Paper to be presented at the

DRUID Society Conference 2014, CBS, Copenhagen, June 16-18

How Search in Science Impacts on the Value of Inventions at Early and Late Stages in the R&D cycle

Karin Beukel
University of Copenhagen
Department for Food and Resource Economics
kab@ifro.ku.dk

Finn Valentin
Coepnhagen Business School
Department of Innovation and Organizational Economics
fv.ino@cbs.dk

Rasmus Lund Jensen
CBS
INO
rlj.ino@cbs.dk

Abstract

The literatures on innovation and organizational learning have identified search in prior science and technology as critical inputs to industrial R&D. Efforts to distinguish the contributions - separately and combined - of these two search orientations are scarce, and quantitative estimates offer contradictory results. The contributions to R&D from science are particularly elusive. To achieve some transparency on these issues we study R&D in biotech drug discovery, where the role of science is pervasive and structured into a recurrent sequence of inventions required to build a drug candidate. The patents filed on these inventions offer, through their citations to prior art, a fine-grained view of the role of science along the R&D cycle. Applying a unique text-mining algorithm we categorize a set of 1,058 patens from Scandinavian drug discovery firms into six types of drug-related inventions. Tests confirm a significantly decreasing presence of science vis a vis technology over the R&D cycle. Effects of the composition of search on the value of single inventions show notable differences. In early R&D increasing predominance of search in science detracts from invention value. In late R&D, inventions increase in value when search includes a scientific orientation. Science-based R&D is cognitively heterogeneous and builds value in forms requiring sophisticated R&D management. Our results add to the theoretical understanding of search and the role of science in innovations. They also explain why an aggregate view, as presented

in prior literature, on the value produced by science in R&D leads to contradictory or insignificant findings.

Jelcodes:O32,O31

1. Introduction

Differences in the sources of the knowledge exploited in Research and Development (R&D) influence firm productivity and economic growth (Griliches, 1984, 1986; Mansfield, 1972; Rosenberg, 1974). The organizational learning literature shows that both explorative and exploitative search is needed to enable good firm performance (Benner and Tushman, 2003; Brown and Duguid, 2001; Gupta et al., 2006; Katila and Ahuja, 2002; March, 1991; Tushman and Oreilly, 1996), and that both science and technology are potentially valuable sources of knowledge for the development of innovations (Brooks, 1994; Jensen et al., 2007; Rosenberg, 1990; Tijssen, 2002). However, search in science and search in technology are fundamentally different, and respond to different problem solving processes during innovation. Search in science is aimed at understanding phenomena while search in technologies is directed to exploiting existing solutions in further innovations (Brooks, 1994; Rosenberg and Nelson, 1994). Despite these essential differences between science and technology, the literatures on innovation and on organizational learning provide strong empirical support for their combined role in the process of innovation. However, little is known about the patterns of these combinations and their effects on innovation value.

Empirical studies of science-based firms differ regarding the role they attribute to science in the context of other sources of knowledge applied in R&D (Bruno Cassiman and Zuniga, 2007; Fleming and Sorenson, 2004). For individual innovations, the orientation of search - towards science vs. technology - may shift across the R&D cycle (Iansiti, 1997; Kline and Rosenberg, 1986). Often the final product is the result of multiple cumulative inventions, whose origins in science and technology search differ. Rosenberg and Nelson (1994) suggested that early in the R&D cycle inventions rely on science search while later innovation stages rely on search in technology. This pattern presents a challenge for the management of science-based innovations. Although different search patterns predominate in different

stages, a *balanced configuration of search* across the R&D cycle may be matter more for innovation performance. This paper investigates questions related to differences in the role of science search across the R&D cycle, and the effects on the value of innovations of the balance between science and technology search. To our knowledge, these issues have not been addressed in the existing literature.

Following a tradition in the organizational learning literature (Fleming and Sorenson, 2004; Yayavaram and Ahuja, 2008) we conceptualize search as an expression of the distinct knowledge sources exploited for problem solving. Drawing on work by Rosenberg and Nelson (1994) and Brooks (1994) on the fundamental difference between science and technology, we propose that search through the lens of science provides answers to questions such as *'why an effect appears'*, while search through the lens of technological application provides answers to *'how an effect appears'*. In science based industries, a core objective of R&D is to transform a scientific discovery providing a solution to a fundamental question, into a commercial invention (Stankiewicz, 2000). Search in science and technology are different approaches applied by firms to inform this innovation task.

Previous studies use backward citations in patents as footprints of the prior science and technology relating to the focal invention. We argue that these footprints also demonstrate how science and technology can be used as *search lenses* to generate inventions. Search based on science is expressed in backward citations in the focal patent, to the scientific literature (referred to as non-patent literature citations) (Cassiman et al., 2008; Fleming and Sorenson, 2004; Harhoff et al., 2003). Search based on technology is expressed in backward citations to prior patents (Harhoff et al., 2003; Lanjouw and Schankerman, 1997).

Although most innovation studies assume that search in science is predominant in early R&D while search in technology prevails later in the R&D cycle, this fundamental difference has not been demonstrated empirically in a large N-sample. Identifying individual inventions at early versus late stages of R&D is problematic in large scale empirical studies. To the best of our knowledge, the present study is

the first to suggest a method to resolve this issue. To identify individual inventions from a string of cumulative inventions requires in-depth knowledge of the types of inventions created during R&D and a method for identifying them. We focus on the biotech industry where types of inventions emerging from R&D are identified in the literature along with their most prevalent sequences. By developing a unique algorithm based on semantic structures and International Patent Codes (IPCs), we can identify different invention types from a large number of patent observations. On this basis we can test our hypotheses in an empirical study utilizing unusually detailed data on 1,058 biotech patents.

Our results support our expectations; search in science gradually becomes less predominant across the R&D cycle and search in technologies become significantly more prevalent during late R&D. Second, we show that invention value is correlated with search in science, however, the sign differs depending on the stage of R&D. With separate tests of invention created early vs. late in R&D we find that in early R&D search in technologies is infrequent in comparison to late R&D processes, but one-sided search in science has a penalizing effect on invention value. At the opposite end, in late stage R&D, search in science is less frequent but nevertheless is associated with increased invention value. We therefore highlight the importance of considering individual stages in the R&D cycle when examining the value and disadvantages associated with search in science.

Our findings add several insights to the management of innovation and organizational learning literatures. First, we empirically demonstrate how, in an intensive science-based industry, search in science varies from early to late stages in the R&D cycle. Second we demonstrate how the value generated for innovation from search in science depends on the way it is balanced against technology search, and that the effects of this balance change across R&D stages. Third, on this basis we resolve the contradictions in the previous literature. Studies have found that the value that research-based firms extract from inflows of strong science vary (Gittelman and Kogut, 2003). But pursed at the level of single

inventions the relationship is either negative or absent (Cassiman et al., 2008). Our findings related to changes in the effects of inputs across the R&D cycle mostly resolve this paradox. Fourth, our findings are based on a novel methodology for advanced profiling of large numbers of patented inventions.

The remainder of the paper is organized as follows: In Section 2, we discuss the mechanisms of search in both early and late R&D processes. We develop hypotheses related to the declining role of search in science across the R&D cycle, associated with increases in technology search. We develop hypotheses related to the effects of innovation value. In Section 3 we present the data and methods. In Section 4 the empirical findings. In the final Section 5 we discuss the results and their implications, limitations, and extensions of the research.

2. Theoretical background and hypotheses

2.1 Changes in search across the R&D cycle

Since its origins, the innovation literature has recognized both science and technology as sources of novelty, and examined how innovations emerge from their interaction and iteration (Gibbons and Johnston, 1974; Kline and Rosenberg, 1986; Schumpeter, 1939). A few sectors including biotech and life sciences, carry out systematic in-house research as the basis for their innovations (Cohen, 2002; Klevorick, 1995; Nelson, 2003; Stankiewicz, 2000), and in these sectors the relationship between scientific and technological advances allows firms to compete on the basis of their ability to extract innovations directly from research (Pavitt, 1991). In these science-based industries, inventions originate from conversations with nature about "why" questions. In this context we expect early stage R&D to be strongly oriented to science. In complex technologies (Cohen et al., 2000) innovations emerge from the accumulation of separately patentable elements. As R&D work progresses, problem-solving shifts to issues that are increasingly technological in nature (Iansiti, 1997; Kline and Rosenberg, 1986). Case studies of the R&D cycle in science-based firms demonstrate this shift from early science-based discovery

toward subsequent stages of development and an increased emphasis on technological knowledge (Cassiman, 2010; Iansiti and West, 1999). There is a change of scientific "why" questions to technological questions about "how" to make things work, e.g. secure stable functionality and compatibility with other technologies.

