Paper to be presented at the

35th DRUID Celebration Conference 2013, Barcelona, Spain, June 17-19

HOW SEARCH IN SCIENCE IMPACTS THE VALUE OF INVENTIONS AT EARLY VERSUS LATE STAGES IN THE R&D CYCLE

Karin Beukel
Copenhagen Business School
Department of Innovation and Organizational Economics
kb.ino@cbs.dk

Finn Valentin
CBS
Department of Innovation and Organizational Economics
fv.ino@cbs.dk

Abstract

Search in science is one way to support knowledge creation in firms. In this paper we examine the conditions under which search in science increases the value of individual inventions. Statistical findings from TOBIT regressions show that the cost-benefit tradeoff of search in science depends on the whether the search is aimed at problem solving at early or late R&D stages. If the search is done at early stages of R&D there is a penalizing effect of searching to a higher degree in science, whereas search in science at later stages of R&D contributes to value. The findings of this paper confirm and extend research on search in science during R&D.

Jelcodes:O32,Z0

HOW SEARCH IN SCIENCE IMPACTS THE VALUE OF INVENTIONS AT EARLY VERSUS LATE STAGES IN THE R&D CYCLE

ABSTRACT

Search in science is one way to support knowledge creation in firms. In this paper we examine the conditions under which search in science increases the value of individual inventions. Statistical findings from TOBIT regressions show that the cost-benefit tradeoff of search in science depends on the whether the search is aimed at problem solving at early or late R&D stages. If the search is done at early stages of R&D there is a penalizing effect of searching to a higher degree in science, whereas search in science at later stages of R&D contributes to value. The findings of this paper confirm and extend research on search in science during R&D.

INTRODUCTION

The literature on organizational learning shows that the presence of both explorative and exploitative search is imperative for firm performance (March 1991; Tushman and Oreilly 1996; Brown and Duguid 2001; Katila and Ahuja 2002; Benner and Tushman 2003; Gupta, Smith et al. 2006). While extensive research at the level of the organization has been conducted, research at lower levels, such as the level of the individual invention, is still absent, even though such studies are expected to be relevant for higher levels, such as the organization (Gupta and Shalley, 2006). At the level of the individual invention we would expect a different pattern of search than what takes place at the organizational level. In an R&D cycle where the final product is the result of multiple cumulative inventions, inventions can occur at both early and late stages, and depending on the stage of the R&D, the nature of the search process will differ (Iansiti, 1997; Kline and Rosenberg, 1986). Early in the R&D cycle inventions rely heavily on search in science, whereas later inventions are based on search in technologies in an effort to make the invention stable and ready for the market. Given these fundamental differences in the underpinning search approaches we theorize that the situations therefore early in R&D versus late in R&D are based on distinct different cognitive maps. Cognitive mapping originates in psychology literature (Tolman 1948) and is an expression for which elements that has been taken into account when considering a situation dependent on its surrounding environment. Cognitive maps are distinct in the sense that they are guided by the lens of the environment that is identified, from which a situation (or a problem) is understood. Therefore the cognitive map is an expression for the distinct knowledge sources a situation has been evaluated from. Applying a cognitive map guided by the lens of science provides answers to fundamental questions of ‘why an effect appears’, while a cognitive map guided by a lens of technological application provides answers as to ‘how an effect appear’.

In an innovation process the objective of the firm is to transform an invention from a scientific discovery answering a fundamental question into a commercial invention. To achieve this, the inventor(s) needs to explore the technological application of the scientific discovery, seeking to answer 'how an effect appeared', to succeed in this search process, the cognitive map guiding the lens must rely on technological application rather than science. On this background, it would be expected that search early in R&D will benefit from applying a lens in which the cognitive map is based on technological application and that there will be a penalizing effect when early inventions have been created based on a cognitive map relying too heavily on science. On the other hand, search in late R&D is guided by a technological based cognitive map, in which inventions have become more predictable and closer to market. At this stage we expect the situation to be very different. After working for years and creating sub-inventions all along, firms could increasingly be caught in familiarity traps (Leonard-Barton 1992; Ahuja and Lampert 2001; Tuahene-Gima 2005), and at this stage, search in science represents a powerful vehicle for their circumvention (Nelson 1982; Fleming 2001). Why presence of cognitive map based on science at this latter part of the innovation process will benefit the value of the individual invention. In essence our argumentation follow the theory on cognitive variety proposed in psychological literature (Lachman, Lachman et al. 1979; Barber 1988) and also argued for being essential for organizations in creating value in dynamic environments in recent management literature (Eisenhardt, Furr et al. 2010). Examining whether cognitive variety in early versus late stages of R&D requires a setup in which "footprints" of the cognitive mapping of both search in science and search in technology can be identified, as well as the possibility of identifying types of inventions created at either early or late stages of R&D. The footprints of both science and technology have been examined in previous literatures using backward citations to the patents held, we argue that these footprints also expresses to which degree each of the different cognitive

maps, science versus technologies have been present in the development of the invention. A cognitive map based on science is expressed in the footprints of science in non patent related backward citations (NPR citations) from patents (Dietmar Harhoff, Frederic M Scherer et al. 2003; Fleming and Sorenson 2004; Cassiman, Veugelers et al. 2008), and the cognitive map based on technologies is expressed in the footprints of technology in backward citations to prior patents (PAT CIT) (Lanjouw J.O. and Schankerman M. 1997; Dietmar Harhoff, Frederic M Scherer et al. 2003). However, identifying individual inventions at early versus late stages of R&D is more challenging and has yet to be done in a large empirical study. To identify individual inventions in 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 overcome this by choosing an industry, namely, the biotech industry, in which the types of inventions appearing in R&D have been identified in the literature. The following types of inventions occur: platform patents, compound patents, utility patents, delivery patents, formulation patents, and process patents (Gupta and Bansal 2002; Kaushal and Garg 2003; Yoo, Ramanathan et al. 2005; Norman 2007). Thereafter, by developing a unique algorithm based on semantic structures and IPC codes, we identify these different invention types in a large number of patent observations. We can thus test our hypotheses in an empirical study utilizing unusually detailed data of 1058 biotech patents.

Our findings add several insights to management of innovation literature. For the first time, we empirically investigate and demonstrate how cognitive maps guiding search at early versus late stages over the R&D cycle influences performance; our results broadly support our expectations that search in science gradually becomes less predominant across the R&D cycle and that linkages to prior technologies become significantly more prevalent during late exploitative

processes. Using this as a stepping stone for further analysis, we identify two types of inventions and do separate tests on how the orientation in search toward science affects the value of inventions at early versus late stages in the R&D cycle. The findings support our contentions; in early explorative R&D processes, cognitive mapping based on technologies are infrequent in comparison to late R&D processes, but a one-sided search in problem solving from science has a penalizing effect on the invention value only. At the opposite end, at late stage R&D, search in science occurs at a lower frequency but is nevertheless associated with an increase in invention value. This means that establishing search based on a cognitive variety in lens applied has a positive effect at both ends of the R&D cycle.

The remainder of the paper is organized as follows: In the first section, we discuss the different invention types that appear during R&D and motivate the definitions used in our study. We then discuss the mechanisms of the underlying cognitive mapping of search in both early and late R&D processes. This is followed by our hypotheses regarding the gradual decrease in search in science and increase in search in technology when moving from early to late R&D, and how cognitive variety in the search approach affects the value of the resulting innovation. Our presentation of data and methods follows, as well as our empirical findings. In the final section we discuss our results and the implications, limitations and further potential extensions of the research.

THE CHARACTERISTICS OF PATENTED INVENTIONS RESULTING FROM R&D

For our study we need information on the following: a) characteristics of the constituent inventions coming out of the R&D processes; b) how these constituent inventions fall into recurrently appearing types, referable to either the early explorative stages of R&D or later

exploitative stages; c) characteristics of each invention regarding reliance on science versus reliance on search in prior technologies; and e) the value of each constituent invention. The biotech industry and its inventions meet these empirical requirements.

In the biotech industry, different types of patented inventions are created during distinct stages of discovery and development of drug candidates. Firms in the biotech industry rely on patents, and not to patent is a less likely scenario (Levin, Klevorick et al. 1987). All patent types are created as part of an overall patent strategy in order to appropriate returns from the invention (Granstrand 1999). We identify six patent types as relevant for biopharmaceutical inventions. They appear during different phases of the R&D process that are either distinctly early or late**:** 1) patents protecting the method of identifying new products. In the biotech industry, these would be platform patents (Yoo, Ramanathan et al. 2005) 2), 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, this would mean composition of matter/compound patents (Norman 2007); 3) patents protecting the specific application of the core structure of the product. In the biotech industry, these would be formulation patents (Norman 2007); 4) patents protecting the methods of utility. In the biotech industry, these would be utility 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)). The different patented inventions have a certain form and focus, contingent on the phase of R&D the invention process has reached; the early processes are characterized by an explorative approach whereas the latter are dominated by exploitation. 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 about here

To further explain the cumulative relationship between these constituent inventions we note that the initial patents of a drug discovery process typically protect a method of identifying products (platform patents). Nightingale (2000) explains how the development of new technologies enabled the pharma industry and especially the biotech industry to conduct large-scale experimentation. For example, Amgen developed methods to span the boundary between the development of automation/information systems and biology and chemistry research (Hamilton, Armstrong et al. 1996), which enabled the company to identify targets much more efficiently.

