383–397.

Kotabe M (1992) A comparative study of U.S. and Japanese patent systems. *J. Internat. Bus. Stud.* 23(1):147–168.

Kuhn T (1962) *The Structure of Scientific Revolutions* (University of Chicago Press, Chicago).

Laursen K, Salter A (2006) Open for innovation: The role of openness in explaining innovation performance among U.K. manufacturing firms. *Strategic Management J.* 27(2):131–150.

Leahey E (2006) Gender differences in productivity: Research specialization as a missing link. *Gender Soc.* 20(6):754–780.

Leahey E (2007) Not by productivity alone: How visibility and specialization contribute to academic earnings. *Amer. Sociol. Rev.* 72(4):533–561.

Leahey E, Reikowsky RC (2008) Research specialization and collaboration patterns in sociology. *Social Stud. Sci.* 38(3):425–440.

Ledoux SF (2002) Defining natural sciences. *Behaviorology Today* 5(1):34–36.

Lerner J (1994) The importance of patent scope: An empirical analysis. *RAND J. Econom.* 25(2):319–333.

Levin SG, Stephan PE (1991) Research productivity over the life cycle: Evidence for academic scientists. *Amer. Econom. Rev.* 81(1):114–132.

Lotka AJ (1926) The frequency distribution of scientific productivity. *J. Washington Acad. Sci.* 16:317–323.

March JG (1991) Exploration and exploitation in organizational learning. *Organ. Sci.* 2(1):71–87.

March JG, Simon H (1958) *Organizations* (Wiley, New York).

Marx M, Strumsky D, Fleming L (2009) Mobility, skills, and the Michigan non-compete experiment. *Management Sci.* 55(6):875–889.

Michel J, Bettels B (2001) Patent citation analysis—A closer look at the basic input data from patent search reports. *Scientometrics* 51(1):185–201.

Miller DJ, Fern MJ, Cardinal LB (2007) The use of knowledge for technological innovation within the diversified firm. *Acad. Management J.* 50(2):308–326.

Narin F, Breitzman A (1995) Inventive productivity. *Res. Policy* 24(4):507–519.

Nelson RR (1959) The simple economics of basic scientific research. *J. Political Econom.* 67(3):297–306.

Nelson RR, Winter SG (1982) *An Evolutionary Theory of Economic Change* (Harvard University Press, Cambridge, MA).

Nersessian NJ (2002) The cognitive basis of model-based reasoning in science. Carruthers P, Stich SP, Siegal M, eds. *The Cognitive Basis of Science* (Cambridge University Press, Cambridge, UK), 133–153.

Nonaka I (1994) A dynamic theory of organizational knowledge creation. *Organ. Sci.* 5(1):14–37.

OECD (Organisation for Economic Co-operation and Development) (1994) *Using Patent Data as Science and Technological Indicators—Patent Manual* 1994 (OECD, Paris).

Pakes A, Schankerman M (1984) The rate of obsolescence of patents, research gestation lags, and the private rate of return to research resources. Griliches Z, ed. *R&D, Patents, and Productivity* (University of Chicago Press, Chicago), 73–88.

Park G, Shin J, Park Y (2006) Measurement of depreciation of technological knowledge: Technology life cycle time approach. *J. Sci. Indust. Res.* 65(2):121–127.

Pavitt K, Robson M, Townsend J (1989) Technological accumulation, diversification and organisation in UK companies, 1945–1983. *Management Sci.* 35(1):81–99.

Pelled LH (1996) Demographic diversity, conflict, and work group outcomes: An intervening process theory. *Organ. Sci.* 7(6):615–631.

Rosenkopf L, Almeida P (2003) Overcoming local search through alliances and mobility. *Management Sci.* 49(6):751–766.

Rosenkopf L, Nerkar A (2001) Beyond local search: Boundary-spanning, exploration, and impact in the optical disk industry. *Strategic Management J.* 22(4):287–306.

Schott K (1978) The rate of obsolescence of technical knowledge–A comment. *J. Indust. Econom.* 26(3):281–283.

Schumpeter J (1934) *Capitalism, Socialism, and Democracy* (Harper & Row, New York).

Shane S (2001) Technological opportunities and new firm creation. *Management Sci.* 47(2):205–220.

Simonton DK (1999) *Origins of Genius: Darwinian Perspectives on Creativity* (Oxford University Press, New York).

Stuart TE, Podolny JM (1996) Local search and the evolution of technological capabilities. *Strategic Management J.* 17(S1):21–38.

Trajtenberg M, Henderson R, Jaffe A (1997) University versus corporate patents: A window on the basicness of invention. *Econom. Innovation New Tech.* 5(1):19–50.

Trajtenberg M, Shiff G, Melamed R (2006) The "names game": Harnessing inventors' patent data for economic research. NBER Working Paper 12479, National Bureau of Economic Research, Cambridge, MA.

von Tunzelmann GN (1998) Localized technological search and multi-technology companies. *Econom. Innovation New Tech.* 6(2–3):231–256.

Wadhwa A, Kotha SB (2006) Knowledge creation through external venturing: Evidence from the telecommunications equipment manufacturing industry. *Acad. Management J.* 49(4):819–835.

Walsh JP (1995) Managerial and organizational cognition: Notes from a trip down memory lane. *Organ. Sci.* 6(3):280–321.

Ward TB, Smith SM, Vaid Y (1997) Conceptual structures and processes in creative thought. Ward TB, Smith SM, Vaid Y, eds. *Creative Thought—An Investigation of Conceptual Structures and Processes* (American Psychological Association, Washington, DC), 1–27.

Woolnough BE (1991) *The Making of Engineers and Scientists* (Oxford University, Oxford, UK).

Woolnough BE (1994) Factors affecting students' choice of science and engineering. *Internat. J. Sci. Ed.* 16(6):659–676.

Zucker LG, Darby MR (1996) Star scientists and institutional transformation: Patterns of invention and innovation in the formation of the biotechnology industry. *Proc. Natl. Acad. Sci. USA* 93(23):12709–12716.