A powerful conceptualization of these differences in cognitive direction of search and problem solving in R&D has emerged from the notion of *cognitive maps*. In the psychological literature, cognitive maps refer to the constructs allowing people to perceive and interpret the current situation in relation to its environment, in their search for suggestions of which orientation to follow (Carroll, 1993; Neisser, 1976; Piaget, 1985). In the management literature, Gavetti and Levinthal (2000) build on this conceptualization of maps: They develop a distinction between two different cognitive problem solving approaches associated with innovations, in the *forward looking* approach, cognitive mapping is concerned with linkages between choice of action and subsequent impact. This type of problem solving centers on the development of hypotheses related to identification of potential causal relationships. Search in the *backward looking* approach is exploitative and uses existing knowledge. We suggest that these two approaches correspond to two different R&D-related search lenses. The forward looking approach is closely associated with science-based search for explanations of 'why an effect appears'. The backward looking approach corresponds to which existing technologies are exploited and define the context.

This correspondence between forward/backward looking cognitive maps and search lenses and science/technology is effectively demonstrated by (Yayavaram and Ahuja, 2008) in their study they exemplify this using the discovery of high-temperature superconductivity in copper-oxide based materials. This discovery led inventors to apply a forward looking cognitive approach to the search for a theoretical model offering a scientific explanation of 'why this effect appeared'. A different experiential

approach was applied to gain alternative and more exploitative insights. Applying a backward looking cognitive approach involves experiments with related materials to answer questions about 'how the effect appears in different settings'. The forward looking science-based approach is critical for understanding the underlying causal mechanisms of high-temperature superconductivity in copper-oxide based materials. A backward exploitative view of existing technological knowledge is required to stabilize and enable industrial applicability, but offers no overall guidance on the critical cause-effect relationships in this new area of technology. Therefore, drawing on Gavetti and Levinthal (2000) and Yayavaram and Ahuja (2008) we relate search in science and search in technology to two distinct different cognitive maps guiding problem solving processes in different directions.

With reference to science-based industries we conjecture that:

Hyp. 1: Inventions in early stage R&D are based on search in science more than are inventions in late stage R&D

Hyp. 2: Inventions in late stage R&D are based on search in technology more than are inventions in early stage R&D.

2.2 Invention value and shifts in the composition of search

Although the composition of search shifts across the R&D cycle the most valuable innovations may not arise from taking these shifts to extremes. Theory suggests that innovation performance is enhanced by cognitive variety (Brown and Eisenhardt, 1997). In the present context that implies R&D search based on a combination of the two cognitive maps with scientific and technological orientations. Management research demonstrates that more valuable problem solving emerges from combining different mental models (Brown and Eisenhardt, 1997; Smith and Tushman, 2005), partly because it enables flexible recombination of individual cognitive maps (Eisenhardt et al., 2010).

However, based on the theory underpinning Hypotheses 1 and 2 presented in Section 2.1, we expect cognitive variety to affect innovation performance based on *different configurations* of scientific and technological search at the opposite ends of the R&D cycle. In early stage R&D, the predominant pattern of search is based on the cognitive map of science. At the same time, the learning loops required by the chain-linked nature of R&D (Kline and Rosenberg, 1986) suggest that innovations will be more valuable if they anticipate technological conditions appearing in later R&D stages. Also, search in technologies provides a window on prior related learning in other firms, and combines the focal inventions with proven principles incorporated in previous technologies, which increases their value by reducing the underlying uncertainty (Nemet and Johnson, 2012) and economizing by reusing knowledge (Langlois, 1999).

To reap these benefits the cognitive map of early R&D must maintain a technological orientation, suggesting that a search pattern that relies predominantly on science-based search will detract from the value of early stage R&D. The logic of this argument suggests that this penalty will increase particularly if science-based search begins to dominate inventive search. Hence we propose:

Hyp 3: Early stage R&D decreases in value with increasing orientation towards science in the search underpinning inventions

The circumstances in the later stages of R&D are different. The input to these stages of R&D is not "raw" scientific opportunity but an invention, which has already progressed through the pipeline for several years. Drug development takes around 8-12 years (Achilladelis and Antonakis, 2001; Chandy et al., 2006; DiMasi, 2001; Mansfield, 1991, 1998). The aim in the later processes of R&D is to further modify and optimize the invention to ease its approval in from regulatory authorities process, to ensure no or small side effects, to prove the drug to be more efficient than the drugs already available on the

market, and to create new patent protection to extend the patent protected life of the drug. In this setting the experiential knowledge accumulated by firms is based on problem-solving using technology based cognitive maps (Pisano, 1996). The risks of falling into familiarity traps (Ahuja and Lampert, 2001) are particularly acute in this phase of the cycle, and finding an effective balance among different mental models by applying cognitive variety involves securing inflows of new information and knowledge (Atuahene-Gima, 2005; Leonard-Barton, 1992). Prior research shows that a science based cognitive map is a powerful way to avoid familiarity traps (Fleming, 2001). Therefore, we propose that in the late stages of R&D, where technology search predominates, the addition of elements from science based cognitive maps will enhance the value of inventions.

Hyp 4: Late stage R&D increases in value if an orientation towards science is part of the search underpinning inventions.

3. Data and Methods

3.1 Differences in patenting across the R&D cycle

We draw on research on innovation and focus on the level of the individual patent as the unit of analysis. Essentially we see each invention (expressed in one type of patent) as coming out of a research project which is distinct from the projects required to invent the other patent types protecting the same drug candidate. The distinctiveness of these projects partly grows out of the different types of expertise required for different patent types (e.g. formulation inventions draw heavily on pharmacological expertise, which is not the case for early inventions of e.g. compound and utility patents). Prior research has applied this strategy to map the firm's knowledge base (Katila and Ahuja, 2002; Yayavaram and Ahuja, 2008), and to characterize its science search (Fleming and Sorenson, 2004) and the effects on innovation value. We contribute to these methodological approaches by a) identifying the constituent inventions that ultimately combine to form a single marketable innovation (in our case a drug), and b) by

categorizing these constituent inventions into recurrently appearing types, related to either the early exploratory stages of R&D or the later exploitative stages. To these solutions we add methodologies proposed in the literature such as characterizing each invention in terms of its reliance on search in science, prior technologies, and value. The drug discovery segment of the biotech industry has a pattern of R&D, and paper trails that allow us to apply these methodological approaches. The final dataset consists of 1,058 patents applied by 110 Scandinavian drug discovery firms established between 1987 and 2003.

In the biotech industry, patents are critical for effective appropriability (Levin et al., 1987). In drug discovery, different types of patented inventions are created during distinct stages of discovery and development of drug candidates. We identify six patent types as relevant for biopharmaceutical drug inventions. They tend to appear in a distinct sequence in the R&D cycle (Gupta and Bansal, 2002; Kaushal and Garg, 2003; Norman, 2007; Sternitzke, 2010, 2013; Yoo et al., 2005)**:** 1) patents protecting the *method of identifying new products.* In the biotech industry, these are *platform patents* (Yoo et al., 2005) which cover inventions created through very early and explorative problem solving on industrialized high-throughput screening platforms (Nightingale, 2000); 2) patents that protect *the core structure of a product* in the biotech industry, or *composition of matter/compound patents* (Norman, 2007); 3) patents protecting *methods of utility*. In the biotech industry, these are *utility patents* (Norman, 2007), (note that in the US utility patents also have a broader meaning referring to the "normal" invention type of patents, as distinct from for example plant or design patents); 4) patents protecting *the specific application* of the core structure of the product. In the biotech industry, or *formulation patents* (Norman, 2007); 5) patents protecting the *method of manufacturing* (*process patents* (Norman, 2007)); and 6) patents protecting the *method of delivery* (*instrument patents* (Gupta and Bansal, 2002; Kaushal and Garg, 2003))*.* Different patented inventions take a certain form and focus that is contingent on the phase of R&D in the

invention process. Fig. 1 provides a visualization of four selected patent types, which are the most common inventions brought together to form a drug candidate.

Insert Figure 1 here

3.2 Identifying inventions in early vs. late R&D

This paper uses a text mining methodology developed by the authors to identify six patent types in Derwant patent abstracts, using semantic structures, keywords, and IPC codes. The text-mining method was developed in three phases: 1) generating a text-mining algorithm, 2) testing and adjusting the text-mining method, and 3) validating the results of the text-mining algorithm with industry players. Each step is described below [a separate report with in-depth details on the text-mining algorithm is available from authors upon request].

3.2.1 Generating a text-mining algorithm

First, a literature review of patent types in the biotech industry was conducted. Since the biotech industry developed out of the pharmaceutical industry (Hopkins et al., 2007), the patent types utilized are closely related to both industries. After identifying patent type, we interviewed industry stakeholders to ensure that our patent type identifications were in line with industry praxis. To develop the text mining method, we first randomly selected 85 biotech patents and categorized them "manually" into the six types. Iterations allowed identification of differentiating semantic structures (keywords and sentence pieces) and IPC codes. A range of prior biotech patent related studies helped us to identify keywords and IPC codes and categorize the patents: the OECD methodology for identifying biotechnology patents (OECD, 2005), the results of the biotechnology comparative study on patent rights by EPO, USPTO, and JPO (EPO et al., 1998), the patent search literature focused on bioscience (Yoo et al., 2005), and Dirnberger's (2011) case of the human recombinant insulin patent landscape.

