

Types of learning in complex technological innovations

Authors

Lars Alkaersig, Technical U. of Denmark, lalk@dtu.dk

Karin Beukel, CBS, kb.ino@cbs.dk

Giancarlo Lauto, U. of Udine, giancarlo.lauto@uniud.it

Finn Valentin, CBS, fv.ino@cbs.dk

Submission #15398 accepted for the 2013 Academy of Management Annual Meeting

Types of learning in complex technological innovations

ABSTRACT

This paper examines how the knowledge base of a firm affects the generation of combinatorial innovations. Recent insights in the literature on organizational learning suggest that the degree of integration among bodies of specialized technological knowledge affects the outcomes of the innovation process. However, the literature is still in its infancy in terms of the relationship between the structure of the knowledge base and the degree of technological complexity of innovations. The present study aims at understanding this relationship.

We take as our empirical case hydrocracking, a rather mature technology widely applied in the oil industry. Innovations in hydrocracking are based on the combination of elementary technologies relative to three technology areas: feeds and products, catalyst preparation, and refinery processing.

Using detailed patent information, we find that a specialized knowledge base enables the generation of innovations characterized by both high and low levels of complexity – i.e. encompassing both few and many multiple technological domains, while related knowledge matters only for low-complexity innovations.

INTRODUCTION

The view of innovation as a recombination of resources, including knowledge of products, markets, and technologies has deep roots in the literature (Schumpeter 1939; Nelson and Winter 1982). More recently, scholars have addressed the issue of how technological inventions come about (Fleming and Sorenson 2001), suggesting possible interpretations in terms of scope and depth of learning (Katila and Ahuja 2002; Schilling and Green 2011) and combinatorial capabilities (Kogut and Zander 1992; Danneels 2002). This research has greatly advanced our understanding of how the exploration of unfamiliar technological domains sustains the generation of combinatorial innovations. Focusing in particular on high technology industries such as microelectronics and biotechnology, this stream of studies has taught us how firms benefit from search in knowledge domains that are distant from their existing knowledge base, or absorb inputs originated by other firms and actors (Sorensen and Stuart 2000; Rosenkopf and Nerkar 2001; Makri, Hitt et al. 2010). This literature has emphasized that scientific knowledge drives firms' exploration of the technological landscape for new combinatorial inventions (Fleming and Sorenson 2004). In contrast to innovations in rapidly changing knowledge environments, innovations in more stable settings, such as mature industries, tend to build on cognitive domains that are proximate to firms' existing knowledge bases (Van den Bosch, Volberda et al. 1999). In such a context, familiarity with the technology domain plays a greater role in combinatorial invention than mechanisms for absorption of distant inputs (such as scientific knowledge) because it permits a more efficient local search (March 1991). This suggests looking inside the firm to understand its innovative performance (Argyres and Silverman 2004).

The literature on understanding the internal knowledge bases of firms as drivers of innovation has shifted its focus. While the traditional organizational learning perspective emphasizes experience in a narrow range of activities as a driver of knowledge accumulation (Argote 1993; Argote 1999), later contributions have suggested that a variation of tasks – rather than a specialization – also adds to the process of learning in a core area (e.g. (Schilling, Vidal et al. 2003; Boh, Slaughter et al. 2007; Eggers 2012)) and shapes different learning sequences (Bingham and Davis 2012). In theorizing the different models of learning, Schilling et al. (2003) introduce the distinction between specialized, related, and unrelated learning: the former resembles the traditional learning curve concept, while the latter two identify learning processes that benefit from experience in problem domains that are cognitively related to the core area of practice and a learning model based on the alternance of unrelated problem areas, respectively.

A useful way of conceptualizing the relatedness of learning in the context of specific firms is offered by Simon's theory on problem architectures (Simon 1962): The type of learning processes activated by an organization to solve a problem depend on the nature of the problem itself. For example, the learning processes activated to generate a combinatorial innovation depend on the kind of innovation (here a combinatorial innovation) the firm aims at producing. Combinatorial inventions can be conceived as complex systems, as they combine their constituent elements according to original architectures in which each technology is embedded, i.e. bodies of specialized technological knowledge. A classic contribution by Simon (1962) conceptualizes the structure of systems in three categories – highly, nearly, and non-decomposable – expressing different intensities in the degree of integration among elements or clusters of elements. The nearly-decomposable type, which is characterized by the existence of distinct clusters that are to some extent interconnected, is the one that best captures

combinatorial inventions. In the original conception, Simon refers to decomposability as an attribute of problems that an individual or an organization faces and attempts to solve. Recently, this framework has been adopted and extended by (Yayavaram and Ahuja 2008), who applied the concept of decomposability to a firm's knowledge base, i.e. they considered it to be an attribute of the *problem solver*. Their study examined the relationship between the structure of a knowledge base and innovation performance, finding that a decomposable knowledge base increases the usefulness of the inventions generated from it.

However, even though this stream of literature is well developed, there is a significant gap in our understanding of what level of decomposable knowledge bases (same, related, unrelated) is favorable for generating inventions characterized by different levels of technological complexity (lower vs. higher). Therefore, this paper brings together the literatures on combinatorial invention, complex systems, and organizational learning, examining the relationship between the structure of the knowledge base of a firm and the degree of decomposability of the innovations it generates. We submit that the knowledge base of a firm is the result of its learning processes and that it thus mirrors its accumulated experience in technology development. Therefore, the domain and the degree of complexity of the stock of technologies developed by a given firm at each point in time express the decomposability and relatedness of its knowledge base. For instance, a firm's experience in generating innovations that cover two distinct technological areas indicates the presence of a nearly-decomposable knowledge base; combining three technology areas in generating an invention signals that the firm has built a highly decomposable knowledge base. In this paper, we examine what type of knowledge base firms operating in a stable knowledge environment exploit to address complex problems, such as the generation of a highly technologically complex invention.

We take the empirical case of hydrocracking to explore these issues. Hydrocracking is a mature technology widely used as part of the oil refinery process to transform crude oil into high value petroleum products. Innovations in hydrocracking technology are developed by vertically integrated oil companies, refinery operators, and chemical firms, and are anchored in three main technological areas and any combinations thereof. Thus, this technology offers an ideal setting to compare the learning processes developed to generate innovations based on a single technology area, a combination of two technology areas, and a combination of multiple combinations (triple inventions), expressing different degrees of integration. Our empirical results broadly support our theory, showing that a nearly decomposable knowledge base (characterized by learning processes grounded in two technology areas) enables firms to generate innovations characterized by lower levels of decomposability (characterized by innovations combining two technology areas). However, this does not enable the generation of highly decomposable innovations (those combining all three areas). Highly technologically complex innovations require firms to develop a highly decomposable knowledge base.

The remainder of this paper is organized as follows: in the first section, we illustrate the concept of complexity of the knowledge base; then we discuss the role of two types of learning – specialized and related – in combinatorial innovation. Our empirical analysis follows, and we conclude with a discussion of implications, limitations, and potential extensions of the research.

THEORY

How firms search for multi-technological innovations

Generating a multi-technological innovation can be seen as a challenge of recombining different knowledge bases. Even a knowledge base composed of only a few elements may give rise to a

high number of innovations, because different patterns of combinations of those elements may give rise to many alternative potential innovations (Fleming 2001; Yayavaram and Ahuja 2008). Firms' search strategies in the combinatorial space have traditionally been conceived in terms of "search scope", but more recently the dimension of "search depth" has been integrated in the framework (Katila and Ahuja 2002). Search scope captures the proximity or distance of firms' search behavior from a cognitive or organizational point of view. Firms search locally when they build their innovations on technologies that are related to the domains in which they are specialized (Stuart and Podolny 1996) and when they rely on their internal knowledge base. A broad search scope generally refers to investigations in domains in which firms lack prior experience or competencies (Schilling and Green 2011) and wish to develop new capabilities (Danneels 2002). Local search permits firms to strengthen capabilities in a given technological area, but on the other hand, provokes technological path dependence by locking the firm into a technological trajectory (Dosi 1988; Arthur 1989). Search depth captures the extent to which firms repeatedly use their existing knowledge while searching for new solutions. The application of a stable set of knowledge elements permits firms to develop search routines and to gain a deeper understanding of the knowledge itself. This opens up to the possibility of identifying non-obvious linkages among concepts and disclosing opportunities for recombination. However, excessive reliance on the existing knowledge base binds the firm to a specific technological trajectory (Katila and Ahuja 2002).

The nature of the knowledge environment influences firms' search strategies. In turbulent knowledge environments firms tend to prioritize exploratory search and to adopt strategies suitable to increasing their combinatorial capabilities so that they can capture the opportunities offered by the changing technological landscape by reconfiguring their knowledge base. By

contrast, in a stable knowledge environment, such as a mature industry, firms focus on exploitation of their existing knowledge base (Van den Bosch, Volberda et al. 1999; Cardinal, Turner et al. 2011). This means a search pattern that is local, i.e. confined within the boundaries and tied to the existing knowledge base of the firm, and cognitively deep, i.e. addressing technological domains that are closely related to a firm's knowledge base. While the literature has extensively analyzed the drivers of competitive advantage in high-technology industries characterized by a fast pace of innovation, we lack empirical evidence relative to stable environments. This paper aims at filling this gap by analyzing the different capabilities that drive combinatorial innovation in a mature technological setting.