Basic product patents (compound patents), which protect the core structure of a product, build in most cases on efforts made by utilizing the platform technology, but can also be a result of other contributing factors. Later in the R&D process, the specific application of the invention is patented in the form of formulation patents. These have by definition a strong relationship to compound patents (Norman 2007), and in this sense build strongly on exploitative learning processes. Moreover, formulation patents are more exploitative in nature, not only because they build on insights previously acquired from the same drug candidate, relating to its composition of matter and its utility patents (MPA 2012).

Process patents are in most cases developed subsequently to the invention types identified above. Both instrument and utility patents tend to be invented subsequent to compound patents. These patent types are not as dependent on prior inventions. Utility, instrument, and process patents can to some extent be cumulative to either compound or formulation patents, but this is not necessarily the case. These types of patents do not protect the core of a product, but rather the

surrounding inventions. Below, we will outline how this relates to problem solving in science or technology.

Search in science in early R&D versus search in technology during late R&D

Several contributions have added to our understanding of search in science versus search in technology at early versus late R&D. Since its infancy, the innovation literature has recognized the presence of and the iterations between both sources of novelty: science and technology (Schumpeter 1939; Gibbons and Johnston 1974; Kline and Rosenberg 1986). Prior research has shown that the extent to which search is done in science is industry dependent (Cohen et al, 2002; Klevorick et al, 1995 ; Nelson, 2003; Stankiewicz, 2000; Pavitt, 1991). We know that only a handful of sectors, including the biotech and life sciences sectors, undertake systematic research based on close correspondence and integration with global scientific advances (Cohen et al, 2002; Klevorick et al, 1995), as science only delivers this support from a few of its disciplines (Nelson, 2003; Stankiewicz, 2000). However, where effective support from science to knowledge formation is available, industries do their upmost to reap the full benefits for their R&D (Pavitt, 1991). For these truly science-based industries we therefore expect that early inventions will have a strong orientation toward scientific research, as these inventions will be based on promising outcomes from front-end scientific research. As R&D work with research-based inventions progresses, it shifts to issues that to a larger extent are technological in nature (Iansiti, 1997; Kline and Rosenberg, 1986). Case studies of the R&D cycle in science-based firms demonstrate this shift in emphasis from early science-based discovery toward subsequent stages of development with increasing emphasis on technological knowledge (Cassiman et al, 2010; Iansiti and West, 1999).

Therefore, following a long tradition in the study of innovation (Schumpeter 1939) we conceptualize innovation as a process of combining existing components, recognizing that science-based innovations come about as a result of an accumulation of multiple "sub-inventions". We propose that during this cumulative process a shift occurs in the orientations of search toward science versus earlier technology. In science-based innovation the early discovery stage is based on explorative research, and hence search in these initial sub-inventions relies heavily on science. If these initial inventions are to be transformed into reliable technologies, they must be made stable and predictable within a range of relevant contexts, and the methods for their production must be developed. Search in these stages thus shifts the emphasis to the technological realm, and search and solutions become more explicitly relatable to previous technologies. Based on this understanding, and our expectation of identifying the footprints of science and technologies in backward citations in patents, we propose that:

> Hyp. 1: Patents protecting inventions from early stage R&D have more citations in absolute numbers to scientific literature than do patents protecting inventions from later stages of R&D

> Hyp. 2: Patents protecting inventions from late stage R&D have more citations in absolute numbers to prior technologies than do patents protecting inventions from early stages of R&D

The impact on patent value of search in science at early versus late stages of the R&D cycle

Hypotheses 1 and 2 are concerned with differences in orientation toward science as opposed to technology at early and late stages of R&D. It was argued that in science-based industries, explorative R&D is more strongly associated with search in science whereas later stage R&D is

more strongly associated with search in technology. In our overall argument these two hypotheses are stepping stones leading to the key question concerned with the impact of science on value at either of the two ends of the R&D cycle: early or late R&D.

As noted previously, the main mechanism that plays a role in early versus late R&D inventions is the cognitive mapping applied in either of the two distinct different search processes. At early R&D stages the input in the process the firm has is a scientific discovery, and the aim of working with this scientific discovery is to mature the discovery and make it an invention, which is both technical stable and industrial applicable. Patents from early R&D stages are therefore, as all ready in earlier chapters explained, based on a high degree of search in science, which also means that it is a science based cognitive map that has been guiding the problem solving. Such science based problem solving results in questions and answers concerning 'why the effect appear' to be solved. This type of cognitive mapping is in line with how Gavetti and Levinthal (2000) define the forward looking cognitive approach: a cognitive map coding elements about the linkage between choice of action and subsequent impact. On the other hand problem solving utilizing a technology based cognitive map provide answers to fundamental questions of 'how the effect appears in different settings'. Gavetti and Levinthal (2000) argues that such search is backward looking and exploitative in its use of existing knowledge. An example used by Yayavaram and Ahuja (2008) to explain the differences in the two distinct cognitive mapping approaches is the discovery of high-temperature superconductivity in copper-oxide based materials. This discovery led inventors to search for the theoretical underlying model of high-temperature superconductivity to provide answers to fundamental science based question of 'why this effect appeared'. On the other hand, and very differently, an experiential approach could also be used to gain alternative and more exploitative insights. Such a cognitive approach could be done by

experimenting with other related materials, in order to answer questions of ‘how the effect appears in different settings’. Whereas the first science based cognitive map is significantly important for basic research and for understanding the underlying mechanism of high-temperature superconductivity in copper-oxide based materials, the latter approach, the technology based cognitive mapping can help stabilize and enable industrial applicability of the scientific discovery into a readymade product for the market, but might not give new scientific insights on a choice of action and a subsequent impact. Therefore, search in science and search in technology is two distinct different cognitive maps guiding problem solving processes into different direction. According to cognitive psychologists cognitive mapping is a precondition to enable any perceiving and thinking in order to come up with suggestions of an orientation to take, taking a current situation in relation to the environment into account (Neisser 1976; Piaget 1985; Carroll 1993). The mechanism to consider when both search in science and search in technologies are present in generating a new innovation is present, is therefore a search conducted based on cognitive variety. In management research cognitive variety has shown to create value due to contradictions arising from applying these significant different mental models (Brown and Eisenhardt 1997; Smith and Tushman 2005), recent literature also emphasize that value might not only be driven by contradictions but also because cognitive variety enables flexible recombination of the individual cognitive map (Eisenhardt, Furr et.al 2010). Therefore if search in science and technology represents two distinct different cognitive maps, search based at a cognitive variety at either early or late R&D phases represents two different situations, in early stages search is based on a science based cognitive map and benefits will therefore arise if also exploring a technology based cognitive map guiding a different set of problem solving during this early process, whereas in later stages of R&D search is based on a technology based cognitive map, where adding elements of a science based cognitive map will be beneficial.

Therefore, in cases where we identify developments coming out of firms whose main purpose is to reach the market with a product, the firm has already considered the benefits it can reap from the invention and is to a much higher degree concerned with stabilizing and enabling an industrial application, eventually transforming it further from scientific discovery into a drug to introduce on the market. Whereas the initial transformation of the scientific discovery into a patent required a science based cognitive map in which forward looking elements in the problem solving answered relationships concerning the choice of action and the subsequent outcome, value to a larger degree will be present if also subsequent problem solving, including essential parts of a stabilization and application of the invention is included in the invention at this stage. For this type of process a technology based cognitive map is required. We therefore argue, in line with literature on cognitive variety, that synergies will arise when search based on a cognitive map using the lens of prior technologies replaces searching in science in the early R&D stages, and inventions based on cognitive variety will be more likely to generate value for the subsequent invention.