In total, 608 semantic structures were identified as belonging to a patent type: Platform (142), Compound (129), Process (98), Formulation (87), Instrument (77), and Utility (75). A further 255 IPCs were identified as belonging to a patent type: Compound (98), Formulation (10), Instrument (29), Platform (28), Process (87), and Utility (3). Each patent was examined to see how it matched the 608 semantic structures and 255 IPCs, frequencies of 'hits' for each patent type were noted for number of semantic structures and IPCs. The frequencies of IPCs and semantic structures were then weighted. The highest score indicated the type of patent. Approx. 10 % of patent scores were equally high in several categories and were assigned to a 'mixed-patent' category.

3.2.2 Testing the automatic text-mining method

The above categorization was translated into an algorithm to allow patent type categorization to be handled automatically by a script. Two tests were performed to check whether the machine method categorization (the algorithm) produced the same result as manual categorization. First, a test of the 1,079 patents from 107 biotech firms was conducted. The 1,079 patents were read and categorized manually into patent types and the results were compared with the results of the automatic machine coding: In 91% of cases the classifications were the same. Second, we tested patents from a large biopharmaceutical company, Novo Nordisk: 5% of their patent portfolio (1,937 patents) was randomly selected and manually classified; for 92% of these patents manual and automatic categorizations were the same.

3.2.3 Validating the text-mining method with industry experts

To ensure that the results of the automatic text mining method correspond to what the industry experts would identify as certain patent types, three external tests with two small biotech firms and one large pharmaceutical firm were conducted. In biotech firm A, the firm's IP counsel's categorization matched the automatic categorization in 9 out of 11 patents. In biotech firm B, the firm's IP counsel's

categorization matched the automatic categorization in 9 out of 13 patents. In the large pharmaceutical firm, H. Lundbeck A/S, 50 patents – representing one-sixth of the total patent portfolio – were randomly chosen for the test; 35 patents were categorized similarly using the automatic categorization. Thus the output from the machine patent type categorization is satisfactory.

3.3 Variables

Identifying valid proxies for search behavior and for the firms' underlying cognitive maps in the process of creating new inventions represents a number of challenges. Management researchers have relied on seeing patents as the outcome of search behavior and characterized the search process by means of backward citations inserted in patents (Benner and Tushman, 2002; Katila and Ahuja, 2002; Laursen, 2011; Laursen et al., 2010; Phelps, 2010). Studies of U.S. patents have pointed to the noise involved when backward citations, including those inserted by patent examiners, are used as proxies for search behavior in innovation (Alcacer and Gittelman, 2006; Alcacer et al., 2009; Jaffe et al., 2000; Roach and Cohen, 2012). Since backward citations are handled differently in the European and U.S. contexts (Criscuolo and Verspagen, 2008), the validity of utilizing backward citations in the European context has been analyzed. Duguet and MacGarvie (2005) provide evidence that backward citations in EPO patents are a fairly good measure of firm level activities, such as M&As, R&D cooperation, etc. However, in line with the recent literature, we do not use individual backward citations as precise proxies but rather as count and ratio-measures for type of search, either scientific or technological, in the innovation process of a patented invention. There are two distinct types of backward citations: a) non-patent literature citations which for chemical and pharmaceutical patents predominantly are citations to scientific journals; these we use as a proxy for the underlying cognitive map guiding the search, namely the *science based cognitive map*, and b) backward citations to prior patents, which we use to proxy the *technology based cognitive map guiding the search.*

In our estimations in Table 3 in Model I and II, we use different types of backward citations as dependent variables: NPL_CIT is the number of backward citations to non-patent literature, PAT_CIT is the number of backward citations to prior patents.

The main independent variables in Table 3 are dummy variables that are coded 1 for each of the four patent types, and are described briefly here [in-depth descriptions available from the authors]:

1) *Platform Patents Category*: Methods for analyzing and selecting compounds and entities; methods for identifying molecules, etc.

2) *Compound Patents Category:* Compounds, molecules, proteins, peptides, enzymes, receptors, derivatives, analogues, and variants, etc.

3) *Utility (Use) Patents Category:* The usage of the compounds and the technologies developed; treatments and methods to alleviate the symptoms of various diseases; and the use of an entity to prepare or manufacture another useful product.

4) *Formulation Patents Category:* New dosage forms, pharmaceutical preparations, etc.

Instrument and process patents were removed from the sample, as they account for less than 50 patents. In Table 3 Model III to VI patent value is a dependent variable. We identify patent value by employing two distinct value correlates:

$$PATVAL = St(Familysize) + St(Forward\ citations)$$

According to Harhoff, Scherer, and Vopel (2003) and Gambardella, Harhoff et Al. (2008), patent value indicators carry noise, this implies they should be applied with caution. The literature suggests a one-sided approach to valuing individual patents by counting forward citations. One way to improve this one-sided approach is to measure patent value using by several indicators (Lanjouw and Schankerman, 2004). However, the Lanjouw and Schankerman (2004) measure include backward citations in the value

equation, we cannot do this as the backward citations are used as identifiers for our independent variables. We therefore apply a measure that incorporates two distinct patent value correlates: family size and forward citations. In Lanjouw and Schankerman (2004) these two measures are the only variables presented as positive and significant for patents in the drug industry, when regressed as determining patent litigation and patent renewal. Family size and forward citations have been included in various empirical studies of value indicators, and are shown to be less noisy (Gambardella et al., 2008). Another reason for choosing these two dimensions is that they contribute to our understanding of patent value in different ways. Family size is the number of countries in which the patent has been applied for, and is a proxy for how core the patent is for the firm. Firms apply for patents in a range of countries depending on the importance to them of the patent: if the patent is non-core, the firm will apply for protection in a smaller number of countries; if it is core, the firm will apply to a higher number of countries. The correlate of forward citations is measured by the total number of forward citations the patent has received, a proxy for how important the patent has been for subsequent inventions within the area.

Inventions are cumulative in their protection of a given drug candidate. For example, a formulation patent is likely to build on a prior compound patent, whereas compound patents rarely cite their formulation counterparts. Thus inventions coming out of late stage R&D will tend to receive fewer citations. We therefore apply both value proxies and conduct tests to check their robustness. The two proxies for forward citations and family size have low correlation (0.22). We standardize each proxy to ensure an equal balance of value. We standardize using the following approach, example is family size:

$$St(Familysize) = Familysize - Mean\ Familysize / Standard\ deviation\ Familysize.$$

Two different independent variables are employed in Table 3 Model III to VI. The independent variable (NPL_RATIO) estimates to what degree the science based cognitive map has been replaced by

the technology based cognitive map; the measure is created by taking the total number of non-patent literature citations and dividing them by the total number of backward citations, i.e. scoring 1 when search is in science only. The second measure for the linkage to science is a dummy variable that takes the value 1 if the invention has one or more non-patent literature citation(s) and 0 if there are no non-patent citations in the patent. This variable indicates whether any search in science has been undertaken.

3.4 Estimation Approach and Control Variables

To analyze the occurrence of citations to non-patent literature citations (NPL) and patent citations (PAT CITs) in patents protecting different types of inventions (Platform, Compound, Utility and Formulation) we estimate the following models (Hypotheses 1 and 2):

$$1)\ Y_{pft} = \alpha + \beta_1(Platform) + \beta_2(Compound) + \beta_3(Utility) + \beta_4(Formulation) + \delta_{pft}$$

The outcome variable Y is a count of either non-patent literature (NPL) or patent citations (PATCITs), the unit of analysis is the individual patent (p) applied by firm (f) at a given time (t). Since we are interested in non-patent literature and patent citations influenced by other possible patents, firms, and time variables, we control for the number of variables represented by δ_{pft}. Our approach with our control variables focuses on controlling for two types of potentially confounding factors, those that affect the overall volume of searching as this could influence approaches sought in the individual searches which we analyze, and those related to the underlying unobserved heterogeneity of patents and firms. As patent level controls we include the total number of backward citations, the technology scope of the patent measured by number of IPC codes, the log of patent family size, whether the patent has been granted, and whether the patent has been withdrawn. As firm-level controls, we use a firm size variable measuring the number of employees in the firm, the year of patent application, the number of the firm's patents applied for before the year of patent application, firm type measured by whether the

firm is focused on small or large molecules, and firm age (whether it is less than four years old) at the time of the patent application. For the time variable, we control for year of patent application. For the estimations, we considered a Poisson regression since the dependent variable is a count variable. However, the variance is higher than the mean, which indicates that the data are overdispersed. This suggests a negative binomial regression in preference to a Poisson regression (Wooldridge, 2009a). This choice is confirmed by our results, which show that α values are significantly different from zero. We employ robust estimators to avoid heteroskedasticity, and to test for multicollinearity we perform VIF tests.