In the case of a decomposable knowledge base (Yayavaram and Ahuja 2008), innovations can be generated at the architectural level (Henderson and Clark 1990) by establishing novel relationships among constituent technologies or modifying existing relationships. The number of technological domains that can potentially be coupled and the volume of interdependencies among their elements increase enormously the space of possible inventions (Simon 1969; Fleming 2001) and the complexity of the innovation process. As hydrocracking technology builds upon three distinct core technological domains – feeds and products, catalyst preparation, and refinery processing – this empirical setting seems ideal to analyze what internal capabilities firms deploy to deal with the complexity of combinatorial innovation. Inventions in this technology typically build on only one of these three core domains or on combinations of them; little contribution is made by inputs referring to other technologies. In Figure 1, we exemplify the hierarchy of combinatorial innovation in an industry characterized by three core technologies. We argue that the degree of complexity entailed in a combinatorial innovation

increases with the elementary technology areas involved, as the firm has to learn and define the linkages among a greater number of elements and interfaces.

Insert Figure 1 about here

Consistently with previous literature (Argyres and Silverman 2004; Mayer and Argyres 2004; Sampson 2005; Kotha, Zheng et al. 2011), we use patents to capture the technological capabilities of firms and their antecedents. Each patent is characterized by one or more technological domains. We consider that the technological domains associated with the stock of patents accumulated by a firm up to a certain point in time represent the areas of its technological capabilities. The literature on organizational learning (Argote 1993) suggests that a firm's knowledge base is built through the accumulation of capabilities that are formed through experience and learning-by-doing. In this sense, resources and capabilities are the repository of firms' learning.

Yayavaram and Ahuja (2008) introduce the concept of "malleability" to characterize the capacity of a knowledge base to change. Firms reconfigure their knowledge base by elaborating new conjectures about the relationships existing between its elements. This process permits firms to replace existing linkages between components of the knowledge base with new ones. The authors suggest that a nearly-decomposable knowledge base enables firms to search for technological innovations. We suggest that it is possible to interpret malleability of a knowledge base as the result of a learning process. Firms' decisions about what technology areas to couple are driven by routines in technological search. Learning about a technology provides guidance to firms in establishing new linkages and removing old ones, restricting the space for possible recombinations.

An important aspect of learning concerns how firms learn to deal with complex problems. This process can be divided into two types of learning processes: horizontal learning and vertical learning. *Horizontal learning* refers to the identification of analogies and differences between technological problems that have the same level of complexity, while *vertical learning* refers to the understanding of the interrelationships between the levels on which a problem is structured. When firms extend their knowledge base and understand the connections of the new element with the existing ones they develop a different type of knowledge that is more general and abstract in nature. This “meta knowledge” eases combinatorial processes at higher levels of complexity. In other words, the translation of knowledge developed by combining technologies from technology areas A and B to technology areas B and C can be regarded as an example of horizontal learning, as the combinatorial activity requires an analogous effort. Instead, the effort to extend the learning developed in combinatorial inventions in areas A and B to areas A, B and C is higher, as it requires developing knowledge relative to the interrelationship among all three areas. We refer to specialized learning as the accumulated experience in innovating in a *specific* technological area, while we conceive related capabilities as the accumulated experience in innovating *across* related yet different technological fields. The development of complex innovations necessitates activating the knowledge base relative to the technological areas as well as the competencies accumulated through “meta-learning” relative to the production of combinatorial inventions. Instead, firms can more easily transfer their accumulated experience “horizontally” to related technological areas when innovations are characterized by a lower level of internal interconnectedness.

It is also important to note that the levels of learning are characterized by a different degree of tacitness of the knowledge involved. Innovations building on the combination of many

technology areas are characterized by a higher level of tacit knowledge, as the specific patterns of combination of constituting elements are the result of learning processes specific to each firm. For these reasons, we expect that specialized knowledge bases produce innovations characterized by both high and low levels of complexity, while related knowledge matters only for low-complexity innovations.

Insert Figure 2 about here

In line with the literature on organizational learning suggesting that subsequent attempts at complex technological innovations have a higher chance of success with increasing returns, we expect:

HYP 1a The probability of a firm generating a patent encompassing two technology domains is positively associated with the firm's previous experience in patenting across those two technology domains.

HYP 1b The probability of a firm generating a patent encompassing all three technology domains is positively associated with the firm's previous experience in patenting across those three technology domains.

The use of existing capabilities (in our case, the exploitation of a specialized knowledge base in a given domain) to build a new capability – such as the development of knowledge based on the integration of multiple elements – has been defined as “competence leveraging” (Danneels 2002; Miller 2003; Danneels 2007). Learning how to recombine specialized technological capabilities, firms develop second-order “combinative capability”. This concept was introduced by (Kogut and Zander 1992; Zander and Kogut 1995; Kogut and Zander 1996), who refer to “the

intersection of the capability of the firm to exploit its knowledge and the unexplored potential of the technology”. By means of combinative capabilities, firms can use their knowledge base to pursue technological opportunities that are distant from it. We note that in the case of multi-technological innovations, firms may rely on combinative capabilities in two circumstances:

First, when a firm develops a multi-technological capability relative to the coupling of two specific technological domains – e.g. it knows how to deal with the interdependencies and complementarities of two given technologies – and applies it to generate recombinant innovations involving other domains. In this case, the firm is experienced in generating combinatorial innovations but lacks specific knowledge within the technological domains that are coupled. The theory on “competence leveraging” suggests that cumulated experience in technological coupling in a given domain can be fruitfully applied in another domain. Therefore, we expect:

HYP 2 The probability of a firm generating a patent encompassing two technology domains is positively associated with the firm’s previous experience in patenting across the related two technology domains

To exemplify: the probability of generating a double invention is positively associated with previous experience in patenting in related double inventions. The other possible use of combinative capabilities refers to the situation in which firms couple different technologies in the absence of multi-technological capabilities. In such a case, the firm is familiar with each technology, but lacks knowledge of their mutual relationships. We suggest that combinatorial capabilities permit firms to overcome this gap. However, strong technological specialization in a narrow set of domains may cause core rigidities (Leonardbarton 1992) and competency traps

(Levitt B and March J G 1988; Levinthal and March 1993), preventing firms from further exploring novel technological domains. For this reason, we expect:

HYP 3a The probability of a firm generating an innovation building on two technology domains is negatively associated with the firm's previous experience in patenting in the individual constituent technological domains.

HYP 3b The probability of a firm generating an innovation building on all three technology domains is negatively associated with the firm's previous experience in patenting in the individual constituent technological domains.

One of the effects of core rigidities is to bound firms' search patterns along narrowly defined technological domains. The presence of a specialized knowledge base thus hinders firms' ability to apply knowledge developed in a given technological field to a different distinct technological area. For this reason, it is expected that specialization of the knowledge base – indicating a high propensity for a firm to develop core rigidities – limits the conductivity of a related knowledge base to combinatorial innovation. For this reason we expect:

HYP 4 The relationship between having a related knowledge base and generating a double invention is negatively moderated by having a specialized knowledge base.

DATA

To examine our hypotheses, this study draws on a unique dataset consisting of the collective patent applications within *hydrocracking*. This patent data is classified into the distinct technology areas within this industry on the level of each innovation, and uses combinations of these classifications to indicate whether innovations combine different technology areas, and which areas are combined.

Previous studies looking into technology combinations in patent data have predominantly relied on the *International Patent Classification* (IPC) system to determine the technological scope of the patent (Lerner 1994; Dietmar Harhoff, Frederic M. Scherer et al. 2003). This discrepancy between studies using IPC codes as a measure for the breadth of a patent is in part caused by the nature of the IPC classification itself. Each patent is assigned by the patent examiner to a number of nine-digit IPC classes. While examiners within the same patent office are coherent in their application of IPC codes, key differences exist between regions. A prime example is the difference between the US and the EU; the primary classification of more than 50 percent of US patents is changed after examination by the German patent office (Dietmar Harhoff, Frederic M. Scherer et al. 2003). This lack of clarity in the patent classification is further apparent when examining the use of multiple IPC classes for a single patent. In a study examining the use of co-classification of IPC codes at the three- and four-digit levels in a sample of 138,681 patents, Leydesdorff (2008) found a weak relation between the classifications, where a co-relation of patent citations was found to be a stronger indicator of relation than a co-classification of patents. Furthermore, previous studies have found that SIC and IPC codes have a weak correlation, indicating that a technology area cannot easily be defined in IPC codes, and that a technology area might cover multiple IPC codes from different parts of the classification (Cohen, Nelson et al. 2002).

In this study we offer an alternative to counting IPC classes to determine the breadth of the technology area(s) of a patent. With the aid of a technical expert within the hydrocracking field, we identified three distinct technology areas within the industry and attributed to each technology area the relevant nine-digit IPC classes. These three technology areas are used to classify all patents within our sample as patents covering either a single technology area or a

combination of multiple technology areas. During this process it became apparent that IPC classes are a weak indicator of technological proximity because they specify both what technologies are covered by a patent and what are not. In the literature, two nine-digit IPC subclasses are considered similar if they share the first seven digits. However we found that this apparent similarity cannot be taken for granted. It is not uncommon that two IPC subclasses sharing the first seven digits describe conflicting processes or the use of technologies within a process. Thus, two IPC subclasses that may seem highly similar due to being grouped within the same seven-digit class can be mutually exclusive and essentially describe conflicting or competing technologies. Therefore, the full informative potential of the IPC classification can only be extracted by means of expert advice.