> Hyp 3: Inventions from early stage R&D increase in value the more a cognitive map based on technologies replaces the overemphasized cognitive map based on science

Whereas the early stages of R&D poses one distinct situation in terms of cognitive maps guiding problem solving the situations at the later stages of R&D is different. At his point in time the input is not a scientific invention, but a ‘well-developed’ invention, which has been worked on for years and years (research shows that a classic drug development process takes between 8-12 years), the aim at the later processes of R&D is to specify the invention further to ease the process through FDA, to ensure no or low side effects, as well to prove the drug to be more efficient than the drugs available, as well as to create new patent protection to extent patent

protection lifetime of the drug if at all possible. In this setting the experiential knowledge accumulated by firms is based on problem-solving through the lens of a technology based cognitive map. For this reason the risks of falling into familiarity traps (Ahuja and Lampert 2001) are particularly acute at this end of the cycle, and finding an effective balance of different mental models applying cognitive variety largely becomes a question of securing inflows of new information and knowledge to avoid these traps (Leonard-Barton 1992; Tuahene-Gima 2005). A science based cognitive map represents a powerful vehicle for the circumvention of familiarity traps (Fleming 2001), we therefore propose, in line with the arguments of cognitive variety, that at the late stages of R&D replacing the technology based cognitive map with elements of a science based cognitive map will enhance the value of inventions.

Hyp 4: Inventions from late stage R&D increase in value if a science based cognitive map has supported the technology based cognitive map during problem solving

DATA AND METHODS

Sample and Methods

For three related reasons we have selected the Scandinavian drug discovery firms as our sample: first, the patents applied for by this subset of biotech firms are related to drug discovery, enabling us to identify types of inventions and their corresponding patents. Second, drug discovery firms are dependent on patents as a method of appropriation (Levin, Klevorick et al. 1987), rendering strategic choices on patenting a focal issue of R&D in these firms. Third, the types of invention and the related patent types are identified in the literature, providing a solid starting point for our methodology. The final dataset consists of 1076 patents applied by 110 drug discovery firms originating from Scandinavia and established between 1987 and 2003.

Identifying inventions in early explorative vs. late exploitative R&D processes

This paper uses a proprietary text mining method to identify seven patent types in Derwant patent abstracts, using semantic structures, keywords, and IPC codes. The text-mining method was developed through 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 of the steps is described further below.

Generating a text-mining algorithm

Initially, a literature review of patent types in the biotech industry was conducted. As the biotech industry was developed out of the pharmaceutical industry (see Hopkins (2007) for a review), the patent types utilized are closely related in the two industries. Norman (2007) identifies four patent types in the drug development industry: 1) compound patents (also termed composition of matter patents), 2) formulation patents, 3) utility patents, and 4) process patents. Additionally, two other patent types have been mentioned relating to biotech drug discovery: 1) platform patents, protecting *"industrialised high*-throughput screening (HTS) platforms. These platforms together with most notably combinatorial chemistry and new informatics systems enabled the generation and rapid screening of very large chemical libraries against the larger number of new *targets"*p.571 (Hopkins, Martin et al. 2007) and 2) instrument patents. Instruments refer to different ways of delivering drugs, one example being osmotic. These systems form a major segment of drug delivery products, a different invention type (Verma 2000), protected by different types of patents; namely instrument/drug delivery patents (Gupta and Bansal 2002; Kaushal and Garg 2003).

After identifying the six patent types, we interviewed industry stakeholders to ensure that our identifications of patent types were in accordance with the industry approach. Then we developed the actual text mining method. Initially, we randomly selected 85 biotech patents and categorized

them “manually” into the six types. Through recurrent iterations we identified differentiating semantic structures (keywords and sentence pieces) and IPC codes. A range of prior biotech patent related studies helped us identify the keywords and IPC codes and thereby categorize the patents: the OECD methodology for identifying biotechnology patents (OECD 2005), the results of the biotechnology comparative study on patent rights done by EPO, USPTO, and JPO (EPO, USPTO et al. 1998), patent search literature focused on bioscience (Yoo, Ramanathan et al. 2005), and Dirnberger’s (2011) case on 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). Furthermore, a total of 255 IPC (International Patent Classification) codes were identified as belonging to a patent type: Compound (98), Formulation (10), Instrument (29), Platform (28), Process (87), and Utility (1934).). Each patent was then examined to see whether it matched any of the 608 semantic structures and 255 IPC codes and given weighed scores if matching occurred.

Testing the automatic text-mining method

The text mining method of patent type categorization described above was then put into an algorithm to make it possible for the patent type categorization to be handled automatically by a script. To test whether the machine method categorization (the algorithm) of patent types equaled a manual categorization, two tests were performed. First, a test of 1079 patents from 107 biotech firms was conducted. The 1079 patents were read and categorized manually into patent types and the results were then compared to the results obtained from the automatic machine coding. Results show that 91% were classified identically. Second, a test of a large biopharmaceutical company, Novo Nordisk, was performed. Five percent of their patent portfolio was randomly

selected and manually classified; 92% of these patents were assigned the same categorization as the automatic categorization.

Validating the text-mining method with industry players

To ensure that the results of the automatic text mining method corresponded to what the industry players 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 of the total of 11 patents. In biotech firm B, the firm's IP counsel's categorization matched the automatic categorization in 9 of the total of 13 patents.

In the large pharmaceutical firm, 50 patents – one sixth of the total patent portfolio – were randomly chosen for the test; of these, 35 patents were categorized in the same way as the automatic categorization. Results show a satisfactory output from the machine patent type categorization.

Variables

Identifying valid proxies for search behavior and the underlying cognitive maps of firms in the process of creating new inventions is difficult; as a result, management researchers have relied on observing the innovative outcome of a given search behavior, patents, and related the search process to that of the backward citations inserted in the patents (Benner and Tushman 2002; Katila and Ahuja 2002; Laursen, Leone et al. 2010; Phelps 2010; Laursen 2011). Backward citations have some elements of noise as proxy for search behavior during innovation (Alcacer and Gittelman 2006; Roach and Cohen 2012). However, in line with recent literature, we refrain from going into the details of the individual backward citations, using them instead in count and ratio measures. There are two distinct types of backward citations : a) a non patent related citation

(in the chemical and pharma industry, these are most often citations to scientific journals, in our robustness checks we examine this closer), which we use as a proxy for the underlying cognitive map guiding the search, namely the science based cognitive map and b) a backward citation to a prior patent (PAT_CIT), which we identify as a proxy for a technology based cognitive map guiding the search.

In our estimations in Tables 3 and 4, we use different types of backward citations as dependent variables: NPR_CIT is the number of backward citations to non patent related literature, PAT_CIT is the number of backward citations to prior patents.

The main independent variables in Tables 3 & 4 are dummy variables that are coded 1 for each of the four patent types:

1) Platform Patents Category: This category includes methods for analyzing and selecting compounds and entities (e.g. nuclear receptors, phages); methods for identifying molecules (e.g. chemical compounds, peptides, ligands, modulators, prodrugs); the generation of new libraries (e.g. phage display library) or other assays (e.g. sandwich assay); construction and utilization of new plasmids, constructs, expression vectors or cassettes, and retroviral or recombinant vectors involved in specialized tasks; specialized techniques for generating transgenic, congenic organisms, or creating new mutants or cells; techniques for generating recombinant proteins or viruses; and other technologies, including site-directed mutagenesis, hybridoma, and phage-display library.

2) Compound Patents Category: This category includes new compounds, molecules, proteins, peptides, enzymes, receptors, derivatives, analogues, and variants; new finished products,

cosmetic compositions, and skin-treating preparations; and mixtures and preparations giving known components a novel usage.

3) Utility Patents Category: This category includes 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: This category includes patents describing new dosage forms, pharmaceutical preparations; immunogenic components, diluents, aqueous solutions or other forms of formulations such as; carriers, synthetic membrane vesicles; and vehicles for controlled delivery of active substances or sublingual compositions or for parietal administration.

Instrument and process patents were removed from the sample, as they each account for less than 50 patents.

In Tables 5, 6, and 7, the patent value is a dependent variable (PAT_VAL). We identify the 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, which should be carefully dealt with. One solution to the one-sided approach to valuing patents in the literature through the use of forward citations to patents could be further improved by also assessing patent value by several estimators (Lanjouw and Schankerman 1999). We therefore apply a measure that incorporates two distinct patent value correlates: family size and forward citations. In Lanjouw and Schankerman (1999) these two measures are the only two presented as positive and significant for patents in the drug industry, when regressed as determining patent litigation and patent renewal, and also later identified as

measures shown to be the better indicators in the large portion of value indicator empirical studies have suggested, shown to be less noisy (Gambardella, Harhoff et al. 2008). Furthermore, the reason for choosing these two dimensions is that they contribute to the 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 to a firm. Firms apply for patents in a range of countries depending on how core the patent is: if the patent is non-core, the firm will apply in a small number of countries; if it is core, the firm will apply in a high number of countries. The correlate of forward citation is measured by the total number of forward citations the patent has received, a proxy for how important subsequent inventions within the area have found the patent to be. The two proxies have low correlation (.2216). We standardize each proxy to ensure an equal balance of value. The independent variable (NPRcit_RATIO) in Tables 5 and 6 estimates to what degree the emphasized science based cognitive map, has been replaced with technology based cognitive map, linked to our hypothesis three on cognitive variety. NPRcit_RATIO is the number of NPR citations divided by the total number of backward citations.