To understand how differences in the composition of search between early and late R&D affects patent value (hyp 3 and hyp 4) NPL, we estimate:

$$2)\ PATENTVALUE_{pft} = \alpha + \beta_1 (SCIENCE) + \delta_{pft}$$

The outcome variable is patent value and the unit of analysis is patent (p) applied for by firm (f) at a given time (t). The independent variable 'SCIENCE' is analyzed in two ways: considering the relative prevalence of scientific search (the NPL_RATIO), and applying the dummy NPL_DUM indicating whether a science based cognitive map has been utilized during innovation activities. We run separate models, for early inventions (Compound patents) and for late inventions (Formulation patents). We are careful in our indication of early and late R&D, as we are aware that not all types of patents are equally distinct. The one area where these distinctions may sometimes be less distinct is in the separation of compound and utility patents, essentially because the spillovers between these two research agendas are considerable. We also see their profiles being more similar in terms of search in science and technology. That is why we conduct our test based on comparison of compound and formulation patents, which require quite different and separate projects.

Again, we are concerned with the influence of firm and time on patent value, and employ a number of variables to control for this, represented by δ_{pft}, and cluster for the effects of the single firm

behind the invention. To estimate the model we initially use the Tobit technique since the dependent variable is lower limited (Wooldridge, 2009a); however, the empirical results for Compound and Formulation patents show very few observations at the lower limit (1 observation for Compound patents and 3 for Formulation patents). We therefore present the results of an OLS regression in Table 3. The results do not differ much from the Tobit estimates but are easier to read. Since the OLS and Tobit regressions are mean centered, we may not capture the differences in the effects at different locations in the patent value distribution and also may not be able to explain the extreme observations in the tails. To explore the degree of 'extreme tails' we follow Koenker (Koenker, 2005; Koenker and Bassett, 1978) and use percentile regressions to investigate the effect at different locations in the patent value distribution (Table 4 and Figure 2).

4. Results

4.1 Descriptive results

Table 1 provides descriptive data for all variables.

Insert Table 1 here

Table 1 shows several interesting features. The patent type referring to compounds accounts for 52% of the total sample. This predominance is explained not just by the protection of the core technology of a drug candidate (Norman, 2007). Compounds are invented early in the drug development process, and the high attrition rate in drug discovery means that development may be discontinued. Compound patents outnumber initial platform patents because a single platform is capable of multiple targets.

The correlation matrix in Table 2 shows only a few significant relationships. Multicollinearity (coefficient > 0.70) is found only where it would logically be expected, e.g. between the total number of

backward citations and backward citations to patents or the non-patent literature, and the number of non-patent literature citations and x citations to non-patent literature or citations to patents and x citations to patents.

Insert Table 2 here

4.2 Shifts in search patterns from early to late R&D processes

Hypotheses 1 and 2 address shifts in the R&D cycle related to its orientation to science vs. technology. Calculations (not presented here) show that for Platform patents 50% of all backward citations are to science, the share drops to 37-40% for Compound and Utility patents and declines further to 20% for Formulation patents in late R&D. So from a descriptive viewpoint shifts do occur in the role of science. These shifts occur in a setting where search throughout R&D predominantly has a technological orientation.

Hypothesis 1 proposes a decrease from early to late R&D in the role of science as a cognitive map. Model I in Table 3 takes as the dependent variable the number of backward citations to science (NPL) found in each patent. Differences across the R&D cycle are tested taking Platform patents as the benchmark, i.e. the patent type appearing at the front end of the drug R&D cycle. Negative significant estimates for the other patent types in Model I indicate their overall lower foundation in science. Differences between these other types are brought out by Wald tests. These tests show significant downward shifts in the estimates from Platform to Compound (see model I) and Compound to Formulation patents (chi sq=50.90***), confirming Hypothesis 1. The results for Utility patents do not follow the pattern suggested by a general R&D process although utility patents most often are filed just after compound patents, utility patents rely more heavily on NPLs than compound patents (chi sq=14.22***). The reason for Utility patents citing non-patent literature significantly more than

Compound patents could be due to the underlying technological mechanisms expressed in utility patents. Utility patents usually refer to basic biomedicine scientific literature for configurations of usage in humans. For inventions created during late R&D, Hypothesis 2 predicts a high number of citations to prior technologies. The number of these references forms the dependent variable in Model II. Comparisons among patent types use platform patents as the benchmark. Wald tests show a significant increase in citations to technology in the shift from Platform to Compound patents (see Model II) and from Compound to Formulation Patents (chi sq= 43.91***), confirming Hypothesis 2. Models I and II control for the number of backward citations in each patent; Hypotheses 1 and 2 are confirmed for any level of search activity.

Hypotheses 3 and 4 concern the effects on patent value of different compositions of search. Models III and IV in Table 3 examine the case of early R&D as represented by Compound Patents. The composition of search is indicated by the ratio of citations to science over the total number of backward citations (NPL RATIO). The significant negative estimate for this ratio indicates decreasing value for an increasing share of citations to science, confirming Hypothesis 3. We also test the relationship (in Model IV) using a dummy that takes the value 1 for the presence of at least one NPL citation. A negative, significant estimate also is obtained for this more conservative indicator.

Further analysis of the ratio indicators offers insights into differences in the effect on patent value for different levels of science search predominance. To examine these differences we use percentile regressions and examine difference across the distribution by means of Wald tests. The results in Table 4 and Figure 2 show significant differences. When science search increases by 1 percentile, the coefficient estimates of the independent variable non-patent literature ratio (NPL RATIO) decrease for every percentile until 0.5. The effects are not significantly different from each other (-0.13 to -0.18) for lower levels of search in science (between values of the ratio from 0.1 to 0.5). However, after 0.5 the

penalizing effect increases for every percentile, being four times more negative for very strong predominance of search in science (ratio value of 0.9) compared to moderate levels. Wald tests also show significant differences after 0.5 (q50 to q90: 6.73***, q 50 to q80:3.30**, q60 to q90:5.04** and q70 to q90: 3.07**). In short, patent value decreases to the same moderate extent up to the level where search in science defines roughly half of the cognitive maps guiding search; thereafter increases in the orientation towards science detract from patent value at an increasing rate.

Insert Table 4 & Figure 2 here

Turning to the effects on value of inventions made in late stage R&D, Models V and VI examine the case of Formulation patents. Two mechanisms are proposed as influencing the effect of science on the value of late stage inventions. Following the literature on cognitive variety would indicate that having a blend of both science and technology inputs would have a positive influence on outcome, while familiarity traps theory (Fleming, 2001) would suggest that a positive effect on value would be conjectured from the presence of just an element of search in science. The descriptive data for late stage inventions (n=110), Formulation patents, on average show a total of six backward citations, of which only 1.5 are to science and the rest are to prior technology (see Table 1). The estimate in Model V shows that the NPL ratio is not significant, thus not supporting the arguments about cognitive variety. However, the positive estimate of the dummy in Model VI indicates the positive effects on patent value conjectured in Hypothesis 4. This estimate is significant only at the 10% level, and is obtained without controlling for the total number of backward citations. Supplementary models that include this control (not presented here) retain a positive sign, while significance drops to the 16% level. Overall, the results are mixed and suffer from the restrictions imposed by small sample size for this patent type. However, it seems that

Hypothesis 4 is confirmed. Inventions from late stage R&D increase in value with the addition of elements of search in science.

4.3 Robustness Checks

The results estimated in Models I to VI broadly support our hypotheses. Additional tests (not presented here, available upon request) were performed to examine the robustness of our findings. Due to only a few censored observations, Models III-VI apply OLS rather than Tobit. Alternative estimates were performed applying Tobit (Wooldridge, 2009b), using the tests indicated by Wiersema and Bowen (2009) and Bowen (2012). We also ran Model III-VI excluding the total number of backward citations to examine whether the effects rely on total scope of search. In all cases the results remain the same as in the main models.

To examine the robustness of our patent value measure (Model III-VI) we estimate the models splitting the dependent variable into its components of forward citations and family size (see Table 5). The results remain significant for Model III-IV for both robustness checks; however we find no support for hypothesis 4 in either of the single value estimations. This might indicate that when using patent value indicators as the dependent variable in empirical hypothesis testing with relatively low numbers, combining several well selected value indicators based on industry characteristics would enhance the accuracy of the indicator, and possibly remove noise from other indicators. We also estimate our models inserting dummies for each firm, the results change for compound patents the effects become stronger and the significance level increases using this model specification (Model III NPR ratio = -0.502*** and Model IV the NPR dummy=-0.294***), however, for formulation patents both estimates are insignificant, though remaining negative coefficients. In addition, in regards to the trunctation of forward citations, we estimate models III to VI using the recent published OECD/EPO dataset (Squicciarini M. et al., 2013) and find that in models where the patent value measure is constructed using either a 5 or 7 year window for forward citations the results for Model III, NPR ratio equals -0.571***(5 year window) and -

0.527***(7year window), in Model IV the NPR dummy equals -0.305***(5year window) and -0.269***(7 year window). Model V and VI for Formulation patents shows negative coefficients however non-significant results (the p-value is p=0.5 for the NPR ratio measure and p=0.18 for the NPR dummy measure), this means that the results for Formulation patents aren't confirmed using this model specification.

