Impact of knowledge search practices on the originality of inventions: a study in the oil & gas industry

Quentin Plantec, Pascal Le Masson, Benoit Weil

IMPACT OF KNOWLEDGE SEARCH PRACTICES ON THE ORIGINALITY OF INVENTIONS: A STUDY IN THE OIL & GAS INDUSTRY

Quentin PLANTEC[1,2], Pascal LE MASSON[1] & Benoît WEIL[1]

Academy of Management (AoM) Conference 2020, Vancouver, Canada

1. MINES ParisTech, PSL University, Centre for Management Science (CGS), i3 UMR CNRS, 60 Boulevard St Michel, 75006 Paris, France[1]

2. Institut National de la Propriété Intellectuelle (INPI), 15 rue des Minimes, 92677 Courbevoie CEDEX, France

ABSTRACT

The paper suggests a new taxonomy of knowledge search modes to describe the creative process of new invention design, in particular how firms combine knowledge components from their own knowledge base—taking into account both the components and the structures of knowledge bases—with those from newly acquired or newly internally developed. Using network theory techniques, we defined four knowledge search modes: (1) refinement, (2) clustering, (3) absorption and (4) recomposition. We conducted an exploratory study on the oil & gas industry, reviewing 50,776 utility patents filed by 16 major firms between 1989 and 2016. The results showed, first, that firms relied to varying extents on different knowledge search modes in their invention design processes. Second, reviewing the technological originality of the designed inventions showed that simply absorbing new knowledge components, without major changes in knowledge base structure, was associated with low technological originality, but constituted one of the main knowledge search modes used by the analyzed firms. In contrast, major changes in knowledge base structure favored technological originality, with or without new knowledge components, but were nevertheless the least used mode. Understanding organizational learning practices associated with the phenomena described here can foster innovation performance in firms.

Keywords: *Knowledge search; Patent; Oil & gas; technological originality; knowledge base.*

Ccorresponding author: Quentin Plantec, quentin.plantec@mines-paristech.fr, Centre for Management Science (CGS), i3 UMR CNRS, 60 Boulevard St Michel, 75006 Paris, France

Impact of Knowledge Search Practices on the Originality of Inventions: a Study in the Oil & Gas Industry

Technological novelty is an important driver of innovation (Arthur, 2007; Fleming & Sorenson, 2004), which is presumed to have a large impact on organization competitiveness and market position (Cooper & Schendel, 1976; Henderson & Clark, 1990). However, innovation is a complex process with high uncertainty across different stages, including the knowledge search supporting the design of new inventions. Knowledge search activities aimed at selecting and combining knowledge components to design new inventions (eg. Arts and Fleming, 2016; Chesbrough, 2003; Cohen and Levinthal, 1990; Hatchuel and Weil, 2003; Nelson and Winter, 1982; Nonaka, 1994). Therefore, one question for innovation management is to better understand how a particular set of knowledge components mobilized by a given inventor (or group of inventors) can conduct to design an original invention.

To help further the understanding of this issue, scholars have analyzed knowledge search practices from the perspective of the global technological landscape, exploring to what extent inventors were combining already known or new knowledge components and what the effects were of those practices on the value of the designed inventions (Lobo & Strumsky, 2019; Strumsky & Lobo, 2015; Verhoeven, Bakker, & Veugelers, 2016). Furthermore, scholars have also reviewed what were the effects on the quality of inventions to either, design it with knowledge components that the inventor already master or to explore new knowledge components to design it (Arts & Fleming, 2016; Fleming, 2001). Exploring new paths is leading, in average, to an increase of the technological originality while refinement of already mastered knowledge components increase the value and robustness of inventions (Arts & Fleming, 2016).

Here, we depart from those analyses by focusing only on firms' inventors. Hence, a firm-level analysis should be conducted because inventors of firms are embedded in particular ecosystems and then have distinct design capabilities (Ahuja & Lampert, 2001; Nooteboom, Van Haverbeke, Duysters, Gilsing, & van den Oord, 2007). The latter depends of the knowledge components mastered by the firm, referred here as firm's knowledge base (Grant, 1996). According to the literature on recombinant innovation (eg. Ahuja and Lampert, 2001), the knowledge of the firm is also embedded in its combinative capabilities (Kogut & Zander, 1992). Hence, the knowledge base has a particular structure; that is, different couplings exist among all knowledge components mastered by the firm based on the ability of the firm to combine them (Le Masson, Weil, & Hatchuel, 2010a; Yayavaram & Ahuja, 2008). As firms support their innovation efforts through distinctive innovation capabilities, team management practices, and exploration practices (Lawson & Samson, 2001; Le Masson, Weil, & Hatchuel, 2010b), firms have different knowledge base conducting to a plurality of design capabilities for their inventors.