In Table 7, we investigate the effect of applying a science based cognitive map in problem solving for inventions coming out at later R&D stages (formulation patents), where technology based cognitive maps are overrepresented. To do so we use a dummy variable as an independent variable, which takes 1 when the invention has one or more NPR citation(s) and 0 when no NPR citations are included in the patent. In this way the variable indicates whether a search in science has been undertaken.

Estimation Approach and Control Variables

To understand how variations of types of inventions (Pat_type_P/ C/U/F) affect the distribution of NPR (XNPR cits) and PAT CITs (XPAT cits) we estimate the following models (Hypotheses 1 and 2):

$$Y_{pft} = \alpha + \beta_1(Pattype) + \beta_2(Pattype) + \beta_3(Pattype) + \beta_4(Pattype) + \delta_{pft}$$

The outcome variable Y is a count of either NPR or PAT CIT, the unit of analysis is patent (p) applied by firm (f) at a given time (t). As we are concerned with NPR and PAT CITS being influenced by other possible patents, firms and time variables, we control for a number of variables presented by δ_{pft}. As patent level controls we include the total number of backward citations (BW_CIT), the number of IPC codes (TECH_SCOPE), the log of the family size of the patent (LOG_FAM_SIZE), whether the patent has been granted (GRANT), and whether the patent has been withdrawn (WITHDRAW). As firm-level control variables we include the number of employees in the firm the year a patent is applied for (FIRMSIZE), the number of accumulated patents the firm has applied for before the year of patent application (AAC_PAT), whether the firm is focused on small or large molecules (FIRM_TYPE), and whether the firm is a young firm (less than four years old) at the time of the patent application (AGE_DUM). For the time variable, we control for the year of the patent application (YEAR). For the estimation mode, we first 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 is over dispersed. This suggests that a negative binomial regression would be a better choice than a Poisson regression (Wooldridge 2009). 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 scientific or technological based cognitive mapping in problem solving affects the PATVALUE, we estimate a second model, Hypothesis 3, for early R&D inventions. The NPR_cit_Ratio is indirectly also a measure of technological impact as the citations that are not to NPRs will be to prior patents (PAT_CIT):

$$3)\ PATVAL_{pft} = \alpha + \beta_1(NPR\_RATIO) + \delta_{pft}$$

We also estimate a third model, Hypothesis 4, for late R&D stage inventions:

$$4)\ PATVAL_{pft} = \alpha + \beta_1(NPR\_DUM) + \delta_{pft}$$

The outcome variable in model 3 and 4 is patent value and the unit of analysis is patent (p) applied by firm (f) at a given time (t). This model is used for each of the types of inventions; however, only selected results for certain types of either early or late stage R&D inventions are presented. Again, we are concerned with the influence of firm and time on PATVAL and employ a number of variables to control for this, presented by δ_{pft}, and cluster for the effects of the single firm behind the invention. In these models we divide the time period into three: patents before 2001, patents after 2002 but before 2004, and patents after 2004. As our estimation mode we use TOBIT, as the dependent is lower, limited by zero (Wooldridge 2009). As interaction effects in non-linear models are troublesome (Ai and Norton 2003) we choose to investigate each type of patent separately, and are extra careful in interpreting β results. Also in these models we apply robust estimators to avoid heteroskedasticity, and perform VIF tests to check for multicollinearity.

The TOBIT regression is mean centered; as a result, it might not capture differences in the effect at different locations in the patent value distribution and especially might not be able to explain the extreme observations in the tails. To explore the degree of 'extreme tails' we therefore follow

the work of Koenker (Koenker and Bassett 1978; Koenker 2005) and use percentile regressions to investigate the effect at different locations of the patent value distribution.

RESULTS

Descriptive results

Table 1 provides descriptive data for all variables.

Insert Table 1 about here

Several interesting features of our sample emerge from this table. From our sample we can see that certain patent types occur frequently while others occur infrequently. Compound stands out by accounting for 52% of the total sample, the other patent types are also unevenly distributed; Platform accounts for 22%, utility, 15% and formulation, 10%. There are several explanations for this uneven distribution. The patent types taken out early in the R&D process, platform patents and finally, compound patents are the most frequent as these invention types are the ones that occur first in R&D projects. Technological change is highly uncertain (Macher 2006) and consequently, R&D projects may be closed before the innovation process addresses the main issues of formulation, instrument, utility and/or process patents. Additionally, compound patents may also be more attractive, since they protect a core technology and are, especially if it is not possible to obtain formulation patents, the main protection for a drug candidate. According to Norman (2007) *"patents claiming products and product derivatives provide the greatest scope* for enhancing patent protection. Follow on applications, or applications claiming enantiomers, salts, solvates, crystalline forms or prodrugs, are all valuable in effecting patent protection *strategies for key products"* (p.12). The reason for platform patents being outnumbered by compound patents is also straightforward: a single platform is capable of generating many

targets, which can then be developed into inventions protected by compound patents, resulting in one enabling technology leading to many potential compound patents.

Table 2 presents correlation matrix variables. Few correlations are significant; unexpectedly, none approaches a common 0.70 criterion for multicollinearity. However, as expected, there is multicollinearity between NPR_CIT and X_NPR citations, PAT_CIT and X_PAT_CIT as well as BWCIT and both NPR_CIT and PAT_CIT.

Insert Table 2 about here

Knowledge sources in inventions created in early R&D processes

Our empirical results are summarized in Tables 3-7. We first test the presence of a high number of scientific citations (NPRs) in inventions created early in R&D processes (Hypothesis 1). Negative significant values for all patent types other than platform patents in Model 1 reflect the fact that platform patents cite NPR significantly more than all other types. To further explore the relationship between the different types of inventions with regard to problem solving in NPR during early R&D phases, we also conduct Wald tests for compound patents, comparing them to formulation patents. Compound patents are composition of matter patents, meaning a core patent type that protects the core substance of a drug. Without compound patent protection a firm's appropriation opportunities of a new drug would be affected negatively. Significant Wald tests indicate that compound patents cite NPR significantly more than formulation patents (chi(2)=53.85***), confirming the hypothesized relationship.

Insert Table 3 about here

Knowledge sources in inventions created in late R&D processes

For inventions created during late R&D we expect a high number of citations to prior technologies (PAT CITs), as stated in Hypothesis 2. The negative significant values presented in Table 4 for all other patent types than formulation patents reflects the fact that formulation patents cite patents significantly more than inventions created earlier in the R&D process, which confirms Hypothesis 2.

Insert Table 4 about here

We thereby confirm for the first time to our knowledge the general underlying assumption in management of innovation literature that early R&D processes are conducted mainly by relying on scientific linkages and less on prior technologies, whereas late R&D processes generate inventions on the basis of linkages to prior technologies and less on the basis of scientific linkages as compared to earlier inventions. We thereby also emphasize that the number of either scientific citations (NPR) or citations to prior technologies (PAT CIT) is an indicator of whether an invention has been created during early explorative R&D processes or late exploitative R&D processes.

Patent value and knowledge sources in exploitative and explorative R&D processes

The remainder of our empirical analysis focuses on how the application of science based cognitive map versus a technology based cognitive map differs for early and late inventions. We present the results for compound patents as early inventions and formulation patents as late inventions. For inventions created in early R&D, the distribution of citations is as follows: On average a compound has 7.28 backward citations, of which on average 37% (2.7 citations) are citations to science (see Table 1 for descriptive statistics on compound patents). This means that

compound patents cite prior technologies to a much greater extent than science. In our first estimations of the value drivers behind explorative inventions, we restrict our analysis to express when science is cited to a high degree (a ratio of NPR over total backward citations). The results presented in Table 5 below use an aggregation of patent value measures as a dependent variable with lower censored observations and are therefore estimated using the Tobit technique. The empirical results for compound patents (early inventions) show that any increase in the share of citations to NPR has a penalizing effect.

Insert Table 5 about here

As the TOBIT regression performed might not capture a difference in the effect at different locations in the patent value distribution, and especially might not be able to explain the extreme observations located in the tails, we use percentile regressions as well as a Wald test between the different parts of the distribution to explore how the upper end extreme tails differ from an average search in science. In Table 6 and Figure 3 we present the results of the percentile regressions for NPRCit_RATIO for compound patents. The results show significant differences as to the end of the tails. When the search in science is increased by a percentile, the coefficient estimates of the independent variable NPRCit_RATIO decrease for every percentile until 0.5. The effects are not significantly different from each other (around -.13 to -.18); however, after 0.5 the penalizing effect increases for every percentile, being four times as negative if the search is extremely focused at 0.78 than at any level from 0.1 to 0.5. 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**). This means that patent value declines with increasing NPR_Cit RATIO and particularly steeply when the ratio is above 60%. This in turn means that a predominance of NPR citations, a science based cognitive map has a strong negative effect on early inventions. We thereby show

that inventions from early stage R&D benefit from a mix of science based and technology based cognitive maps, cognitive variety, and their value decreases the more search is based on a science based cognitive map only, thereby confirming Hypothesis 3.