We also test Models I and II using only backward citations marked X by the patent examiners as referring to *"Particularly relevant documents when taken alone (a claimed invention cannot be considered novel or cannot be considered to involve an inventive step)"* (p.8 Webb et al., 2005). Prior studies argue that restriction to X-marked citations reduces their noise as indicators of knowledge flow (Criscuolo and Verspagen, 2008). Results are largely the same (see Table 6).

One endogeneity issue to consider in the current study is concerned with the possibility that some firms strategically might aim for higher valued patents, introducing a firm effect into our findings. To address this possibility we include a clustering effect, and also test the results including firms as dummy variables in the regression, the results does not change (results not included here).

5. Discussion and Conclusions

In examining a science-based industry, this paper studied orientations in the cognitive maps applied across the R&D cycle. First, we investigated whether different steps in the R&D cycle have different emphasis on the two orientations of a science based cognitive map and a technology based cognitive map. Second, even if shifts occur across the R&D cycle in the composition of search, both orientations may be present in all the stages. We examined whether the two orientations exhibit orthogonal presence, balanced only in different proportions in different parts of the cycle, and whether shifts in these balances affect the value of the inventions associated with that particular part of the R&D cycle.

To address these issues we capitalized on the clearly structured steps characterizing R&D in biotech firms specialized in drug discovery. Given the tight and aggressive appropriability regime of this industry (Levin et al., 1987), firms tend to file patents on inventions coming out of specific steps in R&D leading to a single drug, for example, on the compound, its utility, or its formulation. On the basis of the fairly standardized sequence of these inventions, we made a clear distinction between early R&D (directed e.g. at compounds or at the platform developed for their discovery) and late R&D (directed e.g. at formulation).

The drug industry leaves detailed patent paper trails from each stage of the R&D. Particularly important for our purposes are the footprints left by each patent of the search associated with each invention. The backward citations in a patent to the non-patent literature indicate the extent to which the search behind the invention is guided by a science based cognitive map, or directed by a technology based cognitive map, indicated by backward citations to previous patents (Benner and Tushman, 2002; Katila and Ahuja, 2002; Laursen, 2011; Laursen et al., 2010; Phelps, 2010). As expected, science-based maps are more prevalent in the forward-looking explorative orientation of search in early R&D, while cognitive maps dominated by a technological orientation are more predominant in late R&D. However, in both cases we observed a blend of cognitive maps, rather than total replacement of one by the other. In essence, what differs is the balancing point between the two different approaches. Although these relations seem intuitive, they are not established in the theoretical innovation literature and have previously not been made accessible for systematic large N-based statistical tests.

We next examined the effects on value associated with applying cognitive variety within each invention. Since orientations towards science and technology are balanced differently in the early and late stages of R&D, we designed separate tests. For early stage R&D, we found that increases in the prevalence of science in search detract from invention value. At the point when the increase in the

science based cognitive map approaches total predominance, the decline in value accelerates markedly. Conversely, late stage R&D with its search pattern dominated by the technology based cognitive map, achieves increases in invention value if elements of science based cognitive search are added.

The main contribution of this paper is to the literature on sources of knowledge for innovation. Prior studies emphasize that both science and technology are potentially useful sources for R&D (Brooks, 1994; Jensen et al., 2007; Rosenberg, 1990; Tijssen, 2002). Our results confirm this and extend the literature by empirically testing the variation in each source of knowledge at different stages of R&D, and the contribution of knowledge sources, relative to their position in R&D. We contribute to organizational theory by proposing how search during different stages affects performance in dynamic environments, and add to the research on the micro-foundations of superior performance in dynamic environments since the tests in this paper confirm the value of cognitive variety discussed in prior research (Brown and Eisenhardt, 1997; Eisenhardt et al., 2010; Smith and Tushman, 2005) supporting recent contributions in organizational science, that highlight how understanding the micro-foundations increases our understanding of organizational performance (Abell et al., 2008; Gavetti, 2005; Teece, 2007; Zahra and Wright, 2011). We respond to recent calls for research on organizational learning processes and their interrelatedness (Argote and Miron-Spektor, 2011).The methodological contributions of the paper include its use of citations as indicators of search direction and our input to the more general debate on the role of science. There is a small but growing literature on the contributions to value creation from science when it is an integral part of the problem solving in industrial R&D. So far these effects have been studied in the form of the direct impact of science citations/linkages on patent value (Cassiman et al., 2008; Harhoff et al., 2003; Nemet and Johnson, 2012). Previous studies provide inconclusive results from this approach. In contrast, we find fairly strong effects when applying a contingency design, conjecturing effects for early R&D as the reverse of those from late R&D. We find a negative, significant effect when the orientation towards science becomes predominant in the overall pattern of search.

When this predominance is such that there is only marginal attention to earlier technology, the negative effects on patent value accelerate when the share of citations to earlier technology drops to less than half of its average level. Conversely, the *positive* effects of including science in the search pattern emerge strongly in the context of late stage R&D. This result confirms the prior literature that claims that science is beneficial when coping with potential familiarity traps during later stages of R&D (Fleming, 2001; Nelson, 1982).

Regarding generalizability of our findings, at least two dimensions are interesting to consider, namely geography and industry. As for geography we believe that the results apply worldwide. However, replications of the study should be attentive to the way backward citations are added to patents in different systems (e.g. differences between the US and the European systems). In terms of industry, this study is based on data from drug discovery biotech firms. Large pharmaceutical firms discover and develop drugs based on the same sequence of invention-types presented here. Our findings arguably apply to them as them as well. However, empirical tests if that expectation would require a much stronger focus at the level of the individual project, as focus and resources vary across projects. More generally we expect the main patters identified in this study to apply to indisputably science-based sectors (as defined above) but not to industries in which innovations are based principally on the combinations of existing, well established technologies (Stankiewwicz, 2000)

5.1 Limitations

As discussed in the method section, using backward citations as an indicator of search in either science and/or technology has some limitations. An alternative approach might be to reconstruct the cognitive maps applied in science-driven R&D using interviews or on site observation of inventors. This approach is not feasible for a large N dataset, and would introduce other biases into the econometric analysis, since

individuals would be biased by familiarity with and perceptions of the social structures in the scientific community. Utilizing patent citations, and including patent examiners' citations, would reduce these biases. We tried to overcome the shortcomings of backward citations by relying on an industry where the link to science is strong, and where both patenting and scientific publishing is goals.

Our robustness checks do not control for self-citations, since this indicator is not available in our data. However, because we rely on EPO patents this should not have an effect on our findings. Criscuolo and Verspagen (2008) show that self-citation does not have any effect on qualitative results with regard to backward citations to EPO patents. However this would not apply to analysis based on U.S. patent data.

Finally, we do not control for "type" of backward citations. In the non-patent literature a deeper understanding of the type of science behind the individual citation would have had more explanatory power, and would complement recent work aimed at understanding the nature of science and its interrelatedness with industry (Sauermann and Stephan, 2013).

The approach in the present paper could be extended in several ways. One important direction for further work would be to try to link individual types of inventions (platform, compound etc.) as contributions to the same drug. So far we have not identified a methodology allowing these linkages to be reliably established in a large N dataset. In the current setting however we do not expect that controlling for the project for which the individual invention is generated under would influence the results, this is mainly due to the nature of the dataset. The dataset which mainly consists of small biotech firms does not have resources to engage in many different projects at the same time, often they would focus on one or maximum two projects.

6. References

Abell, P., Felin, T., Foss, N., 2008. Building micro-foundations for the routines, capabilities, and performance links. Managerial and Decision Economics 29, 489-502.

Achilladelis, B., Antonakis, N., 2001. The dynamics of technological innovation: the case of the pharmaceutical industry. Research Policy 30, 535-588.

Ahuja, G., Lampert, C.M., 2001. Entrepreneurship In The Large Corporation: A Longitudinal Study Of How Established Firms Create Breakthrough Inventions. Strategic Management Journal 22, 521-543.

Alcacer, J., Gittelman, M., 2006. Patent citations as a measure of knowledge flows: The influence of examiner citations. Review of Economics and Statistics 88, 774-779.

Alcacer, J., Gittelman, M., Sampat, B., 2009. Applicant and examiner citations in US patents: An overview and analysis. Research Policy 38, 415-427.

Argote, L., Miron-Spektor, E., 2011. Organizational Learning: From Experience to Knowledge. Organization Science 22, 1123-1137.

Atuahene-Gima, K., 2005. Resolving the Capability-Rigidity. Paradox in New Product Innovation. Journal of Marketing 69, 61-83.

Benner, M.J., Tushman, M.L., 2002. Process management and technological innovation: A longitudinal study of the paint and photography industries. Administrative Science Quarterly 47, 676-698.

Benner, M.J., Tushman, M.L., 2003. Exploitation, exploration, and process management: The productivity dilemma revisited. Academy of Management Review 28, 238-256.

Bowen, S., 2012. Testing Moderating Hypotheses in Limited Dependent Variable and Other Nonlinear Models: Secondary Versus Total Interactions. Journal of Management 38, 860-889.

Brooks, H., 1994. THE RELATIONSHIP BETWEEN SCIENCE AND TECHNOLOGY. Research Policy 23, 477-486.