Hydrocracking is a technology surrounding a late stage of the process of refining crude oil to high value petroleum products. The forerunner of the technology was developed in 1927 to hydrogenate distillates from coal, and in the present day version, hydrocracking is a catalytic cracking process that converts heavy hydrocarbons into higher added value, lower molecular weight compounds under hydrogen pressure (Billon and Bigeard 2001). This is a technology aimed at increasing the yield of high value products from crude oil, and converts the lower value lubricating oils and heavy gas oils that invariably result from refining crude oil into higher value products such as low-sulfur diesel fuel and jet fuel. While hydrocracking is a mature technology, the continued application of the technology in modern refineries ensures that the technology is continuously developed.

Insert Figure 3 about here

With expert aid, we identified three distinct technology areas within hydrocracking. Process technologies (A) are primarily associated with how the process of hydrocracking is integrated into the overall refinery process, and therefore includes technologies concerning the flow of petroleum based liquids, including valves, pipes, and the associated controllers. Catalyst preparation (B) is concerned with the manufacturing process of the catalyst necessary for hydrocracking to take place. This includes both the manufacture of the carrier of the catalyst (the base to which the active component in the catalyst is applied) and the application of the active component onto the carrier in the manufacturing process. The area of feeds and products (C) is concerned with the chemical nature of the raw materials of the refineries (feeds), and the chemical reactions that change specific feeds into specific products. These three technology areas find application both alone and in combination with each other¹, where, for example, patents combining the development of the active component (C) with the manufacturing technology (B) are a common combination (BC). The table below shows how IPC classes refer to technology areas within hydrocracking.

 Insert Table 1 about here

However, while the subclasses are generally related, as seen above, exceptions exist where a few specific IPC classes are mutually exclusive from the rest of the subclass. For instance, in the above table we exclude the IPC subclasses C10G-47/24-30 from C as these constitute a competing technology to hydrocracking, which utilizes an entirely different technology and cannot be compared to the hydrocracking process.

¹ A combination of different technology areas is based on the nature of the innovation and the technologies utilized, not on the patent claim itself.

Our patent sample within hydrocracking consists of 3,902 patents from 1977 to 2007 collected from the Derwent Innovation Index. We identified 26 firms with five or more patents in this period from the assignees of these patents, for which we have collected firm level data. This yields a data set of 2,416 patents associated with these firms, with the remaining patents assigned to individuals, universities or firms with fewer than five hydrocracking patents. In order to obtain a measure of patent value, we linked our patent data to the OECD 2010 citations database, which contains citation data for all WPTO and EPO patents. However, these data are not complete, as far from every patent is submitted to the WPTO or EPO. It is not uncommon that patents are submitted to a national patent office only, and not to the international patent offices. Patents not submitted to the WPTO and EPO are commonly patents of little or no commercial value or patents that are not important enough for the firm to have worldwide coverage. It is particularly firms in the US, Japan, and China that submit numerous patents to their national patent offices only. Therefore, when each patent family is counted only once, when only patents that are applied for by applicants with more than five patents in total are used, and when the patent families are linked to the OECD citations dataset, the result is a total of 936 patent families. The findings presented in this paper may therefore be biased towards firms with significant investments in the technology, in that they have five or more patents published by the WPTO/EPO. Below is a list of the patent holders in the hydrocracking industry.

Insert Table 2 about here

Measures

The base unit of analysis in this paper is the individual patent, which is associated with patent citation data, firm level data, and the IPC classes and patent family data associated with the patent itself.

Dependent variables

Whether a patent is a combination of exactly two technology areas is indicated by the dummy variable `DOUBLE_INVENTION`, which assumes the value of 1 if a patent combines, for example, technology areas A and B ('AB') without including C. To capture the capabilities required to generate combinations covering all three technology areas, a dummy variable is generated, `TRIPLE_INVENTION`, indicating whether the patent is a triple combination of A, B, and C technologies, ABC.

Independent variables

In the first model, in which we measure the effect of prior experience with combining technology areas on the likelihood of a given patent being a combination (e.g. AB), two variables are generated. `SAME_DOUBLE_KNOWL` is a count of the number of previous patents with the same combination as the given patent (e.g. AB), and gives an indication of the degree to which the firm is experienced with combining exactly these two technologies.

`RELATED_DOUBLE_KNOWL` is a count of the number of previous patents with other combinations of two technologies, different from that of the focal patent (e.g. AC or BC), which indicates the overall experience of the firm in combining technologies. To isolate the effects of having in-depth knowledge in the technology areas that constitutes a combination (e.g. A and B), `SAME_SINGLE_KNOWL` is a count of the number of previous patents covering one of the two technology areas. This measure includes both patents that utilize a single technology area and

patents that utilize combinations of technology areas, and provides a proxy for the overall experience of the firm with the individual technology areas.

In the second model, in which we measure the effect of having prior experience with either singles, doubles, or triples on the likelihood of a given patent being a triple invention (ABC), three variables are generated: TRIPLE_KNOWL is a dummy variable taking 1 if the firm has applied for patents combining all three technology areas within the last two years.

DOUBLE_KNOWL is a dummy variable showing whether the firm has had experience within the last two years prior to the patent application in combining two technological areas (AB, AC or BC). To isolate the effects of having a specialized knowledge base we use a dummy variable that takes 1 if the firm has applied for specialized patents within two years prior the patent application and 0 if not (SINGLE_KNOWL).

Control variables

We include firm level country dummies to control for regional differences. We use one dummy for EU countries as a benchmark and two dummies for the non-EU countries included in the analyses. The country dummies included are: DEVELOPING_COUNTRY (when firms are from China, Brazil, or South Africa, and OTHER_WESTERN countries, when firms originate from the United States (US) and Japan (EPO, USPTO et al.). The individual firm is associated with the country where it is listed at the stock exchange.

At the firm level, control variables are included for firm size as the log of the number of employees (LN_FIRMSIZE) and the level of firm internationalization as the log of the number of branch locations (LN_FIRMINTER). These are included as our data cover both large fully-integrated oil firms and smaller firms with a narrower focus. The size of a firm patent portfolio is included (FIRM_HCPATYEAR) as a proxy for the overall activity and potential R&D

expenditures of a firm in the given year within hydrocracking. At the patent level, a number of control variables are included: AGE_PAT is a count variable indicating the number of years since the patent was applied for. To control for patents receiving input from external scientific sources such as universities, the control NPL_CITS is included, a dummy variable measuring whether the focal patent cites non-patent related literature. We also control for the number of inventors (INVENTORCOUNT) as well as the number of assignees (ASSIGNEECOUNT).

The degree of firm specialization (SPECFIRM) is defined as the degree to which a firm has more patents within a single technology area or a combination of technology areas than the majority of the firm population in a given year. The majority of the population is defined as 90% of the firms, meaning that a firm needs to have a higher share of patents within a specific technology or technology combination than 90% of the population to be identified as a specialized firm. As the threshold is identified by the type of patent, it reflects the relative prevalence of the type. This variable changes over time, with some firms starting out as highly specialized but losing this label as they accumulate patents in different technology areas. The reverse also occurs, when a firm starts out with a broad patent portfolio, but switches to developing primarily a single technology or technology combination above all else.

METHODS

We articulate our empirical analysis in two steps. First, we explore the type of previous knowledge base that supports the generation of a patent based on the combination of two technology areas (DOUBLE_INVENTION), then we focus on the type of experience generating complex innovations (innovations covering three technological areas – TRIPLE_INVENTION). After presenting an initial descriptive analysis, we test our hypotheses using a series of

regression models. Since DOUBLE_INVENTION and TRIPLE_INVENTION, the two dependent variables that are used to describe the type of technology combination that characterizes the focal patents, are dichotomous, we employ logit models in the first part of our analysis. As the purpose of our study is to understand how a knowledge base can affect the potential for creating new inventions that are either double or triple inventions, our analyses are performed at the level of the individual invention. The model for DOUBLE_INVENTION can be written in the following way:

$$\Pr(\text{generate a double invention}=1 | sdk, rdk, sk, rdk*sk, c, \beta)$$

where the probability of generating a double invention depends on the firm's knowledge base, having same double knowledge, labeled *sdk*, related double knowledge, *rdk*, single specialized knowledge, *sk*, an interaction between related knowledge and single knowledge, *rdk*sk*, and a vector of the control variables, *c*. The model for TRIPLE_INVENTION can be written in the following way:

$$\Pr(\text{generate a triple invention}=1 | s, d, tr, c, \beta)$$

where the probability of generating a triple invention depends on the firm's knowledge base, prior knowledge in generating singles, labeled *s*, a knowledge base for generating doubles, *d*, a prior knowledge base for triple inventions, *tr*, and a vector of the control variables, *c*. We cluster by firm, and as the dependent variable in both models is binary it dictates a probit or logit model. There were no significant differences in the results; below, we have chosen to present the results of the logit model. The results presented are robust using the Huber-White-Sandwich correction.