Insert Table 6 about here

In the last part of our empirical analysis we focus on how the application of science affects the value of inventions created during exploitative learning processes (formulation patents). Formulation patents (late R&D inventions) have on average six backward citations, of which only one fourth are to science and three fourths are to earlier technology. For inventions in late R&D, benefits of cognitive variety are expected; the results for this are presented in Table 7. Using the sample of 110 exploitative inventions (formulation patents) we regress NPR_DUM (which takes 1 when there are one or more citations to NPR and 0 when there are no NPR citations) on PAT_VALUE and find that the results support our hypothesis: NPR_DUM has a positive statistically significant coefficient. Our results confirm Hypothesis 4, stating that inventions from late stage, exploitation-oriented R&D increase in value with the addition of elements of explorative search.

Insert Table 7 about here

ROBUSTNESS CHECKS

The results estimated in Models 1,3,5,6 and 7 support the proposed hypotheses. However, there are several robustness checks that need to be conducted.

For hypothesis 1 and 2 we conduct several robustness checks, to understand the magnitude of our results. First we descriptively investigate whether the results only are consistent in absolute numbers or if the relative numbers also change, in Figure 2 we present a graph showing the four patent types explaining the change over time, which indicates that not only are the absolute focus of either a science based cognitive map or a technology based cognitive map changing over the R&D cycle, also the relative share shifts on average. In addition we estimate the relative pattern of citations by employing a dependent variable taking the ratio of NPL/PAT_CIT (see Table 8, Model 7) and confirm our results.

Insert Table 8, Table 9 and Figure 2 about here

In addition we investigate whether the relationship also is present when investigating only the X citations. X citations are *"*Particularly relevant documents when taken alone (a claimed invention cannot be considered novel or cannot be considered to involve an inventive step*)"* (Webb, Dernis et al. 2005). They are thus citations that challenge the scope of the invention, and might be a more precise measure of search, as these X citations are less noisy than the total as they only include the most important citations. In Table 3, Model 2 we present the results, here the dependent variable is X_NPR_CIT which is the number of X citations to NPR literature, and in Table 4 Model 4, we present the results for X_PAT_CIT which are the citations to patents made as X citations. These tests consistently support our main results, with only few differences, as the coefficients even decrease slightly, indicating an even stronger negative effect. Further to this we present the relative pattern of X citations and use a ratio X_NPL_CIT divided by X_PAT_CIT and confirm the results.

In models for hypothesis 3 & 4 we used a non-conventional value measure to respond to the challenges indicated with one-sided approach of valuing patents with only one value indicator. To test our results in the more conventional manner we estimate our models by splitting the dependent variable into two parts: forward citations and family size. We use a negative binomial regression model, in regards to hypothesis 3 we find strong support when utilizing the more conventional approaches to valuing patents: In the first test we use forward citations as dependent variable (see Table 10, Model 8) and in the second robustness check family size as dependent variable (See Table 11, Model 9). Both results are supportive, however, the coefficient and significance levels are different for NPR_Cit_RATIO being -0.552** when Forward citations is used as dependent variable, and -0.242* when Family size is used as dependent variable, supporting that neither indicator are different from the results presented in the main results.
In the robustness checks for hypothesis 4, we find surprising results, when utilizing the two conventional indicators to patent value we find no support, the results for NPR_DUM, having science based cognitive map becomes insignificant in both cases: if using Forward citations as dependent variable the coefficient remains positive however non-significant (p=0.226) and the results using Family size as dependent variable remains positive coefficients however in-significant (p=0.239). This could indicate that when using patent value indicators as dependent variable in empirical hypothesis testing with relatively low numbers (here only 110 formulations patents were observed), combining several well selected value indicators based on industry characteristics, can enhance the accuracy of the indicator, and thereby possibly remove noise from other indicators.

Insert Table 10, 11, 12 and 13 about here

DISCUSSION AND CONCLUSIONS

With reference to a science-based industry this paper has studied the orientations in cognitive mapping in early and late R&D. First, we have examined whether different steps of the R&D cycle diverge in their emphasis on the two orientations, a science based cognitive map and a technology based cognitive map. Second, even if on cognitive approach may predominate in their respective stages of the R&D cycle the two orientations may still be present in all stages. We examined whether the two orientations exhibit orthogonal presence, only balanced 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, estimating the effects of cognitive variety at different ends of the R&D cycle.

To address these issues we capitalized on the clearly structured steps characterizing the R&D of biotech firms specialized in drug discovery. Given the tight and aggressive appropriability regime (Levin, Klevorick et al. 1987), firms in the drug industry tend to file patents on inventions coming out of each of these steps, for example, on the compound, its utility, or its formulation (Gupta and Bansal 2002; Kaushal and Garg 2003; Hopkins, Martin et al. 2007). These are in fact "sub-inventions", which, if successful, accumulate over the R&D cycle into a marketable drug. On the basis of this cumulative sequence of sub-inventions, we make 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 the formulation of production processes), building on the previous steps.

Furthermore, the industry is unusually articulate in terms of the patent-based paper trails it leaves from each R&D stage. Particularly important for our purposes are the footprints left in each patent of the search associated with each sub-invention. The backward references of a patent indicate the extent to which the search behind the invention has been directed by a science based

cognitive map (indicated by the backward non-patent references (NPR) of patents, or was directed by a technology based cognitive map, indicated by backward citations to previous patents (Benner and Tushman 2002; Katila and Ahuja 2002; Laursen, Leone et al. 2010; Phelps 2010; Laursen 2011). A predominantly science based cognitive map is, as expected, closely associated with an explorative orientation in problem-solving in early R&D; by contrast, a predominantly technology based cognitive map is guiding problem solving at late R&D, as expected. Although these associations intuitively seem likely, so far they have not been the subject of systematic statistical tests.

Therefore, we examined whether early and late orientations of R&D actually differ in their patterns of search directed by science based cognitive map on the one hand, and by a technology based cognitive map on the other. The results confirmed that the research associated with drug discovery – by many standards the most advanced science driven innovation process to be found in industry - indeed applies different cognitive maps across its different R&D stages. Early R&D has a stronger orientation in its search toward science than do later stages of R&D. Conversely, later stages search with comparatively stronger orientation toward a technology based cognitive map. Still, in both cases we observe a blend of the cognitive maps, rather than a total replacement of one by the other. What differs is the balancing point between the two distinct different approaches. In this way, borrowing from each other's logic, the two orientations are clearly orthogonal, not mutually exclusive.

Next we turned to the effects on value associated with applying cognitive variety within each invention. Since the two are balanced in different proportions in early and late stage R&D, we designed separate tests. For the predominantly science based cognitive map of early stage R&D, we found that further increasing that predominance of science detracted from invention value. At

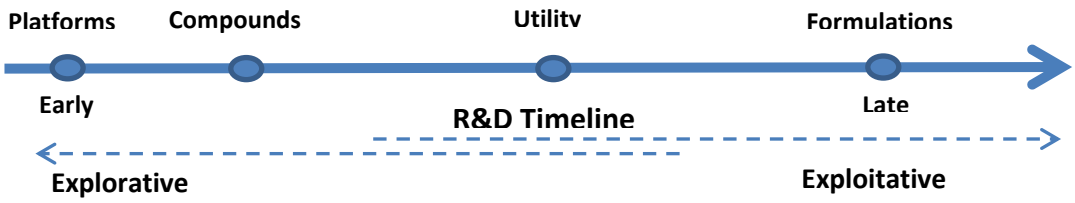
the point when the increase in science based cognitive map approached total predominance, the decline in value accelerated markedly. Conversely, late stage R&D, primarily technology based cognitive map in its orientation, obtains increases in invention value when elements of science based cognitive search are added. The results extend organizational theory by proposing how search during different stages effect performance in dynamic environments, adding to the research on micro-foundations, as the cognitive variety in search approach is empirically tested to have significant effect on the value of new inventions created.

Turning to the methodological contributions of the paper, its use of citations as indicators of search directions allows us to contribute to a more general debate on the role of science. A small, but growing literature has examined the contributions to value creation coming from science when it becomes an integral part of the problem solving in industrial R&D. So far these effects have been studied in the form of direct impact from science citations/linkages on patent value (Dietmar Harhoff, Frederic M Scherer et al. 2003; Cassiman, Veugelers et al. 2008; Nemet and Johnson 2012). Previous studies have obtained inconclusive results from this approach. By contrast, we find fairly strong effects when applying a contingency design, conjecturing effects for early R&D that are the opposite of those of late R&D. Consistent with previous studies, for explorative R&D we find no significant effects on patent value from increases in the absolute number of science citations. But we do find that a negative, significant effect appears when an orientation toward science assumes increasing predominance in the overall pattern of search. When this predominance expands to leave only marginal attention to earlier technology, the negative effects on patent value further accelerate approximately when the share of citations to earlier technology drops to less than half of its average level. Conversely, positive effects of

adding science to the search pattern emerge with particular strength in the context of late stage R&D.