Brown, J.S., Duguid, P., 2001. Knowledge and organization: A social-practice perspective. Organization Science 12, 198-213.

Brown, S.L., Eisenhardt, K.M., 1997. The art of continuous change: Linking complexity theory and time-paced evolution in relentlessly shifting organizations. Administrative Science Quarterly 42, 1-34.

Bruno Cassiman, R.V., Zuniga, P., 2007. In Search of Performance Effects of (in)direct Industry Science Links.

Carroll, J.B., 1993. Human Cognitive Abilities: A Survey of Factor-Analytic Studies. . Cambridge University Press, New York.

Cassiman, B., Di Guardo, M.C., Valentini, G., 2010. Organizing links with science: Cooperate or contract?: A project-level analysis. Research Policy 39, 882-892.

Cassiman, B., Veugelers, R., Zuniga, P., 2008. In search of performance effects of (in)direct industry science links. Industrial and Corporate Change 17, 611-646.

Chandy, R., Hopstaken, B., Narasimhan, O., Prabhu, J., 2006. From invention to innovation: Conversion ability in product development. Journal of Marketing Research 43, 494-508.

Cohen, M., Nelson, R., Walsh, J., 2000. Protecting their intellectual assets: Appropriability conditions and why US manufacturing firms patent (or not). . National Bureau of Economic Research.

Cohen, W.M., Nelson, R.R., Walsh, J.P, 2002. Links and impacts: The influence of public research on industrial R&D. Management Science 48, 1-23.

Criscuolo, P., Verspagen, B., 2008. Does it matter where patent citations come from? Inventor vs. examiner citations in European patents. Research Policy 37, 1892-1908.

DiMasi, J.A., 2001. New drug development in the United States from 1963 to 1999. Clinical Pharmacology & Therapeutics 69, 286-296.

Dirnberger, D., 2011. A guide to efficient keyword, sequence and classification search strategies for biopharmaceutical drug-centric patent landscape searches - A human recombinant insulin patent landscape case study. World Patent Information 33, 128-143.
Duguet, E., MacGarvie, M., , 2005. How well do patent citations measure flows of technology? Evidence from French innovation surveys. Economics of Innovation and New Technologies 14, 375-394.
Eisenhardt, K.M., Furr, N.R., Bingham, C.B., 2010. Microfoundations of Performance: Balancing Efficiency and Flexibility in Dynamic Environments. Organization Science 21, 1263-1273.
EPO, USPTO, JPO, 1998. Trilateral Project 24.1 Biotechnology Comparative Study on Biotehnology Patent Practices Comparative Study.
Fleming, L., 2001. Recombinant uncertainty in technological search. Management Science 47, 117-132.
Fleming, L., Sorenson, O., 2004. Science as a map in technological search. Strategic Management Journal 25, 909-928.
Gambardella, A., Harhoff, D., Verspagen, B., 2008. The value of European patents European Management Review 5, 69-84.
Gavetti, G., 2005. Cognition and hierarchy: Rethinking the microfoundations of capabilities' development. Organization Science 16, 599-617.
Gavetti, G., Levinthal, D., 2000. Looking forward and looking backward: Cognitive and experiential search. Administrative Science Quarterly 45, 113-137.
Gibbons, M., Johnston, R., 1974. ROLES OF SCIENCE IN TECHNOLOGICAL INNOVATION. Research Policy 3, 220-242.
Gittelman, M., Kogut, B., 2003. Does good science lead to valuable knowledge? Biotechnology firms and the evolutionary logic of citation patterns. Management Science 49, 366-382.
Griliches, Z., 1984. R&D, Patents and Productivity. University of Chicago Press: Chicago.
Griliches, Z., 1986. PRODUCTIVITY, RESEARCH-AND-DEVELOPMENT, AND BASIC RESEARCH AT THE FIRM LEVEL IN THE 1970S. American Economic Review 76, 141-154.
Gupta, A.K., Smith, K.G., Shalley, C.E., 2006. The interplay between exploration and exploitation. Academy of Management Journal 49, 693-706.
Gupta, P., Bansal, A.K., 2002. Patent Opportunities in Matrix-Based Oral Controlled Release Drug Delivery Systems, Part I. Pharmaceutical Technology Europe, 14(9): 49.
Harhoff, D., Scherer, F., Vopel, K., 2003. Citations, family size, opposition and the value of patent rights. Research Policy 32, 1343-1363.
Hopkins, M.M., Martin, P.A., Nightingale, P., Kraft, A., Mahdi, S., 2007. The myth of the biotech revolution: An assessment of technological, clinical and organisational change. Research Policy 36, 566-589.
Iansiti, M., West, J., 1999. From Physics to Function: An Empirical Study of Research and Development Performance in the Semiconductor Industry. Journal of Product Innovation Management 16, 385-399.
Jaffe, A.B., Trajtenberg, M., Fogarty, M.S., 2000. Knowledge spillovers and patent citations: Evidence from a survey of inventors. American Economic Review 90, 215-218.
Jensen, M.B., Johnson, B., Lorenz, E., Lundvall, B.A., 2007. Forms of knowledge and modes of innovation. Research Policy 36, 680-693.
Katila, R., Ahuja, G., 2002. Something old, something new: A longitudinal study of search behavior and new product introduction. Academy of Management Journal 45, 1183-1194.
Kaushal, A.M., Garg, S., 2003. An Update on Osmotic Drug Delivery Patents. Pharmaceutical Technology 27(8): 38.
Klevorick, A.K., Levin, R.C., Nelson, R.R., Winter, S.G.,, 1995. On the sources and significance of interindustry differences in technological opportunities. Research Policy, 185-205.
Kline, S.J., Rosenberg, N., 1986. An Overview of Innovation" in Landau, R. and Rosenberg, N.: *The Positive Sum Strategy*., Washington D. C. 1986, excerpt pp. 285 - 298.

Koenker, R., 2005. Quantile Regression. Cambridge University Press.
Koenker, R., Bassett, G., 1978. REGRESSION QUANTILES. Econometrica 46, 33-50.
Langlois, R.N., 1999. Scale, Scope and the Reuse of Knowledge, in: Dow, S.C.a.E., Peter E. (Ed.), Economic Organization and Economic Knowledge. . Cheltenham, Edward Elgar. , pp. 239-254.
Lanjouw, J.O., Schankeman, M., 2004. Patent quality and research productivity: Measuring innovation with multiple indicators. Economic Journal 114, 441-465.
Lanjouw, J.O., Schankerman, M., 2004. Patent quality and research productivity: Measuring innovation with multiple indicators. Economic Journal 114, 441-465.
Laursen, K., 2011. Keep searching and you'll find: What we know about variety creation through firms' search activities for innovation? Forthcoming Industrial and Corporate Change.
Laursen, K., Leone, M.I., Torrisi, S., 2010. Technological exploration through licensing: new insights from the licensee's point of view. Industrial and Corporate Change 19, 871-897.
Leonard-Barton, D., 1992. Core capabilities and core rigidities: A paradox in managing new product development. Strategic Management Journal 13, 111-125.
Levin, R.C., Klevorick, A.K., Nelson, R.R., Winter, S.G., Gilbert, R., Griliches, Z., 1987. Appropriating the Returns from Industrial Research and Development. Brookings Papers on Economic Activity 1987, 783-831.
Mansfield, E., 1972. CONTRIBUTION OF R AND D TO ECONOMIC GROWTH IN UNITED-STATES. Science 175, 477-&.
Mansfield, E., 1991. ACADEMIC RESEARCH AND INDUSTRIAL-INNOVATION. Research Policy 20, 1-12.
Mansfield, E., 1998. Academic research and industrial innovation: An update of empirical findings. Research Policy 26, 773-776.
March, J.G., 1991. Exploration and Exploitation in Organizational Learning. . Organization Science, , 2(1): 71-87.
Neisser, U., 1976. Cognition and Reality: Principles and Implications of Cognitive Psychology. W. H. Freeman, San Francisco.
Nelson, R.R., 1982. The Role of Knowledge in R&D Efficiency. The Quarterly Journal of Economics 97, 453-470.
Nelson, R.R., 2003. On the uneven evolution of human know-how. Research Policy 32, 909-922.
Nemet, G.F., Johnson, E., 2012. Do important inventions benefit from knowledge originating in other technological domains? Research Policy 41, 190-200.
Nightingale, P., 2000. Economies of scale in experimentation: knowledge and technology in pharmaceutical R&D. Journal of Industrial and Corporate Change 9, 315–359.
Norman, P., 2007. Patent Protection Strategies - Maximising Product Revenues., in: Insights, B. (Ed.).
OECD, 2005. A FRAMEWORK FOR BIOTECHNOLOGY STATISTICS.
Phelps, C.C., 2010. A LONGITUDINAL STUDY OF THE INFLUENCE OF ALLIANCE NETWORK STRUCTURE AND COMPOSITION ON FIRM EXPLORATORY INNOVATION. Academy of Management Journal 53, 890-913.
Piaget, J., 1985. The Equilibration of Cognitive Structures: The Central Problem of Intellectual Development. University of Chicago Press, Chicago.
Pisano, G., 1996. Development Factory: Unlocking the Potential of Process Innovation.
Roach, M., Cohen, W., 2012. “Lens or Prism? An Assessment of Patent Citations as a Measure of Knowledge Flows from Public Research.”. forthcoming Management Science.
Rosenberg, N., 1974. SCIENCE, INVENTION AND ECONOMIC GROWTH. Economic Journal 84, 90-108.
Rosenberg, N., 1990. Why do firms do basic research (with their own money)? Research Policy 19, 165-174.
Rosenberg, N., Nelson, R.R., 1994. AMERICAN UNIVERSITIES AND TECHNICAL ADVANCE IN INDUSTRY. Research Policy 23, 323-348.