Few scholars have successfully reviewed the role of changes in a firm's knowledge base—integration of new components and/or new combinations resulting from knowledge search activities—on the global firm's innovation performances (Yayavaram & Ahuja, 2008; Yayavaram & Chen, 2015). They have shown that changes in the combinations between existing knowledge components of firms hurt innovation performance, while new combinations between new and established knowledge components favor innovation performance. Here, we focus only on a particular element of innovation performance: the technological originality which is marked by the significant difference between combinations of underlying knowledge components embedded in a particular invention and the predominant design

Hence, our present contribution **aims to provide a detailed review of the plurality of knowledge search practices undertake by firms and to measure their direct effect on**

technological originality of the designed invention itself. It is supported by the analysis of firm's knowledge base structure including components, combinations and distance between those components that provide a more fine-grained proxy of the inventors' ecosystems. In that sense, we aim at bridging the literature on firms' knowledge base and that on inventors' knowledge search practices in global technological landscape.

In this study, we develop a taxonomy of four knowledge search modes that are used to support the design of novel inventions depending on the firm's knowledge base structure: (1) refinement, (2) clustering, (3) absorption, and (4) recomposition. Our taxonomy was implemented though an exploratory study of 16 major oil & gas firms involving a review of their patent portfolios between 1989 and 2016, comprising 50,776 utility patent applications. We show that, first, the knowledge search modes adopted by firms vary among time and across the firms in the sample. Second, however, refinement was the most used knowledge search mode in the oil & gas industry, but led to inventions with low technological originality. Third, the absorption mode, the second-most used, which focuses on the integration of new knowledge components without major changes in the firm's previous knowledge base structure, scored poorly in terms of technological originality. In contrast, the recomposition knowledge search mode, a design process that generates major changes in firms' knowledge base structure (with or without integration of new knowledge components), conduces to highly original inventions, but was nevertheless the least adopted among the analyzed firms, indicating a major organizational learning challenge. Fourth, firms somewhat rely on a clustering mode as its technological originality score is better than that of the absorption mode.

Finally, the research opens up promising new research paths for the innovation field, emerging from the implication that technological originality, when integrating new knowledge components, depends on implied changes in the firm's knowledge base structure. This also has consequences for organizational learning and innovation design, as it calls for taking into account the dual status of knowledge component distance, that is, considering it both from the firm's and the firm's inventor's points of view.

LITERATURE REVIEW

Knowledge search practices: (re)combination and the design of new inventions

Various streams of research on innovation have highlighted the importance of knowledge search activities, in areas such as organizational learning (e.g., Cohen and Levinthal, 1990; Nonaka, 1994), design science (e.g., Hatchuel and Weil, 2009, 2003), evolutionary theory (e.g., Nelson and Winter, 1982) or open innovation (e.g., Chesbrough, 2003; West and Bogers, 2014). For those authors, firms' knowledge search activities aim at solving problems by combining knowledge components to create new products.

Innovation scholars have mainly operated under the assumption that combining existing technological capabilities or knowledge components in novel ways (potentially including new knowledge for new combinations) is one of the principal sources of inventive novelty (Dosi, 1982; Fleming, 2001; Fleming & Sorenson, 2001; Hargadon & Sutton, 1997; Henderson & Clark, 1990; Schumpeter, 1934; Teece, 1996). Three main ways of combining knowledge can be identified: (1) combining new knowledge only (ie. distant knowledge), (2) designing new configurations of previously combined knowledge (ie. local knowledge), and (3) mixing new knowledge with previously combined knowledge (Fleming, 2001; Yayavaram & Chen, 2015).