FIGURES AND TABLES

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



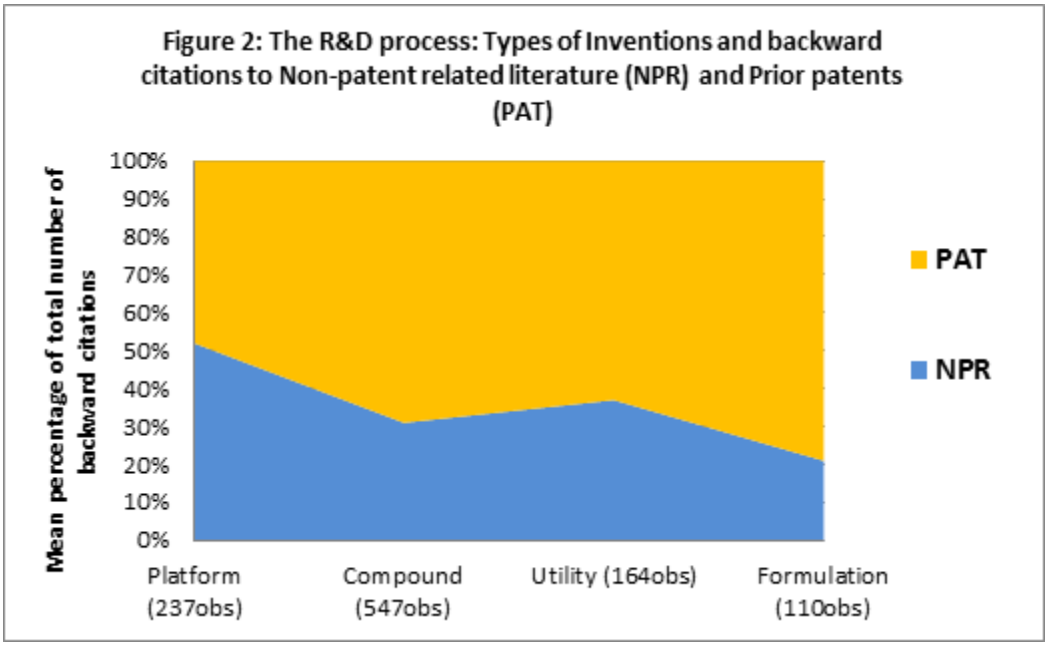


Table 1 Descriptive statistics, all patent types

Variable	Obs	Mean	Std. Dev.	Min	Max
Pat_type_C	1058	.5170132	.4999468	0	1
Pat_type_P	1058	.2240076	.4171242	0	1
Pat_type_U	1058	.1550095	.3620849	0	1
Pat_type_F	1058	.1039698	.3053657	0	1
PAT_VAL	1058	-.1430077	1.40923	-1.424756	17.34493
BW_CIT	1058	7.289225	5.159472	0	35
NPR_CIT	1058	3.034972	3.534357	0	22
PAT_CIT	1058	4.254253	3.195995	0	26
X_CIT_NPR	1058	1.614367	2.913555	0	22
X_CIT_PAT	1058	2.37051	2.977487	0	26
TECH_SCOPE	1058	9.70794	11.282	1	86
ACC_PAT	1058	51.80718	69.37456	0	314
FIRM_TYPE	1058	.4886578	.5001077	0	1
AGE_DUM	1058	.2344045	.4238264	0	1
LN_FAM_SIZE	1058	2.203967	.8350453	0	4.882802
GRANT	1058	.3364839	.4727301	0	1
WITHDRAW	1058	.2381853	.4261745	0	1
YEAR	1058	2003.289	2.80483	1997	2008
Compound patents, descriptive statistics					
BW_CIT	547	7.283364	5.871811	0	35
NPR_CIT	547	2.707495	3.6668	0	22
PAT_CIT	547	4.575868	3.607481	0	26
X_CIT_NPR	547	1.6234	3.25487	0	22
X_CIT_PAT	547	2.621572	3.528706	0	26
NPRCit_RATIO	547	.3080588	.2857164	0	1
Formulation patents, descriptive statistics					
BW_CIT	110	6.036364	3.17654	1	22
NPR_CIT	110	1.536364	2.26937	0	13
PAT_CIT	110	4.5	2.341293	0	13
X_CIT_NPR	110	.5909091	1.467041	0	7
X_CIT_PAT	110	2.363636	2.187233	0	10
NPR_DUM	110	.5636364	.4982036	0	1

Table 2: Pairwise correlation

	PATENT~E	Pat_ty~C	Pat_ty~P	Pat_ty~U	Pat_ty~F	BW_CIT	NPR_CIT
PATENTVALUE	1.0000						
Pat_type_C	0.0573	1.0000					
Pat_type_P	0.0038	-0.5559	1.0000				
Pat_type_U	-0.0667	-0.4431	-0.2301	1.0000			
Pat_type_F	-0.0199	-0.3524	-0.1830	-0.1459	1.0000		
BW_CIT	0.0900	-0.0012	0.0231	0.0449	-0.0828	1.0000	
NPR_CIT	0.0098	-0.0959	0.1718	0.0564	-0.1445	0.7923	1.0000
PAT_CIT	0.1344	0.1042	-0.1528	0.0101	0.0262	0.7381	0.1733
X_CIT_NPR	0.0571	0.0032	0.0602	0.0271	-0.1197	0.7241	0.8316
X_CIT_PAT	0.1319	0.0873	-0.1461	0.0485	-0.0008	0.6742	0.2444
TECH_SCOPE	0.3675	0.1756	-0.0936	-0.0225	-0.1329	0.1705	0.0704
ACC_PAT	-0.1025	0.2343	-0.2205	0.0874	-0.1860	-0.0524	-0.1316
FIRM_TYPE	0.0332	0.2675	-0.2168	0.0881	-0.2463	0.0427	-0.0536
AGE_DUM	0.1410	-0.0635	0.1148	-0.0336	-0.0130	0.1178	0.1215

	PAT_CIT	X_CIT_~R	X_CIT_~T	TECH_S~E	ACC_PAT	FIRM_T~E	AGE_DUM
PAT_CIT	1.0000						
X_CIT_NPR	0.2493	1.0000					
X_CIT_PAT	0.8182	0.3798	1.0000				
TECH_SCOPE	0.1974	0.1422	0.2145	1.0000			
ACC_PAT	0.0611	-0.0008	0.0570	-0.0174	1.0000		
FIRM_TYPE	0.1282	0.0723	0.1274	0.1417	0.5342	1.0000	
AGE_DUM	0.0558	0.0909	0.0571	0.1479	-0.2303	-0.0008	1.0000

Table 3 – Negative binomial regression models Dependent variable is NPR in Model 1 and X_NPR_CIT in Model 2. Negative coefficients indicate decreased probability that knowledge source in patent is scientific literature. Platform inventions are benchmark.

	Model 1	Model 2
	NPR_CIT	X_NPR_CIT
PAT_TYPE_C	-0.462***	-0.537***
	[0.097]	[0.114]
PAT_TYPE_U	-0.260**	-0.391***
	[0.108]	[0.088]
PAT_TYPE_F	-0.886***	-1.109***
	[0.141]	[0.195]
BW_CIT	0.132***	0.174***
	[0.008]	[0.010]
TECH_SCOPE	-0.001	0.008**
	[0.003]	[0.004]
FIRMSIZE	0.000***	0.000
	[0.000]	[0.000]
ACC_PAT	-0.003***	-0.001**
	[0.000]	[0.001]
FIRM_TYPE	-0.175**	-0.032
	[0.083]	[0.157]
AGE_DUM	-0.072	-0.272**
	[0.068]	[0.125]
LN_FAM_SIZE	-0.074*	-0.136*
	[0.038]	[0.077]
GRANT	-0.100	-0.187
	[0.086]	[0.137]
WITHDRAW	-0.062	-0.047
	[0.063]	[0.122]
YEAR	0.001	-0.015
	[0.010]	[0.025]
Constant	-1.438	29.904
	[21.036]	[50.419]
lnalpha		
Constant	-1.555***	-0.276**
	[0.189]	[0.129]
Pseudo LL	-1992.775	-1489.646
No of Obs	1058	1058
Wald-Chi2	811.5829***	693.87***

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

Table 4 – Negative binomial regression models Dependent variable is PAT_CIT in Model 1 and X_PAT_CIT in Model 2. Negative coefficients indicate decreased probability that knowledge source in patent is prior technologies. Formulation inventions are benchmark.