Sauermann, H., Stephan, P., 2013. Conflicting Logics? A Multidimensional View of Industrial and Academic Science. Organization Science 24, 889-909.
Schumpeter, J.A., 1939. Business Cycles.
Smith, W.K., Tushman, M.L., 2005. Managing strategic contradictions: A top management model for managing innovation streams. Organization Science 16, 522-536.
Squicciarini M., Dernis H., C., C., 2013. Measuring Patent Quality: Indicators of Technological and Economic Value. OECD Science, Technology and Industry Working Papers, OECD Publishing 03.
Stankiewicz, 2000. The concept of 'design space', in: Ziman, J. (Ed.), Technological Innovation as an Evolutionary Process Cambridge University Press, Cambridge, pp. 234-247.
Sternitzke, C., 2010. Knowledge sources, patent protection, and commercialization of pharmaceutical innovations. Research Policy 39, 810-821.
Sternitzke, C., 2013. An exploratory analysis of patent fencing in pharmaceuticals: The case of PDE5 inhibitors. Research Policy 42, 542-551.
Teece, D.J., 2007. Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance. Strategic Management Journal 28, 1319-1350.
Tijssen, R.J.W., 2002. Science dependence of technologies: evidence from inventions and their inventors. Research Policy 31, 509-526.
Tushman, M.L., Oreilly, C.A., 1996. Ambidextrous organizations: Managing evolutionary and revolutionary change. California Management Review 38, 8-&.
Webb, C., Dernis, H., Harhoff, D., Hoisl, K., 2005. ANALYSING EUROPEAN AND INTERNATIONAL PATENT CITATIONS: A SET OF EPO PATENT DATABASE BUILDING BLOCKS STI WORKING PAPER 2005/9
Wiersema, M.F., Bowen, H.P., 2009. THE USE OF LIMITED DEPENDENT VARIABLE TECHNIQUES IN STRATEGY RESEARCH: ISSUES AND METHODS. Strategic Management Journal 30, 679-692.
Wooldridge, 2009a. Introductory Econometrics - A Modern Approach. South Western Cengage Learning.
Wooldridge, J., 2009b. Introductionary Econometrics - A modern approach. South-Western Cengage Learning.
Yayavaram, S., Ahuja, G., 2008. Decomposability in knowledge structures and its impact on the usefulness of inventions and knowledge-base malleability. Administrative Science Quarterly 53, 333-362.
Yoo, H., Ramanathan, C., Barcelon-Yang, C., 2005. Intellectual property management of biosequence information from a patent searching perspective. World Patent Information 27, 203-211.
Zahra, S.A., Wright, M., 2011. Entrepreneurship's Next Act. Academy of Management Perspectives 25, 67-83.

7. Figures and Tables

Figure 1: Patent types early vs. late in R&D

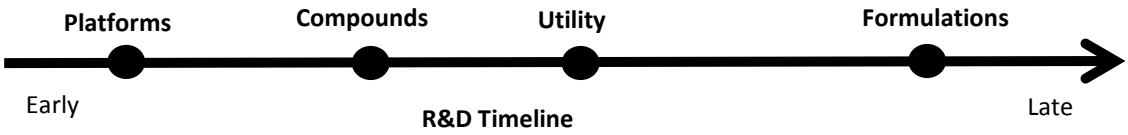

Figure 2: Percentile regression estimates for Non Patent Related citations ratio (NPLCit_RATIO) for compound patents

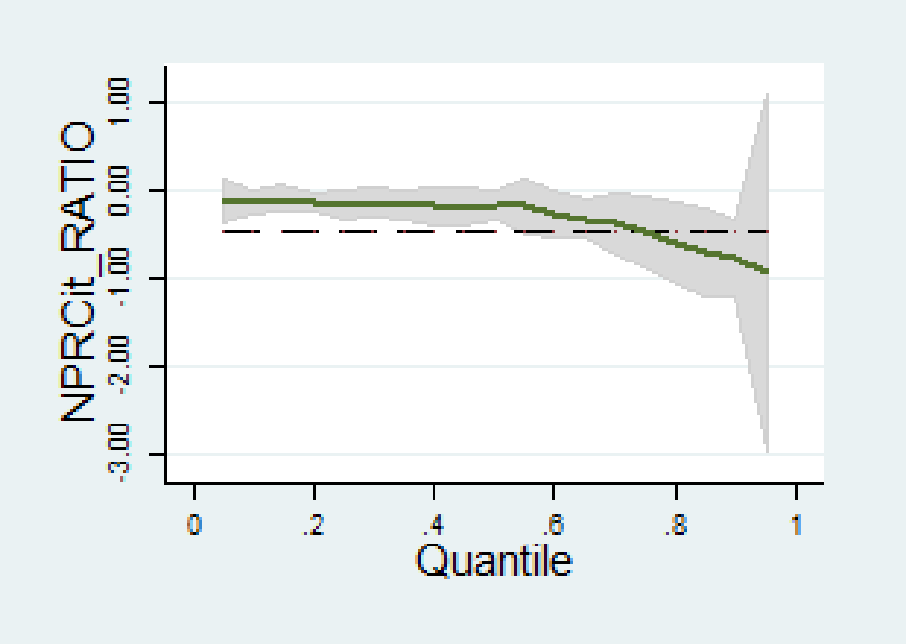

Table 1: Descriptive statistics

All patent types	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
Patent value	1058	-.143	1.40	-1.424	1.734
Compound patents	1058	.517	.49	0	1
Platform patents	1058	.224	.41	0	1
Utility patents	1058	.155	.36	0	1
Formulation patents	1058	.103	.30	0	1
Total number of backward citations	1058	7.28	5.15	0	35
Non-patent literature (NPL)	1058	3.03	3.53	0	22
Backward citations to patents (PAT_CIT)	1058	4.25	3.19	0	26
X-citations to non-patent literature (XNPL)	1058	1.61	2.91	0	22
X-citations to patents	1058	2.37	2.97	0	26
Technology scope	1058	9.70	11.28	1	86
Accumulated patents in the firm	1058	51.80	69.37	0	314
Firm size	1058	91.20	137.94	0	850
Type of firm	1058	.48	.50	0	1
Young firm	1058	.23	.42	0	1
Log of patent family size	1058	2.20	.83	0	4.882
Patent grant	1058	.33	.47	0	1
Patent withdrawn	1058	.23	.42	0	1
Year of patent application	1058			1997	2008
Compound patents					
Total number of backward citations	547	7.28	5.87	0	35
Non-patent literature citations (NPL)	547	2.70	3.66	0	22
Backward citations to patents	547	4.57	3.60	0	26
X-citations to non-patent literature (XNPL)	547	1.62	3.25	0	22
X-citations to patents (XPAT)	547	2.62	3.52	0	26
The total number of non-patent literature citations over total number of backward citations (NPL_RATIO)	547	.30	.28	0	1
Non-patent literature citations – dummy	547	.667	.471	0	1
Formulation patents					
Total number of backward citations	110	6.03	3.17	1	22
Non-patent literature citations (NPL)	110	1.53	2.26	0	13
Backward citations to patents	110	4.5	2.34	0	13
X-citations to non-patent literature (XNPL)	110	.59	1.46	0	7
X-citations to patents	110	2.36	2.18	0	10
The total number of non-patent literature citations over total number of backward citations (NPL RATIO)	110	.210	.24	0	1
Non-patent literature citations – dummy	110	.563	.498	0	1

Table 2: Correlation matrix

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	Patent value	1.00														
2	Compound patents	0.06	1.00													
3	Platform patents	0.00	-0.56	1.00												
4	Utility patents	-0.06	-0.44	-0.23	1.00											
5	Formulation patents	-0.02	-0.35	-0.18	-0.15	1.00										
6	Total number of backward citations	0.09	-0.00	0.02	0.04	-0.08	1.00									
7	Non-patent literature	0.01	-0.09	0.17	0.06	-0.14	0.79	1.00								
8	Backward citations to patents	0.13	0.10	-0.15	0.01	0.02	0.74	0.17	1.00							
9	X-citations to non-patent literature	0.06	0.00	0.06	0.03	-0.12	0.72	0.83	0.25	1.00						
10	X-citations to patents	0.13	0.09	-0.15	0.05	-0.00	0.67	0.24	0.82	0.38	1.00					
11	Technology scope	0.37	0.18	-0.09	-0.02	-0.13	0.17	0.07	0.19	0.14	0.21	1.00				
12	Accumulated patents in the firm	-0.10	0.23	-0.22	0.09	-0.19	-0.05	-0.13	0.06	-0.00	0.06	-0.02	1.00			
13	Firm size	0.04	0.14	-0.08	0.02	-0.16	0.04	-0.02	0.09	0.05	0.10	0.11	0.53	1.00		
14	Type of firm	0.03	0.27	-0.22	0.09	-0.25	0.04	-0.05	0.13	0.07	0.13	0.14	0.53	0.41	1.00	
15	Young firm	0.14	-0.06	0.11	-0.03	-0.01	0.12	0.12	0.06	0.09	0.06	0.15	-0.23	0.15	-0.00	1.00

Table 3: Model I & II: Negative binomial regression models, dependent variable is non-patent literature citations and patent citations – benchmark is Platform patents (earliest invention type)
Model III-VI: OLS regressions the dependent variable is patent value.