Based on these theoretical assumptions, scholars have developed techniques based on patent analytics to explore how inventions are designed. Those models mainly use Patent Classification Codes as a proxy for knowledge components. For example, Lobo and Strumsky (2019) and Strumsky and Lobo (2015) used four categories to describe novel inventions: (1) origination if all the IPC codes used in a given patent have never been used before, (2) novel combination if at least one pair of IPC codes in a given patent contains a code that has never been used before, (3) combination if at least one pair of IPC codes has never been used before in a given patent, and finally, (4) refinement for cases where all IPC codes have already been used before in other patents. They showed that from 1980 to 2014 for United States Patent and Trademark Office (USPTO) patents, combinations and refinements accounted for 47% and 52% of all patents respectively, while origination and novel combinations were very rare. Verhoeven et al. (2016) used another taxonomy: (1) novelty in recombination, when a patent contains at least one pair of IPC codes that was previously unconnected, (2) novelty in technological origins, when a patent contains at least one pair of IPC codes between the focal patent and the patent cited as literature (backward citation) that was previously unconnected, and (3) novelty in scientific knowledge, based on the pair of IPC and Web of Science category classifications of scientific articles cited as literature in the patent (i.e., Non-Patent Literature). By using a sample of patent families, they showed that 7% of patent families scored on novelty in recombination, 22% scored on novelty in technological origins, and 11% of patent families that cited at least one scientific article scored on novelty in scientific knowledge. Overall, 25% of all patent families scored on at least one novelty criteria.

These valuables techniques help us better understand the nature of the knowledge search practices used by inventors to design new inventions and their occurrence in the global technological landscape. Furthermore, scholars have also developed more applied research conducting analyses at firm level, in order to help a given firm identify what new knowledge components could be combined with their established knowledge to design new inventions (e.g., Nakamura et al., 2015; Sarica et al., 2019). Thus, the design of new inventions is closely related to a firm's knowledge base, whose structure matters in this regard.

Firm's knowledge base—components and structures—and its interface with knowledge search practices

The firm's knowledge base describes what the firm knows (Grant, 1996; Jaffe, 1986; Kogut & Zander, 1992). Knowledge base theory has established that knowledge components are the preeminent resource of innovation activities of firms (Grant, 1996). Nevertheless, not only the knowledge components themselves matter: the structure of the knowledge base is also a key element (Kogut & Zander, 1992). Hence, the knowledge base "*can be characterized by the set of knowledge elements that is possesses and the relationships that it has forged between the knowledge domains to which these elements belong*" (Yayavaram & Chen, 2015: 377–378). As a result, firms' knowledge search practices are mainly linked to the firm's knowledge base: local knowledge corresponds to the firm's knowledge base at time *t*, while accessed distant knowledge components fuel the firm's knowledge base and are integrated at time *t+1*, as well as internally developed new knowledge generated through the local space (Kogut & Zander, 1992; Yayavaram & Ahuja, 2008; Yayavaram & Chen, 2015).

Developed patent techniques also exist to explore firms' knowledge base components and structures. At the adopted starting point, scholars used Patent Classification Codes to reveal the knowledge components used by a given firm. For example, Fleming and Sorenson (2001) used USPTO subclass references for all patents filed by a given firm to compute knowledge components of the firm's knowledge base. Yayavaram and Ahuja (2008) and Yayavaram and Chen (2015) provided a more advanced representation of a firm's knowledge base by focusing on knowledge components and couplings at a firm-level analysis. "*[It] suggests that for several reasons it is beneficial to conceptualize knowledge bases as networks of knowledge elements in which even the ties between knowledge elements are important, rather than as simply sets of individual elements.*" (Yayavaram & Ahuja, 2008: 357–358). Yayavaram and Ahuja (2008) showed that first, there is an inverse-U-shape between continuums of structures—from fully decomposed to an integrated knowledge base—and inventions' usefulness as measured through patent citations. Second, they highlighted that firms are pursuing two types of strategies: (1)

adding new knowledge components and (2) adding new combinations of existing knowledge components (i.e., new couplings). Furthermore, Yayavaram and Chen (2015) explored the extent to which changes in knowledge component couplings between (1) familiar knowledge components only and (2) both new knowledge and familiar knowledge components affects innovation performance, using knowledge base complexity as a moderating variable. By reviewing financial and patent data for 1,750 firms between 1976 and 2004, they showed that changes in coupling of knowledge with which the firm is familiar (i.e., local knowledge) hurt the firm's innovation performance, while new couplings, using both local and distant knowledge components, increase the firm's innovation performance. Drawing on the literature in architectural innovation (Henderson & Clark, 1990), they highlighted that changing couplings between local knowledge components in highly complex environments reflects that the firm has overcome interdependency issues and uncertainty regarding which couplings are valuable. This phenomenon then leads to improved innovation performance.

Sourcing knowledge locally or distantly and the quality of invention

As firms rely on different knowledge search modes to source and combine knowledge components in order to design new inventions, with various implications and dependencies regarding firm's knowledge base, one key question is: to what extent does a given firm need to rely on its previous stock of knowledge to foster the design of new inventions (Katila, 2002), that is, to what extent does it hold that "old is gold" (Nerkar, 2003)?