	Model 3	Model 4
MainPT	PAT_CIT	X_PAT_CIT
PAT_TYPE_C	-0.276***	-0.409***
	[0.066]	[0.126]
PAT_TYPE_U	-0.298***	-0.263*
	[0.080]	[0.136]
PAT_TYPE_P	-0.484***	-0.653***
	[0.073]	[0.142]
BW_CIT	0.076***	0.118***
	[0.003]	[0.008]
TECH_SCOPE	0.002	0.007**
	[0.002]	[0.003]
FIRMSIZE	0.000	0.000
	[0.000]	[0.000]
ACC_PAT	0.001**	0.001
	[0.000]	[0.001]
FIRM_TYPE	0.032	0.026
	[0.057]	[0.136]
AGE_DUM	-0.023	-0.010
	[0.064]	[0.114]
LN_FAM_SIZE	0.039	0.062
	[0.025]	[0.050]
GRANT	0.030	-0.147
	[0.098]	[0.134]
WITHDRAW	0.081	-0.041
	[0.055]	[0.091]
YEAR	-0.009	-0.009
	[0.007]	[0.015]
Constant	17.992	17.460
	[13.970]	[30.329]
lnalpha		
Constant	-3.397***	-0.880***
	[0.253]	[0.125]
Pseudo LL	-2185.183	-1930.530
No of Obs	1058	1058
Wald-Chi2	1263.675***	817.5432***

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

Table 5 – TOBIT regression
Dependent variable is PAT_VAL. The presented results are for compound inventions.
Positive coefficients indicate increased probability of a more valuable invention.

	Model 5
model	
NPRCit_RATIO	-0.485**
	[0.231]
FIRMSIZE	0.001***
	[0.000]
ACC_PAT	-0.002***
	[0.001]
FIRM_TYPE	0.112
	[0.150]
AGE_DUM	0.036
	[0.178]
GRANT	0.279
	[0.173]
WITHDRAW	-0.654***
	[0.136]
YEAR_Period	-0.578***
	[0.094]
TECH_SCOPE	0.021***
	[0.004]
Constant	0.960***
	[0.291]
sigma	
Constant	0.982***
	[0.059]
No of Obs	547
Uncensored o~s	546
Log likelyhood	-765.3264
Pseudo R-squ~d	.1582612
F test	227.8537***

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

Table 7 – TOBIT regression
Dependent variable is PAT_VAL. The presented results are for formulation inventions.
Positive coefficients indicate increased probability of a more valuable invention.

	Model 5
model	
NPR_DUM	0.275*
	[0.142]
FIRMSIZE	0.001
	[0.003]
ACC_PAT	-0.002
	[0.006]
FIRM_TYPE	-0.322
	[0.241]
AGE_DUM	0.517
	[0.344]
GRANT	0.244
	[0.296]
WITHDRAW	-0.679**
	[0.341]
YEAR_Period	-0.307*
	[0.162]
TECH_SCOPE	0.096***
	[0.034]
Constant	-0.333
	[0.487]
sigma	
Constant	0.897***
	[0.118]
No of Obs	110
Uncensored o~s	107
Log likelyhood	-142.6831
Pseudo R-squ~d	.1728685
F test	5.818047***

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

Table 6: Coefficient estimates for percentile regression for compound patents

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

Figure 3: Percentile regression estimates for NPRCit_RATIO for compound patents

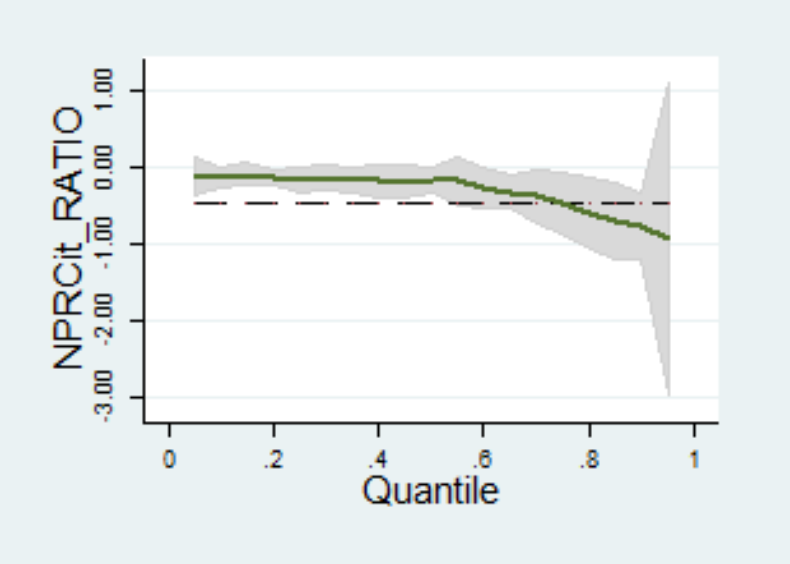


Table 8 – Negative binomial regression models Dependent variable is NPL/PAT_CIT. Negative coefficients indicate decreased probability of a higher degree of NPL pat citations in relation to PAT citations. Platform inventions are benchmark.

	Model 6
NPL/PAT_CIT	
PAT_TYPE_C	-0.671***
	[0.141]
PAT_TYPE_U	-0.412**
	[0.206]
PAT_TYPE_F	-1.382***
	[0.221]
BW_CIT	0.067***
	[0.009]
TECH_SCOPE	-0.009*
	[0.005]
FIRMSIZE	0.000
	[0.000]
ACC_PAT	-0.003***
	[0.001]
FIRM_TYPE	-0.355***
	[0.114]
AGE_DUM	-0.007
	[0.118]
LN_FAM_SIZE	-0.078
	[0.070]
GRANT	-0.110
	[0.135]
WITHDRAW	-0.126
	[0.112]
YEAR	0.003
	[0.021]
Constant	-4.953
	[41.209]
lnalpha	
Constant	-0.679***
	[0.118]
Pseudo LL	-1329.420
No of Obs	1016
Wald-Chi2	455.5041***

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

Table 9 – Negative binomial regression models Dependent variable is X_NPL/X_PAT_CIT. Negative coefficients indicate decreased probability of a higher degree of X NPL pat citations in relation to X PAT citations. Platform inventions are benchmark.

	Model 7
XNPL_XPAT	
PAT_TYPE_C	-0.750***
	[0.206]
PAT_TYPE_U	-0.641***
	[0.163]
PAT_TYPE_F	-1.706***
	[0.245]
BW_CIT	0.110***
	[0.014]
TECH_SCOPE	-0.005
	[0.006]
FIRMSIZE	0.000
	[0.000]
ACC_PAT	-0.001
	[0.001]
FIRM_TYPE	-0.121
	[0.154]
AGE_DUM	-0.162
	[0.179]
LN_FAM_SIZE	-0.152
	[0.108]
GRANT	-0.065
	[0.227]
WITHDRAW	-0.126
	[0.201]
YEAR	-0.027
	[0.035]
Constant	53.759
	[70.871]
lnalpha	
Constant	-0.175
	[0.114]
Pseudo LL	-803.432
No of Obs	728
Wald-Chi2	151.4922***

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

Table 10 –Negative binomial regression
Dependent variable is PAT_VAL measured by total number of forward citations. The presented results are for compound inventions.
Positive coefficients indicate increased probability of a more valuable invention.

	Model 8
FORWARD CITATIONS	
NPRCit_RATIO	-0.552**
	[0.267]
FIRMSIZE	0.001*
	[0.000]
ACC_PAT	-0.002***
	[0.001]
FIRM_TYPE	0.229*
	[0.132]
AGE_DUM	-0.039
	[0.157]
GRANT	-0.099
	[0.124]
WITHDRAW	-0.572***
	[0.176]
YEAR_Period	-0.777***
	[0.114]
TECH_SCOPE	0.012***
	[0.004]
Constant	2.469***
	[0.328]
lnalpha	
Constant	-0.109
	[0.098]
Pseudo LL	-1122.805
No of Obs	547
Wald-Chi2	868.302***

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

Table 11 –Negative binomial regression
Dependent variable is PAT_VAL measured by family size. The presented results are for compound inventions.
Positive coefficients indicate increased probability of a more valuable invention.

	Model 9
Family_size	
NPRCit_RATIO	-0.242*
	[0.130]
FIRMSIZE	0.001***
	[0.000]
ACC_PAT	-0.002***
	[0.001]
FIRM_TYPE	0.045
	[0.122]
AGE_DUM	0.009
	[0.105]
GRANT	0.484***
	[0.073]
WITHDRAW	-0.379***
	[0.115]
YEAR_Period	-0.253***
	[0.070]
TECH_SCOPE	0.008***
	[0.002]
Constant	2.888***
	[0.176]
lnalpha	
Constant	-1.464***
	[0.121]
Pseudo LL	-1759.003
No of Obs	547
Wald-Chi2	691.1537***

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

Table 12 – Negative binomial regression Dependent variable is number of Forward citations. The presented results are for formulation inventions. Positive coefficients indicate increased probability of a more valuable invention.