RESULTS

Table 3 presents the descriptive results for the sample of 936 patents included in the analysis. Several interesting features emerge concerning the generation of innovations within hydrocracking technology. First, 11% of the innovations disclosed have combined refinery processing, catalyst preparation and feeds and products technologies (TRIPLE_INVENTION), indicating that these complex technology combinations are infrequent despite the maturity of the industry and the high firm age². Second, 54% of the patents are combinations of two technological areas (DOUBLE_INVENTION), which highlights how reliant this industry is on combining different technology areas for further innovation. The remaining 34% of the sample are innovations that rely on knowledge from a single technology area. The two variables that indicate the degree to which the firm has experience in combining technology areas are a) experience in combining same technologies (SAME_DOUBLE_KNOWL) and b) experience in combining related technologies (RELATED_DOUBLE_KNOWL), which are both skewed toward zero. Some of these zeros are inevitable as technology combinations are bound to be novel to the individual firm. The variable SAME_SINGLE_KNOWL, a proxy for the overall experience of the firm within a distinct technology area (A or B or C), is also highly skewed, which is expected due to the number of firms in our sample and the differences in their total patent portfolios. Some firms have many years of experience in working with a specific technology, whereas other firms are relatively smaller and have significantly less experience in developing innovations within a certain technological area.

 Insert Table 3 about here

² The majority of firms in the sample are more than 20 years old, and a large number are major oil firms that have operated for 50 years or more within the industry

In Table 4 below we present a correlation matrix for our dependent and independent variables. Some of our independent variables are highly correlated; however, this correlation is expected. For example, having prior experience within a certain technological field (SAME_SINGLE_KNOWL) is highly correlated (.8637 and .8417) with having prior experience with combinatorial innovations (RELATED_DOUBLE_KNOWL or SAME_DOUBLE_KNOWL). This is the case because SAME_SINGLE_KNOWL captures both prior experience with a technology area on its own or in combination with other technology areas.

Insert Table 4 about here

Similarly, R&D activity within hydrocracking technology in Year t (FIRM_HCPATYEAR) also increases the probability of having previous technological (SAME_SINGLE_KNOWL), or combinatorial experience within the field (SAME_DOUBLE_KNOWL) or a related area (RELATED_DOUBLE_KNOWL). No independent variables approach a 0.7 correlation with any of the dependent variables (DOUBLE_INVENTION and TRIPLE_INVENTION). However, we have been conscious of the high correlation between some of the variables when interpreting our results, and perform the necessary tests (including VIF and robustness checks) to ensure the reliability of our results.

Our models are estimated using a logit model; the complete results of the direct effects of our explanatory variables are shown in Tables 5 and 7. In these tables, the significance is based entirely on the values derived from normal Z-statistics rather than T-statistics. Therefore, to estimate the direct effect of our explanatory variables we investigate the marginal effects, as it is

a limited dependent variable we are examining the effects of (Wiersema and Bowen 2009). We use the following equation to estimate it:

Marginal Effect of X =

$$= \frac{\partial \Pr(\text{INVENTION} = 1 | V, \beta)}{\partial X}$$

$$= \frac{\partial \pi(V\beta')}{\partial X} = \pi(V\beta')\beta_x$$

This indicates that the marginal effect of X is proportional to its model coefficient, β_x . Here we can also see that the value and significance of X is not given by the estimates in the model coefficients. We therefore compute, at each observation, the value of the marginal effect, its standard error as well as Z-statistics, in order to test the significance of each value. However, initially we turn to the results presented in Table 5; here, the model significance is ($p < 0.000$), indicating a good fit against an intercept-only model. The r-square is 0.5226. We start our examination with the contribution of multi-technological knowledge bases that are related to two technologies to the development of an innovation coupling precisely those two domains (DOUBLE_INVENTION). Model 2 and Model 5 – which also takes into account the presence of other knowledge bases – show a positive effect of the variable SAME_DOUBLE_KNOWL. However, due to the fact that the estimates have been made in relation to a limited dependent variable, as explained above, we compute, at each observation, the value of the marginal effect, its standard error, and Z-statistics, in order to test the significance of each value to see whether the results of the table hold (Wiersema and Bowen 2009).

We investigate the marginal effects using graphical plots in figure 4. The blue dots represent marginal effect values (recorded on the left axis), and the red dots represent Z-values (as

recorded on the right axis). As shown in the graphical representation the marginal effects of SAME_DOUBLE_KNOWL in generating a DOUBLE_INVENTION range from 0 to 10.8549. The Z-values associated with most marginal effects exceed 1.28 (0.10) in absolute value, except at very high and very low probabilities of generating a DOUBLE_INVENTION. Since the slope of the logistic distribution approaches zero at the extremes, the lack of significance at the extremes is expected (Wiersema and Bowen 2009). The analysis of the graphical representation together with the logistic regression in the table above therefore indicates a positive significant relationship between having a knowledge base in the same double inventions and generating an invention based on double technologies, providing support for H1a. Furthermore, to investigate the relationship between same and related knowledge bases we did a Wald test. The results are presented in Table 6; they indicate that same knowledge is significantly more important in generating new inventions than related knowledge.

 Insert Tables 5&6 about here

In the same way, we investigate whether having the same knowledge base counts for generating the very complex TRIPLE_INVENTION. Here, the results of Models 10 and 11 in Table 7 indicate that hypothesis H1b would be confirmed. In the graphical plot in Figure 4b, we present how having a prior knowledge base in generating triple inventions has a marginal effect on the probability of generating a TRIPLE_INVENTION. The figure shows that for the lower values of the dependent variable, there are more insignificant observations, which are, however, significant at the mean; this is probably due to the effect of the vector. Therefore, the positive impact of TRIPLE_KNOWLEDGE on generating new TRIPLE INVENTION is not strong, except when the value of the dependent variable is above approximately 0.2. In this case, the results are

constantly significant, but very few observations fall within this value distribution. This means that having a knowledge base of experience within the same area has a positive and significant effect on the firm generating yet another invention of the same kind, and this accounts for both levels of complexity: lower, when the invention is based on a double, or higher, when the invention is based on a triple invention. However, the effect is much stronger for double inventions than for triple inventions, as the triple inventions only remain constantly significant when the value of the dependent variable is above 0.2.

The prediction of H2, that the probability of a firm generating a patent encompassing two technology domains is positively associated with the previous experience in patenting across the related two technology domains can be tested only with regard to inventions combining two technology domains. In Models 3, 5 and 6, as well as in the graphical analysis in Figure 4c, we find support for the hypothesis. In other words, we find that combinatorial innovations in a given domain benefit from the experience of a firm in a related domain, signaling the existence of combinatorial knowledge bases, also when controlling for the effect of the same knowledge base.

As regards hypotheses 3, we investigate the proposed negative effect of having specialized knowledge bases in generating both double inventions (H3a) and triple inventions (H3b). In regard to H3a, which concerns the role of a specialized technological knowledge base, singles, Model 4 shows a positive effect of prior experience in the domains that constitute a combinatorial technology. Such an effect should be considered spurious, as the relationship turns negative and significant in Models 5 and 6, where we introduce the variables related to double inventions, the multi-technological knowledge bases, as well as the interaction effect. This means that for any level of multi-technological knowledge bases, increasing focus on specialized technological knowledge bases deteriorates a firm's ability to innovate across technological

domains. This is consistent with the concept of “core rigidities”. The graphical plot in Figure 4d presents the family of marginal effects and the associated Z-values for SAME_SINGLE_KNOWL. The blue dots represent values of marginal effect (recorded on the left axis) and the red dots represent Z-values (as recorded on the right axis). The marginal effects of SAME_SINGLE_KNOWL range from 0 to -0.4944. As shown in the graphical representation, the Z-values associated with most marginal effects exceed 1.28 (0.10) in absolute value except at very high and very low probabilities of generating a DOUBLE_INVENTION. In H3b we hypothesized that the probability of a firm generating a highly complex innovation, a triple invention, would be negatively associated with the firm’s previous experience in patenting in the individual constituent technological domains. We did not find support for this hypothesis in our main models as presented in Table 6. The results show negative coefficients, which are, however, not significant, and in a graphical presentation of the effect no observations are significant. Our robustness checks, which consider only the experience of the last year (whether the firm has had experience with specialized knowledge within the last year prior to the patent application), did, however, produce results. While we find strong support for H3a, hypothesis H3b is only confirmed if we take into account a knowledge base for the last year. This means that only for the level of double invention does increasing focus on a specialized technological knowledge base deteriorate a firm’s ability to innovate across technological domains. This is consistent with the “core rigidities” concept, evidently pointing in the direction that specialization penalizes the firm when it wants to generate inventions across technological knowledge bases.

Insert Table 7 about here

To test the hypothesis that the relationship between having a related knowledge base and generating a double invention is negatively moderated by having a specialized knowledge base, we follow the method presented by Ai and Norton (2003) on true interaction effects. We find that the interaction effect is statistically significant in Model 6 (-0.005), which examines the marginal effect of SAME_SINGLE_KNOWL (moderating variable) on the relationship between RELATED_DOUBLE_KNOWL and generating a DOUBLE_INVENTION. As with the marginal effects examined in the main effects, we apply the graphical plot present in figure 4 to investigate each observation. As shown, the value, sign, and significance of the true interaction effect differ greatly over the range of predicted values for DOUBLE_INVENTION. The values of the true interaction range from -.0296919 to .029494; the Z-values range from -2.680328 to 2.792779. Model 6 shows that the interaction effect is negative and significant; however, figure 5 shows a more complex relationship. As shown in the figure, the interaction effect is mainly positively significant at lower value of doubles, when the probability of generating a double is below .4; however, there are significant dots after .65 also. Even though the effect is strongly negative, there are a few dots that are also significantly positive. The results provide support for a hypothesis stating that having a high number of specialized knowledge bases has a negatively moderating effect on the likelihood of the firm generating new doubles on a prior related knowledge base.