A wide stream of research argues that radical innovation necessarily presupposes the utilization of very recent knowledge (Henderson & Clark, 1990; Katila, 2002; O'Connor, 2008) and knowledge from different industries and technological domains (Dahlin & Behrens, 2005; Dosi, 1982; Katila & Ahuja, 2002; Nooteboom et al., 2007). In principle, such knowledge has not yet been integrated into the firm's knowledge base. By reviewing novelty through patent citations, Ahuja and Lampert (2001) effectively showed that having no backward citations, which can be interpreted as only using new knowledge components, is associated with more radical invention. Furthermore, Arts and Fleming (2016), relying on semantic analysis of patent data, showed that inventors who change fields are more likely to create novel patents due to their exposure to new knowledge components. In addition, drawing on state-of-the-art new scientific knowledge is also associated with the exploration of completely new innovation paths, such as the use of spintronic theories developed by 2007 Nobel laureate Albert Fert for semiconductors and the current industrial development on graphene by Nobel laureate Andre Geim (e.g., Fleming and Sorenson, 2004; Hatchuel et al., 2013). By measuring scientific article novelty and utilization in patent literature, Veugelers and Wang (2019) showed that utilizing more original scientific articles conduces to more novel patents (through patent citation evaluation). Finally, firms that build on newly developed knowledge components more frequently are able to better predict future technological advances and thereby design more novel inventions (Cohen & Levinthal, 1989). The frequent exposure to newly developed knowledge helps firms to sustain their cognitive capabilities, thereby helping future innovation (March, 1991).

Although relying on external knowledge is positively related to the originality of the design of a given invention, extensive reliance on local knowledge has been proven to be detrimental to the global innovation process of firms. First, prior learning and existing paradigms can cause learning myopia or learning traps (Levinthal & March, 1993) and constrain the direction of search due to cognitive path dependence, specific historical pathways, or fixation effects (Agogué et al., 2014; Kaplan & Tripsas, 2008; Sydow, Schreyögg, & Koch, 2009). Nevertheless, it has been underlined by many researchers that firms should exploit and build upon their knowledge and expertise (i.e., their knowledge base) to innovate; indeed, as technological progress is cumulative, inventors inevitably need to draw on their prior knowledge and expertise (Kline & Rosenberg, 1986). Using knowledge components which the firm is familiar with enhances the reliability of the design and its uniqueness, in particular if the firm has a distinctive knowledge base (Katila, 2002; March, 1991). Following a combinatorial logic, a larger knowledge base evidently give more rooms for combinatorial possibilities, but it can also enhance complexity due to interdependency of knowledge components (Fleming, 2001; Yayavaram & Ahuja, 2008).

March (1991) characterizes the essential problem as a choice between exploiting known knowledge or exploring new and distant possibilities. Literature on “innovation ambidexterity” has drawn on those two concepts, advocating for simultaneously pursuing both exploration and exploitation (e.g., Andriopoulos and Lewis, 2010; O’Reilly and Tushman, 2008; Raisch et al., 2009), but mainly through pursuing different projects on similar timeframes. There is also room for the combination of both new and old knowledge components at an invention level and not at a portfolio level (Hatchuel & Weil, 2003, 2009; Le Masson et al., 2010a; Strumsky & Lobo, 2015; Yayavaram & Ahuja, 2008; Yayavaram & Chen, 2015).

LITERATURE GAP & RESEARCH QUESTION

To design novel inventions, firms need to combine knowledge components based on a continuum of search practices, depending on the extent of their reliance on familiar knowledge components. In our approach, we are bridging the literature on inventor’s knowledge search and firm’s knowledge base, focusing solely on invention originality but taking into account that the inventor is embedded in a given firm with distinct innovation practices and knowledge base (Le Masson et al., 2010a). The combination of knowledge searched for to design a given invention is then related to the components, the structure of the knowledge base, and in particular, the distance between knowledge components in the knowledge base. Thus, what would previously have been considered "local" knowledge from a firm's perspective is not necessarily local for a given inventor in that firm.

To our knowledge, we are the first to provide a formal model describing in-depth how firms are designing new inventions depending on their knowledge base structure, the combination of local and distant knowledge components and the distance between those components in firms' knowledge base. As noted by Yayavaram and Ahuja (2008): "*The problem of structuring organizational knowledge represents a significant frontier for organizational research with immense and exciting possibilities*" (Yayavaram and Ahuja, 2008, p. 358). In particular, a better understanding of firms' knowledge search practices helps answer the question of what knowledge search practices may best foster the probability of designing an invention with high technological novelty.