	Model 10
FORWARD CITATIONS	
NPR_DUM	0.178
	[0.147]
FIRMSIZE	0.010**
	[0.005]
ACC_PAT	-0.025*
	[0.013]
FIRM_TYPE	-0.433
	[0.447]
AGE_DUM	0.690**
	[0.276]
GRANT	-0.066
	[0.230]
WITHDRAW	-0.681**
	[0.298]
YEAR_Period	-0.227
	[0.210]
TECH_SCOPE	0.079***
	[0.020]
Constant	0.832*
	[0.494]
lnalpha	
Constant	-0.414
	[0.273]
Pseudo LL	-219.819
No of Obs	110
Wald-Chi2	103.9132***

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

Table 13 – Negative binomial regression Dependent variable is Family size. The presented results are for formulation inventions. Positive coefficients indicate increased probability of a more valuable invention.

	Model 11
Family_size	
NPR_DUM	0.149
	[0.127]
FIRMSIZE	-0.004
	[0.004]
ACC_PAT	0.011**
	[0.005]
FIRM_TYPE	-0.006
	[0.202]
AGE_DUM	-0.068
	[0.196]
GRANT	0.637***
	[0.162]
WITHDRAW	-0.216
	[0.246]
YEAR_Period	-0.360***
	[0.126]
TECH_SCOPE	0.018
	[0.013]
Constant	2.577***
	[0.413]
lnalpha	
Constant	-1.319***
	[0.143]
Pseudo LL	-337.091
No of Obs	110
Wald-Chi2	141.7833***

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

REFERENCES

Ahuja, G. and C. M. Lampert (2001). "Entrepreneurship In The Large Corporation: A Longitudinal Study Of How Established Firms Create Breakthrough Inventions." Strategic Management Journal **22**(6/7): 521-543.

Ai, C. R. and E. C. Norton (2003). "Interaction terms in logit and probit models." Economics Letters **80**(1): 123-129.

Alcacer, J. and M. Gittelman (2006). "Patent citations as a measure of knowledge flows: The influence of examiner citations." Review of Economics and Statistics **88**(4): 774-779.

Barber, P. J. (1988). Applied Cognitive Psycology: An information Processing Framework.

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

Benner, M. J. and M. L. Tushman (2003). "Exploitation, exploration, and process management: The productivity dilemma revisited." Academy of Management Review **28**(2): 238-256.

Brown, J. S. and P. Duguid (2001). "Knowledge and organization: A social-practice perspective." Organization Science **12**(2): 198-213.

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

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

Cassiman, B., R. Veugelers, et al. (2008). "In search of performance effects of (in)direct industry science links." Industrial and Corporate Change **17**(4): 611-646.

Dietmar Harhoff, Frederic M Scherer, et al. (2003). "Citations, family size, opposition and the value of patent rights." Research Policy **32**(8): 1343-1363.

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**(2): 128-143.

Eisenhardt, K. M., N. R. Furr, et al. (2010). "Microfoundations of Performance: Balancing Efficiency and Flexibility in Dynamic Environments." Organization Science **21**(6): 1263-1273.

EPO, USPTO, et al. (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**(1): 117-132.

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

Gambardella, A., D. Harhoff, et al. (2008). "The value of European patents " European Management Review **5**(2): 69-84.

Gavetti, G. and D. Levinthal (2000). "Looking forward and looking backward: Cognitive and experiential search." Administrative Science Quarterly **45**(1): 113-137.

Gibbons, M. and R. Johnston (1974). "The Roles of Science in Technological Innovation." Research Policy **3**: 220-242.

Granstrand, O. (1999). The Economics and Management of Intellectual Property towards Intellectual Capitalism, Edward Elgar: Cheltenham, UK.

Gupta, A. K., K. G. Smith, et al. (2006). "The interplay between exploration and exploitation." Academy of Management Journal **49**(4): 693-706.

Gupta, P. and A. K. Bansal (2002). "Patent Opportunities in Matrix-Based Oral Controlled Release Drug Delivery Systems, Part I." Pharmaceutical Technology Europe, **14(9): 49.**

Hamilton, K. S. (2003). Subfield and Level Classification of Journals, CHI Research Inc.

Hamilton, S., J. Armstrong, et al. (1996). "An overview of automated biotechnology screening." Laboratory Robotics and Automation **8**: 287-294.

Hopkins, M. M., P. A. Martin, et al. (2007). "The myth of the biotech revolution: An assessment of technological, clinical and organisational change." Research Policy **36**(4): 566-589.

Katila, R. and G. Ahuja (2002). "Something old, something new: A longitudinal study of search behavior and new product introduction." Academy of Management Journal **45**(6): 1183-1194.

Kaushal, A. M. and S. Garg (2003). "An Update on Osmotic Drug Delivery Patents." Pharmaceutical Technology **27(8): 38.**

Kline, S. J. and N. Rosenberg (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. and G. Bassett (1978). "REGRESSION QUANTILES." Econometrica **46**(1): 33-50.

Lachman, R., J. L. Lachman, et al. (1979). Cognitive psycology and Information Processing: An Introduction, Hillsdale, NJ.

Lanjouw J.O. and Schankerman M. (1997). "Stylized facts of patent litigation: value, scope, ownership, ." NBER Working Paper No. 6297. National Bureau of Economic Research, Cambridge, MA.

Lanjouw, J. O. and M. Schankerman (1999). "THE QUALITY OF IDEAS: MEASURING INNOVATION WITH MULTIPLE INDICATORS." NBER Working Paper 7345.

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., M. I. Leone, et al. (2010). "Technological exploration through licensing: new insights from the licensee's point of view." Industrial and Corporate Change **19**(3): 871-897.

Leonard-Barton, D. (1992). "Core capabilities and core rigidities: A paradox in managing new product development." Strategic Management Journal **13**(S1): 111-125.

Levin, R. C., A. K. Klevorick, et al. (1987). "Appropriating the Returns from Industrial Research and Development." Brookings Papers on Economic Activity **1987**(3): 783-831.

Macher, J. T. (2006). "Technological development and the boundaries of the firm: A knowledge-based examination in semiconductor manufacturing." Management Science **52**(6): 826-843.

March, J. G. (1991). " Exploration and Exploitation in Organizational Learning. ." Organization Science, : 2(1): 71-87.

MPA. (2012). "Types of Pharmaceutical Patent - Product, Process, Formulation and Method." from http://mpasearch.co.uk/product-process-formulation-patents

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**(3): 453-470.

Nemet, G. F. and E. Johnson (2012). "Do important inventions benefit from knowledge originating in other technological domains?" Research Policy **41**(1): 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. B. Insights.
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**(4): 890-913.
Piaget, J. (1985). The Equilibration of Cognitive Structures: The Central Problem of Intellectual Development., University of Chicago Press, Chicago.
Roach, M. and W. Cohen (2012). "“Lens or Prism? An Assessment of Patent Citations as a Measure of Knowledge Flows from Public Research.”." forthcoming Management Science.
Schumpeter, J. A. (1939). Business Cycles.
Smith, W. K. and M. L. Tushman (2005). "Managing strategic contradictions: A top management model for managing innovation streams." Organization Science **16**(5): 522-536.
Tolman, E. C. (1948). "COGNITIVE MAPS IN RATS AND MEN." Psychological Review **55**(4): 189-208.
Tuahene-Gima, K. (2005). "Resolving the CapabilityGÇöRigidity Paradox in New Product Innovation." Journal of Marketing **69**(4): 61-83.
Tushman, M. L. and C. A. Oreilly (1996). "Ambidextrous organizations: Managing evolutionary and revolutionary change." California Management Review **38**(4): 8-&.
Verma, R. M., B; Garg, S (2000). "Osmotically controlled oral drug delivery." DRUG DEVELOPMENT AND INDUSTRIAL PHARMACY **26**(7): 695-708.
Webb, C., H. Dernis, et al. (2005). " ANALYSING EUROPEAN AND INTERNATIONAL PATENT CITATIONS: A SET OF EPO PATENT DATABASE BUILDING BLOCKS " STI WORKING PAPER 2005/9
Wooldridge (2009). Introductory Econometrics - A Modern Approach, South Western Cengage Learning.
Yayavaram S and Ahuja G (2008). "Decomposability in knowledge structures and its impact on the usefulness of inventions and knowledge-base malleability." Administrative Science Quarterly **53**(2): 333-362.
Yoo, H., C. Ramanathan, et al. (2005). "Intellectual property management of biosequence information from a patent searching perspective." World Patent Information **27**(3): 203-211.