Insert Figure 4 about here

ROBUSTNESS

To further examine the robustness of our findings we also estimated models in which we controlled for time compression. It has been argued that time compression (Dierickx and Cool

1989) and the related absorptive capacity (Cohen and Levinthal 1990) have an impact on the process in which learning takes place in organizations. To test this effect, we added a variable to our model measuring the time (counted in years) from which the firm last made an invention of the same technological type (A,B, C, AB etc). Using such a control variable limits the number of observations to 832, as the observations in which the firm has not had any prior experience within the field are generated as missing variables. The variable TIME_COMPRESSION is positive and significant in generating TRIPLE_INVENTION (.028), indicating that the longer the time since the creation of the last triple invention, the higher the probability of generating one more, TIME_COMPRESSION is not significant in generating DOUBLE_INVENTION (.13). Adding the variable as a control did not change our findings with regard to our explanatory variables.

In the second model, we also investigated the findings when count variables are used for our three explanatory variables (SINGLE_KNOWL, DOUBLE_KNOWL and TRIPLE_KNOWL). The results were confirmed in relation to Model 11; however, the model had issues with high multicollinearity, suggesting a bad fit. Using count variables in Model 8 for SINGLE_KNOWL and in Model 9 for DOUBLE_KNOWL changed the results. In these models, when the degree of prior experience is taken into account, the results became significant and positive. However, as a result of the multicollinearity issues in Model 11, this may be due only to the strong relationship between having a great deal of experience with one of the two variables and having a great deal of experience with triple inventions.

Finally, we also tested for the effects of time compression in Model 2 to investigate the effects on generating TRIPLE_INVENTION by changing the lagged time to one year instead of two years. Our main results were confirmed, as TRIPLE_INVENTION proved again to be

significantly positive; however, having SINGLE_KNOWL became significant and negative. This result suggests that very accurate timing is an issue in learning with specialized knowledge and indicates that reaping the rewards of specialized knowledge is more dependent on timing compared to both knowledge combining technological areas and related knowledge bases.

DISCUSSION AND CONCLUSION

The findings reported above provide support for our contention that innovations characterized by more complex configurations rest upon the prior experience of the firm in *combining* technological knowledge rather than upon the firm's prior knowledge of the technologies that are combined. This puts firms with a specialized knowledge base at a disadvantage in complex technological domains, as these firms lack the capability to generate combinative knowledge. Below, we explore the implications of our findings for innovation literature and its value, also for studies of organizational learning. We also reflect on the new methodology of IPC codes, and speculate as well on some useful directions for future research.

In line with RBV and organizational learning, the initial part of our analysis provides support for the contention that the firm undergoes a process of organizational learning (Argote 1993), in which the firm accumulates different types of knowledge to create a knowledge base that underpins the ability to develop new innovations (Penrose 1959; Nelson and Winter 1982; Wernerfelt 1984). The knowledge base we explore is that of technical capabilities and combinative capabilities (Kogut and Zander 1992) as well as the rareness of these capabilities (Barney 1991; Danneels 2007). The analyses presented in this paper suggest that technical capabilities are less important than combinative capabilities in creating innovations based on a combination of technology areas. This study thus confirms some of the prior studies in organizational learning literature indicating that not all learning is equally difficult and therefore

equally valued (Eggers 2012). This study also reaffirms some of the main conclusions of learning and organizational literature, as it emphasizes that the ability to develop more difficult innovations is dependent on prior learning in an organization, that learning of combinative capabilities can be done in related areas, and that this increases the probability of creating more difficult innovations (Schilling, Vidal et al. 2003; Clark and Huckman 2009). However, we add to the discussion on related combinative capabilities by providing evidence that the effects of related combinative capabilities are dissolved when the firm learns the specific combinative capability in question. Related combinative capabilities can therefore be an important step on the way to learning, but have demarcated effects. However, as the technological complexity of a combination increases, the effect of related combinative capabilities is diminished. When firms attempt to generate a patent encompassing all three technology domains – the most complex type of patent – only prior experience in combining all three technology domains has a positive effect. Furthermore, experience within the individual constituent technology domains has a negative effect on the ability of the firm to generate patents encompassing all three technology domains. This highlights the steep learning curve present in complex technological domains, where only prior experience or pure serendipity will help.

Finally, this also leads us to a discussion of utilizing IPC codes as patent value indicators. We propose that the number of IPC codes in a patent is an imprecise way of measuring the technology scope of patents. When conducting research using IPC codes, we must look beyond them as purely quantitative measures and explicitly understand the meanings of each IPC code. When this approach is taken, IPC codes can provide a better understanding of the technological position and capabilities of the firm, both in relation to competitors, the industry of the firm, and other industries.

FIGURES AND TABLES

Figure 1 Exemplification of the hierarchy of combinatorial innovations

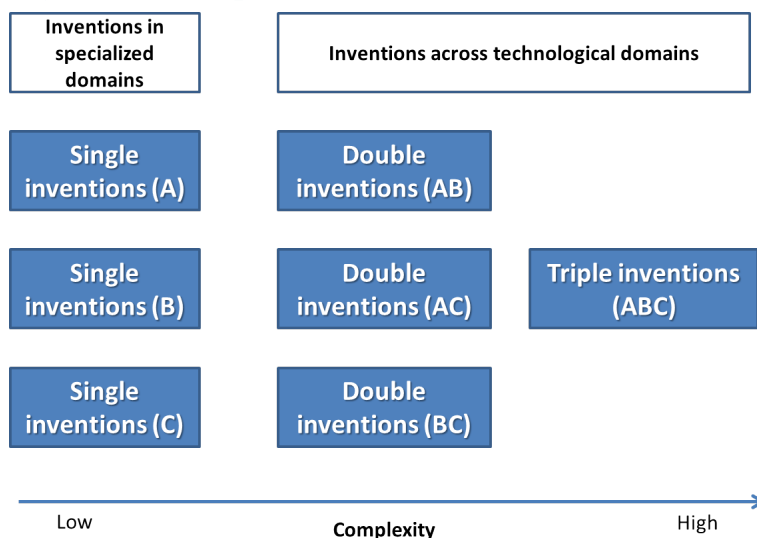


Figure 2: Hypothesized relationship between degree of complexity of combinatorial innovations and learning approach.

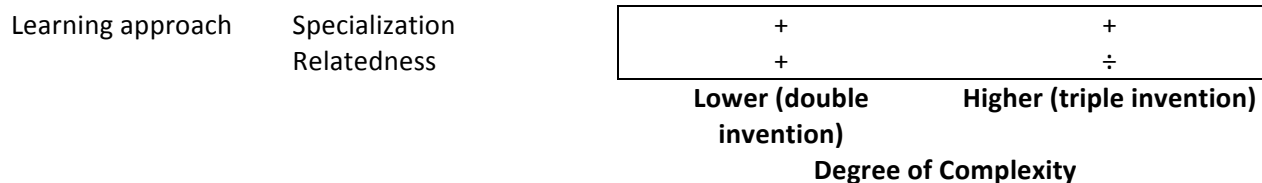


Figure 3 – Patent applications annually within hydrocracking and accumulated hydrocracking patents

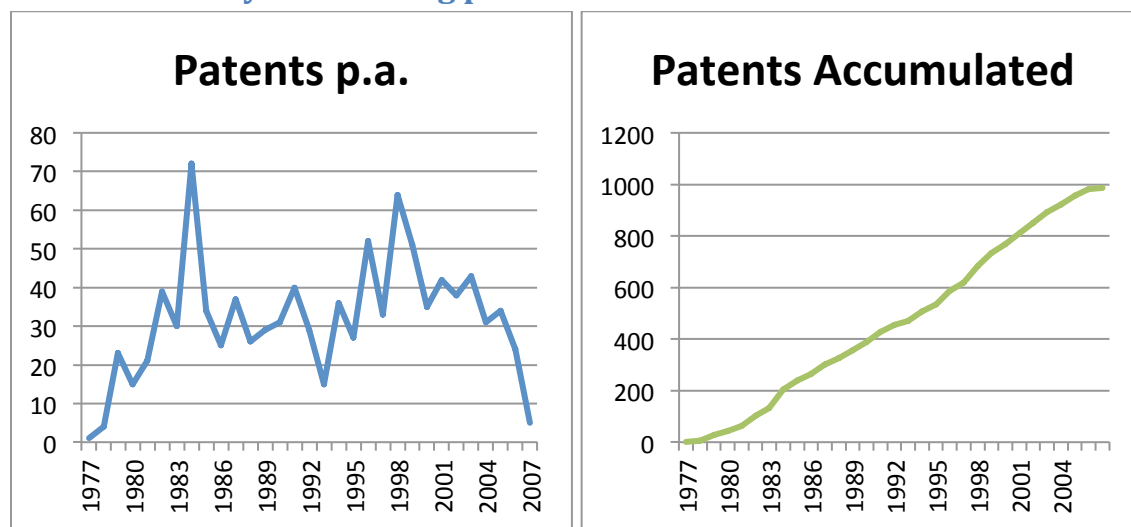


Table 1 – Hydrocracking technology areas and IPC classes³

Technology area	Associated IPC classes	Excluded IPC classes
Process technologies (A)	C10G-065/00 B01J-008/00	
Catalyst preparation (B)	B01J-021/00 to B01J-049/00	B01J-023/76 B01J-029/00
Feeds and products (C)	C10G-045/00 C10G-047/00 C10G-049/00	C10G-045/44 C10G-045/54 C10G-045/58 C10G-047/24-30