Our research questions are then the following (analytical framework and research questions are synthesized in Figure 1):

QR1: *How are knowledge search practices undertaken by firms to design novel inventions related to both their previous knowledge base—components and structures—and the sourcing of new external knowledge?*

QR2: *How are knowledge search practices related to the technological originality of those inventions?*

Insert Figure 1 about here

TAXONOMY OF INNOVATION SEARCH PRACTICES

Designing firm's knowledge base

In order to define a firm's knowledge base, we drew on the common assumption that to design a given invention, firms need to combine knowledge components (Katila, 2002; Nelson

& Winter, 1982). Hence, to design invention i, a given firm needs to combine n-tuples of knowledge components (referring to the sequence of all knowledge components that have to be mobilized to produce the invention). We adopted a pairwise logic, following the methodologies of Strumsky and Lobo (2015), Verhoeven et al. (2016), Yayavaram and Ahuja (2008), and Yayavaram and Chen (2015). Then, by using network theories[2] Graph G_i for invention i is comprised of vertex $V(G_i)$, representing the knowledge components that have to be mobilized to design the invention and edges $E(G_i)$.

At a given time t, the firm may have developed multiple inventions; we are then able to map the knowledge base generated by all those inventions at given time by simply unifying the different graphs of each invention. We made the assumption that, following this operation, the weight of each existing vertex and edge can be considered equal to 1[3]. Over time, a given firm produces diverse inventions by combining previously existing components from time t and newly accessed knowledge components at time $t + 1$.

Generating the taxonomy of knowledge search modes

To help in the understanding of knowledge search practices, we developed a taxonomy based on backward reasoning: what would be the effects of a newly designed invention at time $t+1$ on the previous knowledge base at time t. The taxonomy was designed based on three key dimensions: (1) the integration, or not, of new knowledge components in the firm's knowledge base; (2) the creation, or not, of new combinations of pre-existing knowledge components; and (3) the distance between the knowledge components in the firm's knowledge base. The Figure

[2] See Phelps & al. 2012 for a review.
[3] There is no escalation for knowledge accumulation, as we only focused on knowledge base structure and components.

2 provide a synthesis of the taxonomy and the Appendix A provide the formal network analysis and stylized examples.

Insert Figure 2 about here

Refinement mode.

Inventions relying on the refinement mode are entirely based on the firm's previous efforts to acquire and combine knowledge. Hence, refinement inventions do not add any new vertex or edge to the previous knowledge structure, as those inventions are only reemploying existing knowledge and combination of knowledge. This definition is in line with Strumsky and Lobo (2015). Inventions designed through this mode are related to local search and can be viewed as a proxy for knowledge depth: the firm is incrementally improving its knowledge in an already known and mastered technical discipline or area of expertise.

Clustering mode.

Inventions relying on clustering mode are based on combining knowledge components searched locally, already present in the firm's knowledge base, but not yet combined by the firm. Hence, the design leads to at least one new combination of already "close" existing knowledge components. We consider that the geodesic distance between two knowledge components constitutes an indicator of what can be considered local throughout the firm's knowledge base. For instance, combining knowledge components that have a geodesic distance strictly equal to 2 can be considered local from a firm's inventor's standpoint; the combination is then creating a knowledge cluster. We made the assumption that those elements could describe new product development projects involving, for example, two divisions of a given business unit; this would also guarantee higher modularity for future projects.

Absorption mode.

Inventions based on the absorption mode are those that draw on new knowledge components that were not part of the knowledge structure at time *t-1* and that do not imply new combinations of previously unconnected existing components in the firm's knowledge structure, except for local components (i.e., geodesic distance strictly equal to 2). Hence, those inventions can either be based on a combination of new knowledge components that are completely unconnected to the main knowledge structure, which leads to creation of a new component in the graph (case 1), or they can be based on new knowledge components that are combined with only one existing knowledge component (case 2). This mode refers to the situation in which the firm is sourcing knowledge outside of its knowledge base to complete the design of the invention; most probably, those new distant knowledge components are the results of an explorative knowledge search practice with external partners or of completely new greenfield research projects.

Recomposition mode.

Inventions based on the recomposition mode comprise new combinations of existing knowledge that are not considered local in the firm's knowledge base (i.e., that entail major changes in the structure). Potentially, this category also includes combinations involving new knowledge components. This practice is particularly challenging, as the firm needs to combine knowledge components that are already mastered within the firm but are not considered local from the inventor's point of view. This can occur because of fixation (Agogué & Le Masson, 2014) or lack of communication between silos, for example. This redefinition of the links between familiar and mastered knowledge components can be generated through access to new knowledge components.