	Model I	Model II	Model III	Model IV	Model V	Model VI
	NPL CIT	PAT_CIT	Compound patents		Formulation patents	
Compound patents	-0.462***	0.208***				
	[0.097]	[0.060]				
Utility patents	-0.260**	0.186**				
	[0.108]	[0.082]				
Formulation patents	-0.886***	0.484***				
	[0.141]	[0.073]				
NPL_ RATIO (The total number of non-patent literature citations over total number of backward citations)			-0.482**		0.310	
			[0.233]		[0.248]	
Citations to non- patent literature (dummy)				-0.315**		0.254*
				[0.130]		[0.145]
Backward citations	0.132***	0.076***				
	[0.008]	[0.003]				
Technology scope	-0.001	0.002	0.021***	0.022***	0.102***	0.098***
	[0.003]	[0.002]	[0.004]	[0.004]	[0.036]	[0.036]
Firmsize	0.000***	0.000	0.001***	0.001***	0.000	0.000
	[0.000]	[0.000]	[0.000]	[0.000]	[0.003]	[0.003]
Accumulated patents in the firm	-0.003***	0.001**	-0.002***	-0.002***	-0.003	-0.003
	[0.000]	[0.000]	[0.001]	[0.001]	[0.006]	[0.006]
Type of firm	-0.175**	0.032	0.109	0.104	-0.287	-0.260
	[0.083]	[0.057]	[0.151]	[0.150]	[0.244]	[0.246]
Young firm	-0.072	-0.023	0.034	0.030	0.533	0.494
	[0.068]	[0.064]	[0.180]	[0.178]	[0.364]	[0.363]
Log of patents family size	-0.074*	0.039				
	[0.038]	[0.025]				
Patent grant	-0.100	0.030	0.284	0.300*	0.266	0.248
	[0.086]	[0.098]	[0.175]	[0.175]	[0.312]	[0.307]
Patent withdrawn	-0.062	0.081	-0.644***	-0.638***	-0.644*	-0.615*
	[0.063]	[0.055]	[0.137]	[0.140]	[0.332]	[0.339]
Year of patent application	0.001	-0.009	-0.572***	-0.572***	-0.254	-0.260
	[0.010]	[0.007]	[0.094]	[0.096]	[0.159]	[0.159]
Constant	-1.438	17.508				
	[21.036]	[13.973]				
lnalpha						
Constant	-1.555***	-3.397***	0.948***	1.000***	-0.347	-0.392
	[0.189]	[0.253]	[0.293]	[0.293]	[0.519]	[0.494]
Pseudo LL	-1992775	-2185183				
No of Obs	1058	1058	547	547	110	110
Wald-Chi2	811.5829***	1263.675***				
R-squared			0.408	0.410	0.407	0.414
Adj.R-squared			.3979968	.3998012	.3536971	.3617736
F test			236.6396***	231.9679***	4.871587***	5.23622***

* p<0.1, ** p<0.05, *** p<0.01

Table 3: Coefficient estimates for percentile regression for compound patents

	.1	.2	.3	.4	.5	.6	.7	.8	.9
NPL_ RATIO	-0.147	-0.139**	-0.158**	-0.194*	-0.181	-0.278***	-0.39***	-0.618***	-0.786***
	[0.100]	[0.059]	[0.089]	[0.102]	[0.141]	[0.139]	[0.164]	[0.162]	[0.216]

Table 4 Negative binomial regression models for compound and formulation patents, dependent variables patent value measured by forward citations and family size

	Compound patents				Formulation patents			
	Model VII Forward citations	Model VIII Family size	Model IX Forward citations	Model X Family size	Model XI Forward citations	Model XII Family size	Model XIII Forward citations	Model XIV Family size
NPL_ RATIO (The total number of non-patent literature citations over total number of backward citations)	-0.552**	-0.242*			0.127	0.260		
	[0.267]	[0.130]			[0.300]	[0.213]		
Citations to non patent literature (dummy)			-0.302**	-0.137**			0.178	0.149
			[0.128]	[0.065]			[0.147]	[0.127]
Firm size	0.001*	0.001***	0.001*	0.001***	0.010**	-0.004	0.010**	-0.004
	[0.000]	[0.000]	[0.000]	[0.000]	[0.005]	[0.004]	[0.005]	[0.004]
Accumulated patents in the firm	-0.002***	-0.002***	-0.002***	-0.002***	-0.026*	0.011**	-0.025*	0.011**
	[0.001]	[0.001]	[0.001]	[0.001]	[0.013]	[0.005]	[0.013]	[0.005]
Type of firm	0.229*	0.045	0.227*	0.042	-0.420	-0.039	-0.433	-0.006
	[0.132]	[0.122]	[0.130]	[0.122]	[0.463]	[0.200]	[0.447]	[0.202]
Young firm	-0.039	0.009	-0.046	0.008	0.725***	-0.045	0.690**	-0.068
	[0.157]	[0.105]	[0.158]	[0.105]	[0.270]	[0.196]	[0.276]	[0.196]
Patent grant	-0.099	0.484***	-0.082	0.495***	-0.034	0.653***	-0.066	0.637***
	[0.124]	[0.073]	[0.124]	[0.073]	[0.251]	[0.169]	[0.230]	[0.162]
Patent withdrawn	-0.572***	-0.379***	-0.575***	-0.375***	-0.699**	-0.237	-0.681**	-0.216
	[0.176]	[0.115]	[0.174]	[0.116]	[0.293]	[0.247]	[0.298]	[0.246]
Year of patent application (three periods)	-0.777***	-0.253***	-0.779***	-0.252***	-0.208	-0.359***	-0.227	-0.360***
	[0.114]	[0.070]	[0.114]	[0.070]	[0.216]	[0.130]	[0.210]	[0.126]
Technology scope	0.012***	0.008***	0.013***	0.009***	0.082***	0.020	0.079***	0.018
	[0.004]	[0.002]	[0.004]	[0.002]	[0.021]	[0.013]	[0.020]	[0.013]
Constant	2.469***	2.888***	2.495***	2.895***	0.845*	2.594***	0.832*	2.577***
	[0.328]	[0.176]	[0.326]	[0.186]	[0.479]	[0.443]	[0.494]	[0.413]
lnalpha	-0.109	-1.464***	-0.106	-1.464***	-0.402	-1.310***	-0.414	-1.319***
Constant	[0.098]	[0.121]	[0.103]	[0.122]	[0.270]	[0.151]	[0.273]	[0.143]
Pseudo LL	-1.122.805	-1.759.003	-1.123.403	-1.759.312	-220.107	-337.322	-219.819	-337.091
No of Obs	547	547	547	547	110	110	110	110
Wald-Chi2	868.302***	691.1537***	988.8566***	690.7214***	97.05693***	140.2927***	103.9132***	141.7833***

* p<0.1, ** p<0.05, *** p<0.01

Table 5 : Negative binomial regression models, dependent variable is X-citations non patent related citations and X-citations to patent related citations – benchmark is Platform patents (earliest invention type)

	Model XV	Model XVI	Model XVII	Model XVIII
	XNPL	XNPL	XPAT	XPAT
Compound patents	-0.537***	-0.333**	0.244**	0.366**
	[0.114]	[0.153]	[0.100]	[0.151]
Utility patents	-0.391***	-0.229	0.391***	0.474***
	[0.088]	[0.147]	[0.113]	[0.149]
Formulation patents	-1.109***	-1.226***	0.653***	0.478***
	[0.195]	[0.290]	[0.142]	[0.158]
Backward citations	0.174***		0.118***	
	[0.010]		[0.008]	
	----------Same controls as in Table 3----------			
Constant	29.904	43.429	16.806	43.635
Lnalpha	[50.419]	[72.288]	[30.309]	[53.659]
Constant	-0.276**	0.808***	-0.880***	-0.020
	[0.129]	[0.106]	[0.125]	[0.105]
Pseudo LL	-1.489.646	-1.715.555	-1.930.530	-2.127.699
No of Obs	1058	1058	1058	1058
Wald-Chi2	693.87***	132.9567***	817.5432***	174.7439***

* p<0.1, ** p<0.05, *** p<0.01