Table 2 – Distribution of patents and firm-year observations

Firm name	Total patents	Total years
ExxonMobil	265	30
Shell	190	29
IFP Energies Nouvelles	104	24
Chevron	90	26
UOP	73	23
Dow Chemical	32	8
Albemarle	20	9
Japan Energy	20	12
British Petroleum	19	12
Conoco	17	10
Sinopec	16	10
Total	15	10
Akzo	14	10
BASF	14	10
Mitsubishi	14	6
Ashland	13	4
Grace	13	9
Eurecat	12	9
Eni	11	6
Veba	11	9
IdemitsuKosan	9	7
Haldor Topsoe	7	6
NipponKetjen	7	5
Kellogg Co.	6	5
Petrobras	5	4
Sasol	5	4
Total number of patents	936	

³ IPC classes ending in /00 signify that all nine-digit subclasses within the seven-digit class are included, unless otherwise noted.

Table 3 Descriptive statistics					
Variable	Obs	Mean	Std. Dev.	Min	Max
-----+-----					
TRIPLE_INV~N	936	.1143162	.3183651	0	1
DOUBLE_INV~N	936	.5405983	.4986155	0	1
SINGLE_KNOWL	936	.8247863	.380353	0	1
DOUBLE_KNOWL	936	.7307692	.4437972	0	1
TRIPLE_KNOWL	936	.3344017	.4720331	0	1
-----+-----					
RELATED_DO~L	936	13.40812	21.55927	0	118
SAME_DOUBL~L	936	9.752137	15.00911	0	77
SAME_SINGL~L	936	51.05876	69.6081	0	282
_SINGLE_RE~E	936	1946.33	4083.97	0	24426
ASSIGNEECO~T	936	1.839744	.9545621	1	10
-----+-----					
AGE_PAT	936	13.82372	7.775452	0	29
SINGLE_A	936	.0512821	.2206904	0	1
SINGLE_B	936	.2104701	.4078605	0	1
SINGLE_C	936	.0833333	.2765332	0	1
DOUBLE_AB	936	.0363248	.187197	0	1
-----+-----					
DOUBLE_AC	936	.1314103	.3380291	0	1
DOUBLE_BC	936	.3728632	.4838247	0	1
NPL_CITS	936	.0854701	.2797294	0	1
LN_FIRMINT~T	936	1.433886	1.813451	0	6.302619
LN_FIRMSIZE	936	8.007897	3.687638	0	11.71198
-----+-----					
INVENTORCO~T	936	4.585653	4.321647	1	74
FIRM_HCPAT~R	936	8.202991	7.236674	1	25

Table 4 - Correlation matrix							
	TRIPLE~N	DOUBLE~N	SINGLE~L	DOUBLE~L	TRIPLE~L	RELATE~L	SAME_D~L
-----+-----							
TRIPLE_INV~N	1.0000						
DOUBLE_INV~N	-0.3897	1.0000					
SINGLE_KNOWL	0.0508	0.0319	1.0000				
DOUBLE_KNOWL	0.1045	0.0543	0.7593	1.0000			
TRIPLE_KNOWL	0.2649	-0.0691	0.3267	0.4302	1.0000		
RELATED_DO~L	0.1965	0.3054	0.2444	0.2912	0.4909	1.0000	
SAME_DOUBL~L	0.0100	0.4438	0.2485	0.2942	0.3814	0.6979	1.0000
SAME_SINGL~L	0.3843	0.2628	0.2797	0.3351	0.4910	0.8417	0.8637
_SINGLE_RE~E	0.2437	0.1746	0.2034	0.2553	0.4732	0.9311	0.7384
ASSIGNEECO~T	0.0111	0.0294	-0.0686	0.0066	0.0930	0.0357	0.0287
AGE_PAT	-0.1327	-0.0444	-0.0687	-0.1480	-0.3759	-0.3557	-0.2700
SINGLE_A	-0.0835	-0.2522	-0.0330	-0.0554	-0.0005	-0.1447	-0.1511
SINGLE_B	-0.1855	-0.5601	-0.0447	-0.1061	-0.1549	-0.3213	-0.3357
SINGLE_C	-0.1083	-0.3271	-0.0237	-0.0174	0.0485	-0.1876	-0.1960
DOUBLE_AB	-0.0698	0.1790	-0.0457	-0.0366	-0.0408	0.1121	-0.0607
DOUBLE_AC	-0.1397	0.3586	0.0212	0.0436	0.1466	0.4746	0.0943
DOUBLE_BC	-0.2770	0.7108	0.0357	0.0397	-0.1578	-0.0602	0.4150
NPL_CITS	-0.0498	0.0058	-0.0501	-0.0729	-0.0304	-0.0384	0.0270
LN_FIRMINT~T	-0.0667	-0.0108	-0.1365	-0.2686	-0.1463	-0.1990	-0.1830
LN_FIRMSIZE	0.0232	-0.0693	-0.0861	-0.1296	0.0926	0.0106	0.0387

INVENTORCO~T		0.0360	0.0170	0.0339	0.0801	0.1181	0.1547	0.1286
FIRM HCPAT~R		0.0981	0.0019	0.4015	0.4566	0.3549	0.2858	0.2125
		SAME S~L	~RELA~E	ASSIGN~T	AGE PAT	SINGLE A	SINGLE B	SINGLE C
-----+								
SAME SINGL~L		1.0000						
SINGLE RE~E		0.8789	1.0000					
ASSIGNEECO~T		0.0319	0.0576	1.0000				
AGE PAT		-0.2895	-0.3170	-0.3378	1.0000			
SINGLE A		-0.1706	-0.1109	-0.0676	-0.0240	1.0000		
SINGLE B		-0.3789	-0.2462	-0.0094	0.1982	-0.1200	1.0000	
SINGLE C		-0.2213	-0.1438	0.0020	-0.0404	-0.0701	-0.1557	1.0000
DOUBLE AB		-0.0039	0.0394	-0.0033	-0.0088	-0.0451	-0.1002	-0.0585
DOUBLE AC		0.1349	0.2689	0.0355	-0.2117	-0.0904	-0.2008	-0.1173
DOUBLE BC		0.1782	-0.0232	0.0068	0.1056	-0.1793	-0.3981	-0.2325
NPL CITS		-0.0191	0.0008	0.1395	-0.0506	-0.0711	0.0390	0.0461
LN FIRMIN~T		-0.2447	-0.2379	-0.0465	0.0474	-0.0249	0.0927	-0.0204
LN FIRMSIZE		0.0644	0.0200	0.1709	0.0374	-0.0006	0.0193	0.0703
INVENTORCO~T		0.1365	0.1884	0.3035	-0.3093	-0.0472	-0.0621	0.0572
FIRM HCPAT~R		0.2936	0.2513	-0.0702	0.0466	-0.0119	-0.0543	-0.0266
		DOUBLE~B	DOUBL~AC	DOUBL~BC	NPL CITS	LN FIR~T	LN FIR~E	INVENT~T
-----+								
DOUBLE AB		1.0000						
DOUBLE AC		-0.0755	1.0000					
DOUBLE BC		-0.1497	-0.2999	1.0000				
NPL CITS		-0.0594	-0.0737	0.0804	1.0000			
LN FIRMIN~T		0.0551	-0.0442	-0.0016	0.0447	1.0000		
LN FIRMSIZE		0.0111	-0.0135	-0.0663	0.0695	0.2784	1.0000	
INVENTORCO~T		-0.0078	0.0315	-0.0014	0.0921	-0.0822	0.0016	1.0000
FIRM HCPAT~R		-0.0299	0.0385	-0.0134	-0.1264	-0.2515	-0.0367	-0.0010
		FIRM H~R						
-----+								
FIRM HCPAT~R		1.0000						