IMPLEMENTATION OF THE TAXONOMY TO OIL & GAS INDUSTRY

Oil & Gas industry specifics

We conducted an exploratory study on the energy sector, in particular the oil & gas subsector. The oil & gas industry is key to almost all modern economies because of its major dependency on fossil fuel, significance in the global energy mix and impacts on societies (Korotayev, Bilyuga, Belalov, & Goldstone, 2018). In this study we are focusing on oil & gas industry innovation practices. In this subsector, R&D efforts are dynamically increasing due largely to three major shifts in the market: (1) the decreasing stock of oil & gas resources, requiring the development of new technologies to find and produce hydrocarbons as they become more difficult to source and produce; (2) major disasters such as Exxon Valdez (1989), Brent Spar (1995), or Deepwater Horizon (2010), which have led to increasing R&D efforts to sustain human and environmental safety; and finally (3) the diversification of major players towards more renewable energy alternatives (cf. Perrons (2014) for a detailed review of R&D trends in the sector).

Regarding our research questions, on the one hand, firms in the oil & gas industry have shown a dynamic innovation trend over the last decades, while on the other hand, some actors are old, very well-established companies we can assume have accumulated a large knowledge base over time. The subsector thus constitutes a good candidate for research on knowledge search and the design of new inventions.

Data sample

We only focused on major firms in the worldwide market, as we wanted to select firms with a large knowledge base. Firms were selected using *the Thomson Reuters Top 100 Energy Report* (Thomson-Reuters, 2017) subsection on the top 25 companies for the oil & gas subsector. This sample includes five most prominent oil- & gas-integrated companies:

companies operating in upstream (exploration and exploitation), midstream (transportation, storage, and processing) and downstream (refining, purifying and marketing and commercial distribution of various products, such as natural gas, kerosene, asphalt and other petrochemicals materials). We retrieved patent data using the *Clarivate Derwent* database, which includes patents filed in major patent offices (EP, WO, and US); we used the "Optimized Assignee" function to retrieve patents for the top 25 oil & gas companies. Clarivate Derwent conducted in-depth analyses for the 21,000 top worldwide companies in terms of patents, to retrieve filed patents, identify major subsidiaries, and correct company name spelling issues. We were able to retrieve 16 companies' patent portfolios from among the 25 selected in the sample[4]. With the help of experts from the *Technological Intelligence Unit and the Global R&D Patent Unit* of one of the major Oil & Gas companies present in the sample, we were able to obtain more fine-grained data due to corrections on subsidiaries; non-completely integrated subsidiaries were excluded from the sample. Patent data were initially retrieved for applications between January 1, 1980 and September 9, 2019.

Ultimately, our sample comprises 16 firms accounting for 125,465 unique patents and 48,662 International Patent Documentation (INPADOC) families (i.e., inventions).

Methodology

We used Patent Classification Codes to compute knowledge base components for a given firm at time *t* and applied co-classification data to build the network (i.e., the structure of the knowledge base). We relied on the International Patent Classification (IPC) codes from the

[4] Excluded companies included Bharat Petroleum, Gazprom, Hess Corporation, MOL, Lukoil, PKN Orlen, PTT Public Company Limited, Thai Oil, and S-Oil, as for these companies Clarivate Derwent did not perform in-depth analysis (i.e., did not have optimized assignee functionality); therefore, we considered that patent data for those companies were not reliable enough to be included.

World Intellectual Property Organization, which are more suitable for assessing the recombinant nature of invention than the United States Patent Office Classification (Gruber, Harhoff, & Hoisl, 2013). To design the firm's knowledge base, we relied on the IPC subclasses (i.e., IPC-4) as we considered that we only need to account for consequential knowledge components and combinations. This assumption is consistent with practices in the field (e.g., Verhoeven et al., 2016). To model the base, we looked at all the patent applications of a given firm at time *t*; due to issues related to international extensions and other patenting specificities, we do not account for patents individually but for families of patents, to analyze inventions per se (Martínez, 2011). We used all patent applications, including for patents that were not granted, under the assumption that if the firm is willing to apply for a patent, due to the costs of application procedures, the knowledge components involved in the given invention have been mastered by the company. We interpret it that if a given patent is not issued, it is due to competitors' previous applications, procedural issues, or incomplete review of the previous state of the art, none of which interferes with what knowledge components are mastered or not by the firm.