<i>Table 5: Logit</i>						
<i>Dependent variable is DOUBLE_INVENTION</i>						
<i>Positive coefficients indicate increased probability that innovations with a combination will be innovated.</i>						
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
DOUBLE_INVEN~N						
SAME_DOUBLE~L		0.160***			1.445***	1.368***
		[0.059]			[0.246]	[0.217]
RELATED DOUB~L			0.051***		0.266***	0.264***
			[0.018]		[0.083]	[0.073]
SAME_SINGLE~L				0.012***	-0.199***	-0.128***
				[0.004]	[0.036]	[0.037]
SINGLE_RELA~E						-0.005***
						[0.001]
ASSIGNEECOUNT	0.068	0.245***	0.197**	0.157***	0.311**	0.338*
	[0.059]	[0.073]	[0.077]	[0.059]	[0.159]	[0.188]
AGE_PAT	-0.008	0.050**	0.040**	0.023	0.090***	0.088***
	[0.011]	[0.023]	[0.019]	[0.016]	[0.034]	[0.033]
NPL_CITS	0.049	0.078	0.139	0.086	-0.085	0.125
	[0.212]	[0.355]	[0.278]	[0.256]	[0.531]	[0.513]
LN_FIRMINTER~T	0.013	0.098	0.066	0.087	-0.020	-0.042
	[0.044]	[0.085]	[0.061]	[0.056]	[0.151]	[0.177]
LN_FIRMSIZE	-0.040**	-0.075**	-0.062***	-0.070***	0.012	0.015
	[0.021]	[0.031]	[0.024]	[0.026]	[0.070]	[0.075]
INVENTORCOUNT	-0.001	-0.017	-0.011	-0.008	-0.019	0.004
	[0.009]	[0.027]	[0.024]	[0.018]	[0.027]	[0.019]
FIRM_HCPATYEAR	0.003	-0.049*	-0.030*	-0.024	-0.077***	-0.105***
	[0.008]	[0.027]	[0.016]	[0.019]	[0.015]	[0.027]
SPEC_FIRM	0.085	0.368	0.282	0.279	0.053	0.130
	[0.209]	[0.340]	[0.307]	[0.279]	[0.455]	[0.498]
Constant	0.415*	-1.201**	-0.709	-0.397	-2.582***	-2.872***
	[0.248]	[0.491]	[0.450]	[0.375]	[0.815]	[0.922]
Pseudo LL	-641.979	-481.653	-576.695	-596.163	-308.275	-267.969
No of Obs	936	936	936	936	936	936
Wald-Chi2	16.9584**	66.68478***	35.54628***	40.67581***	81.37533***	143.7592***
* p<0.1, ** p<0.05, *** p<0.01						

<i>Table 6:</i>	
. test SAME_DOUBLE KNOWL RELATED_DOUBLE KNOWL	
(1)	[DOUBLE_INVENTION]SAME_DOUBLE KNOWL = 0
(2)	[DOUBLE_INVENTION]RELATED_DOUBLE KNOWL = 0
chi2(2) =	56.60
Prob > chi2 =	0.0000

Table 7: Logit model					
<i>Dependent variable is TRIPLE_INVENTION</i>					
<i>Positive coefficients indicate increased probability that most complex innovations will be innovated.</i>					
	Model 7	Model 8	Model 9	Model 10	Model 11
TRIPLE_INVEN~N					
SINGLE_KNOWL		0.251			-0.762
		[0.465]			[0.882]
DOUBLE_KNOWL			0.598		0.567
			[0.399]		[0.737]
TRIPLE_KNOWL				1.362***	1.350***
				[0.251]	[0.293]
ASSIGNEECOUNT	-0.149*	-0.142*	-0.141*	-0.040	-0.051
	[0.082]	[0.083]	[0.081]	[0.096]	[0.096]
AGE_PAT	-0.066***	-0.065***	-0.062***	-0.039***	-0.039***
	[0.013]	[0.013]	[0.014]	[0.012]	[0.013]
NPL_CITS	-0.697	-0.699	-0.700	-0.708	-0.705
	[0.486]	[0.488]	[0.490]	[0.470]	[0.470]
LN_FIRMINTER~T	-0.104	-0.103	-0.107*	-0.068*	-0.070*
	[0.070]	[0.067]	[0.062]	[0.040]	[0.040]
LN_FIRMSIZE	0.034	0.034	0.037*	0.057***	0.058***
	[0.022]	[0.022]	[0.021]	[0.017]	[0.018]
INVENTORCOUNT	0.001	0.000	0.000	0.000	0.001
	[0.024]	[0.023]	[0.023]	[0.019]	[0.019]
FIRM_HCPATYEAR	0.031**	0.028*	0.022	-0.012	-0.012
	[0.015]	[0.015]	[0.015]	[0.016]	[0.015]
SPEC_FIRM	-0.286	-0.279	-0.239	-0.215	-0.199
	[0.467]	[0.465]	[0.460]	[0.382]	[0.388]
Constant	-1.267***	-1.499**	-1.796***	-2.469***	-2.256***
	[0.451]	[0.600]	[0.561]	[0.420]	[0.613]
Pseudo LL	-315.413	-315.235	-314.097	-303.790	-303.306
No of Obs	936	936	936	936	936
Wald-Chi2	64.05743***	68.08584***	67.25106***	122.2113***	134.3674***
* p<0.1, ** p<0.05, *** p<0.01					

Figure 4 Main effects

Figure 4a: Analysis of the marginal effect of SAME_DOUBLE_KNOWL on the probability of generating a DOUBLE

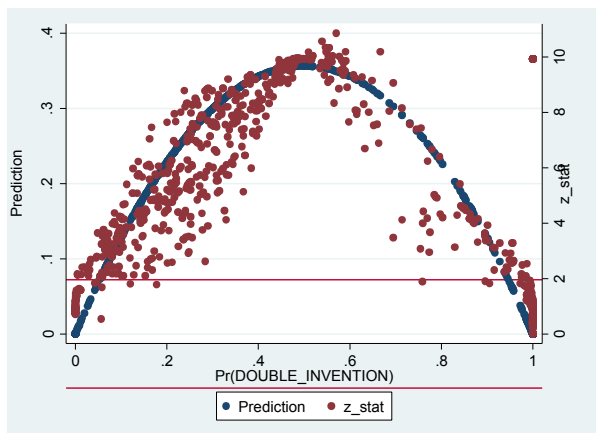


Figure 4b: Analysis of the marginal effect of TRIPLE_KNOWL on the probability of generating a TRIPLE

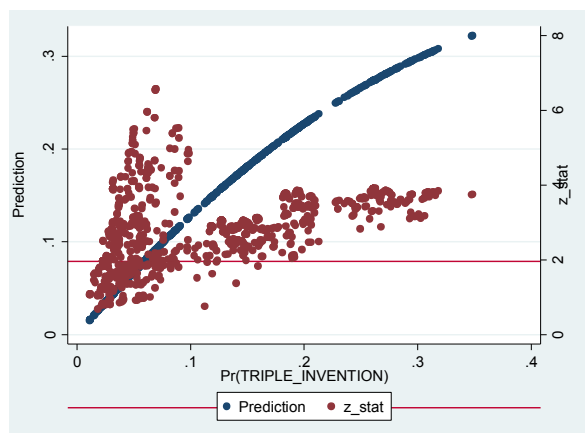


Figure 4c: Analysis of the marginal effect of RELATED_DOUBLE_KNOWL on the probability of generating a DOUBLE invention

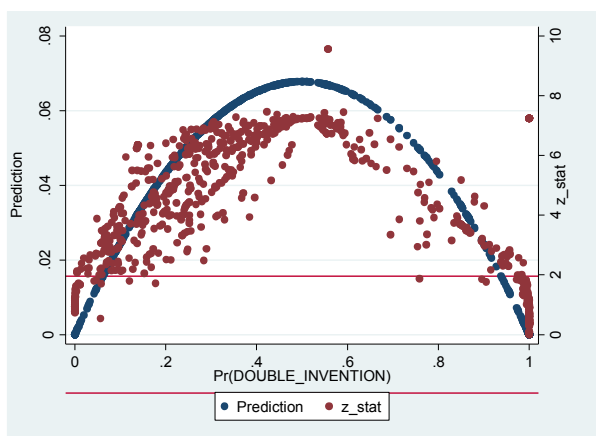


Figure 4d: Analysis of the marginal effect of SAME_SINGLE_KNOWL on the probability of generating a DOUBLE invention

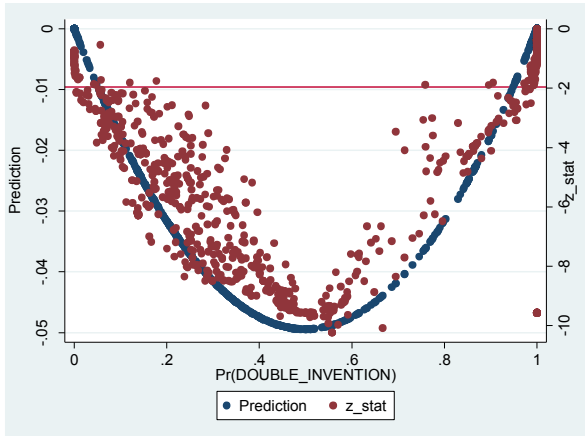
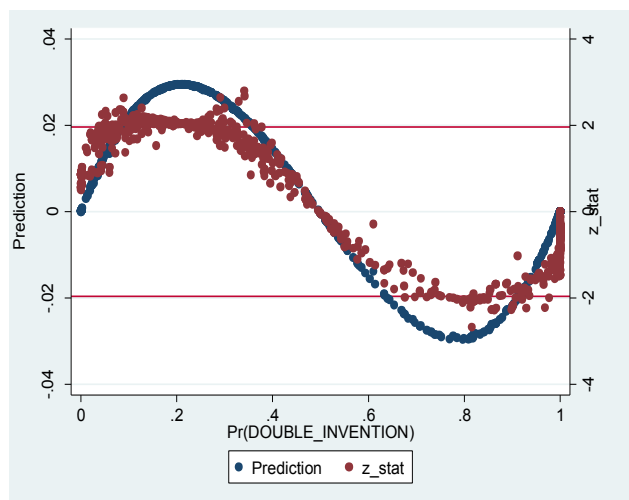