To compute the final version of the model, we had to specifically deal with both (1) time constraints, to implement the network dynamic, and (2) mono-class patents. To avoid a binary model and to take into account time-lag due to investment and research duration required to master given new knowledge, a new vertex would be considered new for a period of 3 years (Yayavaram & Ahuja, 2008; Yayavaram & Chen, 2015). On the other hand, it has been shown that knowledge comprising of filed patents is depreciated over time (Park, Shin, & Park, 2006) and that the depreciation rate is industry dependent. We set a knowledge components and combinations lifetime of 8 years following Park et al. (2006) findings for the studied industry. As for mono-class patents, some inventions only comprise a unique IPC code; studies focusing on combinatorial innovations do not take in account those patents (e.g., Verhoeven et al., 2016).

Our approach is different, as in the initial phase mono-class patents are taken into account for knowledge base computation, knowledge component newness, and lifetime duration calculation. Nevertheless, mono-class patent families are not consistent with our approach to take into account technological originality, and thus are excluded from the final studied sample after knowledge base computation.

Using the detailed data sample, we implemented our taxonomy to define knowledge search modes. We followed the steps described in Figure 3 and detailed in Appendix B. The model was developed using R code (through an R Studio Interface) developed by the authors specifically for this study. The final sample comprised 19,863 INPADOC patent families representing 50,776 patent applications filed between 1989 and 2016 for which the knowledge search taxonomy has been specified. The breakdown per company is detailed in Appendix C.

Insert Figure 3 about here

Invention technological originality measurement

The measure of technological invention originality is a complete innovation management field in itself, and it is not our objective here to provide an exhaustive assessment (see Squicciarini et al. (2013) for a review). Generally, when using patent data, scholars are accounting for patent citations in order to proxy the value of the invention, such as its commercial potential, usefulness, or social welfare benefits (Hall, Jaffe, & Trajtenberg, 2000; Harhoff, Narin, Scherer, & Vopel, 1999; Jaffe & de Rassenfosse, 2010; Trajtenberg, 1990). In this study, our focus was on patent classification classes and then the combinatorial originality of combinations; as a result, we looked for a technological originality proxy using patent classification classes. One of the main advantages of this approach is that the measure would not be affected by citations time lag.

We chose to rely on global Patent Classification Class mapping (Leydesdorff, Kushnir, & Rafols, 2014; Schoen et al., 2012). Specifically, we used an indicator derived from Alstott et al. (2017a, 2017b), who developed a methodology to compute technology network structure and draw a measure of proximity between each pairwise technological combination at IPC-4 level. In more detail, the technology network was computed by reviewing 3,911,050 utility patents issued from 1976 to 2010 by the United States Patent and Trademark Office. The measure of technological proximity aggregated multiple proximity indicators: direct citations, co-citation, cosine similarity, co-occurrence and co-classification, along with several normalization measures in order to correct impinging factors. In particular, the authors showed that on one hand, the technology network is sparse, as very few pairs of technology classes are significantly related, while on the other hand, the network is very stable over time. We called this the Alstott Score: the proximity between two IPC-4 patent classes, which was normalized between [0; 1].

In our study, we used an Alstott Derived Originality Indicator (ADOI). As the technology network is considered stable, we made the assumption that combining two IPC-4 classes that were not close in Alstott et al.'s network was an adequate indicator of technological originality meaning that the combination depart from the technological dominant design. Thus, the ADOI for the pairwise combination of IPC-4s *i* and *j* was computed as follows:

$$ADOI^{i-j} = 1 - Alstott\ Score^{i-j}$$

As an INPADOC patent family can comprise several pairwise combinations, we used an arithmetic mean to compute the ADOI score. The Alstott Score database was sourced from a file made publicly available by those authors following their publication. Our sample contained 496 single IPC-4s, of which 9 had not been computed by Alstott et al., (2017b). A total of 48 INPADOC patent families were impacted by those 9 classes and were excluded from our sample. The final sample thus comprised 61,901 IPC-4 pairwise combinations, of which 7,261 were unique IPC-4 pairwise combinations.