Figure 5 Moderating effects

Figure 5a: Analysis of the interaction effect of specialized knowledge (SAME_SINGLE_KNOWL) on a related knowledge base (RELATED_DOUBLE_KNOWL) and the probability of generating a DOUBLE invention



REFERENCES

- Ai, C. R. and E. C. Norton (2003). "Interaction terms in logit and probit models." *Economics Letters* **80**(1): 123-129.
- Argote, L. (1993). "GROUP AND ORGANIZATIONAL LEARNING-CURVES - INDIVIDUAL, SYSTEM AND ENVIRONMENTAL COMPONENTS." *British Journal of Social Psychology* **32**: 31-51.
- Argote, L. (1999). *Organizational Learning: Creating, Retaining and Transferring Knowledge.*, Kluwer Academic Publishers, Boston.
- Argyres, N. S. and B. S. Silverman (2004). "R&D, organization structure, and the development of corporate technological knowledge." *Strategic Management Journal* **25**(8-9): 929-958.
- Arthur, W. B. (1989). "COMPETING TECHNOLOGIES, INCREASING RETURNS, AND LOCK-IN BY HISTORICAL EVENTS." *Economic Journal* **99**(394): 116-131.
- Barney, J. (1991). "FIRM RESOURCES AND SUSTAINED COMPETITIVE ADVANTAGE." *Journal of Management* **17**(1): 99-120.
- Billon, A. and P.-H. Bigeard (2001). *Petroleum Refining*, Technip, Paris.
- Bingham, C. B. and J. P. Davis (2012). "LEARNING SEQUENCES: THEIR EXISTENCE, EFFECT, AND EVOLUTION." *Academy of Management Journal* **55**(3): 611-641.
- Boh, W. F., S. A. Slaughter, et al. (2007). "Learning from experience in software development: A multilevel analysis." *Management Science* **53**(8): 1315-1331.
- Cardinal, L. B., S. F. Turner, et al. (2011). "Organizing for Product Development Across Technological Environments: Performance Trade-offs and Priorities." *Organization Science* **22**(4): 1000-1025.

- Clark, J. and R. Huckman (2009). Broadening focus: Spillovers and benefits of specialization in the hospital industry.
- Cohen, W., R. Nelson, et al. (2002). "Links and Impacts: The influence of public research on industrial R&D." *Management Science* **48**(1).
- Cohen, W. M. and D. A. Levinthal (1990). "ABSORPTIVE-CAPACITY - A NEW PERSPECTIVE ON LEARNING AND INNOVATION." *Administrative Science Quarterly* **35**(1): 128-152.
- Danneels, E. (2002). "The dynamics of product innovation and firm competences." *Strategic Management Journal* **23**(12): 1095-1121.
- Danneels, E. (2007). "The process of technological competence leveraging." *Strategic Management Journal* **28**(5): 511-533.
- Dierickx, I. and K. Cool (1989). "ASSET STOCK ACCUMULATION AND SUSTAINABILITY OF COMPETITIVE ADVANTAGE." *Management Science* **35**(12): 1504-1511.
- Dietmar Harhoff, Frederic M Scherer, et al. (2003). "Citations, family size, opposition and the value of patent rights." *Research Policy* **32**(8): 1343-1363.
- Dosi, G. (1988). "SOURCES, PROCEDURES, AND MICROECONOMIC EFFECTS OF INNOVATION." *Journal of Economic Literature* **26**(3): 1120-1171.
- Eggers, J. P. (2012). "All experience is not created equal: learning, adapting, and focusing in product portfolio management." *Strategic Management Journal* **33**(3): 315-335.
- EPO, USPTO, et al. (1998). Trilateral Project 24.1 Biotechnology Comparative Study on Biotechnology Patent Practices Comparative Study.
- Fleming, L. (2001). "Recombinant uncertainty in technological search." *Management Science* **47**(1): 117-132.
- Fleming, L. and O. Sorenson (2001). "Technology as a complex adaptive system: evidence from patent data." *Research Policy* **30**(7): 1019-1039.
- Fleming, L. and O. Sorenson (2004). "Science as a map in technological search." *Strategic Management Journal* **25**(8-9): 909-928.
- Henderson, R. M. and K. B. Clark (1990). "Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms." *Administrative Science Quarterly* **35**(1): 9-30.
- 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.
- Kogut, B. and U. Zander (1992). "Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology." *Organization Science* **3**(3): 383-397.
- Kogut, B. and U. Zander (1996). "What firms do? Coordination, identity, and learning." *Organization Science* **7**(5): 502-518.
- Kotha, R., Y. F. Zheng, et al. (2011). "ENTRY INTO NEW NICHEs: THE EFFECTS OF FIRM AGE AND THE EXPANSION OF TECHNOLOGICAL CAPABILITIES ON INNOVATIVE OUTPUT AND IMPACT." *Strategic Management Journal* **32**(9): 1011-1024.
- Leonardbarton, D. (1992). "CORE CAPABILITIES AND CORE RIGIDITIES - A PARADOX IN MANAGING NEW PRODUCT DEVELOPMENT." *Strategic Management Journal* **13**: 111-125.
- Lerner, J. (1994). "THE IMPORTANCE OF PATENT SCOPE - AN EMPIRICAL-ANALYSIS." *Rand Journal of Economics* **25**(2): 319-333.
- Levinthal, D. A. and J. G. March (1993). "THE MYOPIA OF LEARNING." *Strategic Management Journal* **14**: 95-112.
- Levitt B and March J G (1988). "ORGANIZATIONAL LEARNING." *Annual Review of Sociology* **14**: 319-340.
- Leydesdorff, L. (2008). "Patent classifications as indicators of intellectual organization." *Journal of the American Society for Information Science and Technology* **59**(10): 1582-1597.

- Makri, M., M. A. Hitt, et al. (2010). "COMPLEMENTARY TECHNOLOGIES, KNOWLEDGE RELATEDNESS, AND INVENTION OUTCOMES IN HIGH TECHNOLOGY MERGERS AND ACQUISITIONS." Strategic Management Journal **31**(6): 602-628.
- March, J. G. (1991). "EXPLORATION AND EXPLOITATION IN ORGANIZATIONAL LEARNING." Organization Science **2**(1): 71-87.
- Mayer, K. J. and N. S. Argyres (2004). "Learning to contract: Evidence from the personal computer industry." Organization Science **15**(4): 394-410.
- Miller, D. (2003). "An asymmetry-based view of advantage: Towards an attainable sustainability." Strategic Management Journal **24**(10): 961-976.
- Nelson, R. and S. Winter (1982). An Evolutionary Theory of Economic Change, Cambridge, Harvard University Press.
- Penrose, E. (1959). "Theory of the Growth of the Firm."
- Rosenkopf, L. and A. Nerkar (2001). "Beyond Local Search: Boundary-spanning, Exploration, and Impact in the optical disc industry." Strategic Management Journal **22**(4): 287-306.
- Sampson, R. C. (2005). "Experience effects and collaborative returns in R&D alliances." Strategic Management Journal **26**(11): 1009-1031.
- Schilling, M., P. Vidal, et al. (2003). "Learning by doing something else: variation, relatedness and the learning curve." Management Science **49**(1).
- Schilling, M. A. and E. Green (2011). "Recombinant search and breakthrough idea generation: An analysis of high impact papers in the social sciences." Research Policy **40**(10): 1321-1331.
- Schilling, M. A., P. Vidal, et al. (2003). "Learning by doing something else: Variation, relatedness, and the learning curve." Management Science **49**(1): 39-56.
- Schumpeter, J. A. (1939). Business Cycles.
- Simon, H. (1962). "The Architecture of Complexity." Proceedings of the American Philosophical Society **106**(6): 467-482.
- Simon, H. A. (1969). The Sciences of the Artificial. In Frenken, K. (2006). A fitness landscape approach to technological complexity, modularity, and vertical disintegration, Structural Change and Economic Dynamics **17**(3): 288-305., MIT Press, Cambridge, MA & London.
- Sorensen, J. B. and T. E. Stuart (2000). "Aging, obsolescence, and organizational innovation." Administrative Science Quarterly **45**(1): 81-112.
- Stuart, T. E. and J. M. Podolny (1996). "Local search and the evolution of technological capabilities." Strategic Management Journal **17**: 21-38.
- Van den Bosch, F. A. J., H. W. Volberda, et al. (1999). "Coevolution of firm absorptive capacity and knowledge environment: Organizational forms and combinative capabilities." Organization Science **10**(5): 551-568.
- Wernerfelt, B. (1984). "A RESOURCE-BASED VIEW OF THE FIRM." Strategic Management Journal **5**(2): 171-180.
- Wiersema, M. F. and H. P. Bowen (2009). "The use of limited dependent variable techniques in strategy research: Issues and methods." Strategic Management Journal.
- Yayavaram, S. and G. Ahuja (2008). "Decomposability in knowledge structures and its impact on the usefulness of inventions and knowledge-base malleability." Administrative Science Quarterly **53**(2): 333-362.
- Zander, U. and B. Kogut (1995). "KNOWLEDGE AND THE SPEED OF THE TRANSFER AND IMITATION OF ORGANIZATIONAL CAPABILITIES - AN EMPIRICAL-TEST." Organization Science **6**(1): 76-92.