RESULTS

Knowledge search practices

As per our taxonomy, we estimated the proportions of the respective knowledge search mode in the sample: (1) refinement, (2) clustering, (3) absorption, and (4) recomposition. Results are presented in Figure 4. The refinement knowledge search mode was the most used (47% of the sample). This indicates that firms in this industry are mainly developing in-depth knowledge in or applying existing scientific and technological disciplines in order to incrementally develop new inventions. The second most common knowledge search mode was the absorption mode (38% of the sample). This reflects the promotion of open innovation projects across the industry. As noted, due to major changes in the field and the increase in firm innovation efforts, companies are looking more to greenfield projects; impacts on invention may be driven, for example, by new start-up acquisitions that became fully integrated in the organization, bringing non-traditional research areas for the company. Finally, clustering and recomposition modes were the least used modes. Those knowledge search modes face particular silo issues because they require new combinations of the mastered knowledge components in the firm's knowledge base, which probably indicates that the knowledge components are mastered by a different team or business unit and not by the one currently handling the invention.

Insert Figure 4 about here

We also looked at the breakdown of knowledge modes by firm and at representations of each mode in terms of patent family application portfolios. As a result, we were able to highlight huge variations in knowledge search practices across the firms in the sample, as shown in Figure

5 below. As, there was variation in the number of patents filed each year per firm, we also detailed total number of patent families filed through each knowledge search mode and total number of associated INPADOC patent family applications in Appendix D.

Insert Figure 5 about here

Technological originality of inventions

We computed the ADOI score for each patent family in the sample. There were only a few patent families with high ADOI scores: 2.1% had an ADOI score of higher than 0.9, which indicates high technological originality. Very low scores were common: 59.3% of the patent families had an ADOI score below 0.2 (Figure 6). This is consistent with the findings of Alstott et al. (2017b), which show using a large set of patents that only a few classes are connected together.

Insert Figure 6 about here

Knowledge search modes and technological originality

In this section we analyzed what knowledge search mode conduced to the highest originality score for the designed inventions. We crossed the two variables Knowledge search mode, computed according to our taxonomy, and ADOI per patent family applications. Thus, for the oil & gas industry, we obtained a clear hierarchy for which knowledge modes gain the highest scores for technological originality. The refinement mode is the least helpful in this regard, with a very low average score of 0.16; at the opposite end, recomposition is the most helpful mode for technological originality in this industry, and is also the mode that accounted for the fewest patent families (3.5% of all the patent families in the sample). Furthermore, we also showed that inventions designed through the recomposition mode without including new knowledge components also recorded high technological originality scores (Appendix E). Clustering and absorption modes were respectively the second- and third-most helpful knowledge search practices in terms of outcome technological originality. The detailed analysis results and t-tests are presented in Figure 7.

Insert Figure 7 about here

DISCUSSION & CONCLUSION

Theoretical implications

In this research, we examined firms' knowledge search practices to support the design of new inventions. We departed from the extant literature by (1) proposing a formal taxonomy of how firms combine knowledge components to design new inventions depending on their degree of reliance on their own knowledge base components and structure and (2) by exploring the relationship between the design of technologically original inventions and the selected knowledge search mode. The formal model was operationalized using patent data and network theory techniques and was applied to a set of companies in the oil & gas industry. The study findings have several important implications for research in the innovation and design field in this industry and in general, as explained next.

First, extent innovation literature shows that firms can combine knowledge components searched locally or distantly in order to design new inventions. However, in the literature, no in-depth analysis of how firms rely both on their previous knowledge base—components and structure—and on sourcing knowledge externally to design new inventions was found. We developed a taxonomy based on four knowledge search modes describing knowledge component combinations for a given invention: (1) refinement, (2) clustering, (3) absorption and (4) recomposition. The taxonomy was elaborated through backward reasoning by exploring the effects of a new invention on the firm's previous knowledge base structure. Our model sheds light on both the design of new inventions and the associated creative process in relation to the firm's knowledge base.

Second, we showed to what extent firms relied on the four knowledge search modes in a specific industry, oil & gas. Based on the analysis of 16 key players in this market, we showed that firms mainly designed new inventions by relying solely on their already mastered knowledge components and pre-existing combinations, as the refinement mode accounted for 47% of our sample. This gives new insight on Strumsky and Lobo (2015), who found that, across all firms in all industries, 52% of designed inventions reuse a set of similar knowledge components and associated structures existing in the USPTO database. Here, we showed that in addition, firms were reusing sets of similar knowledge components and associated structures existing in their own knowledge base. It has to be highlighted that innovation in the oil & gas industry has been driven by incremental innovation for a long time; it has even been considered an industry of "*medium-low R&D intensity*" (Moncada-釘ternò-Castello, Ciupagea, Smith, Tübke, & Tubbs, 2010: 527). When reviewing the technological originality of the refinement knowledge search mode, we found that it scored the lowest of the four tested modes as well as very low in absolute terms (0.16). This means that the set of combined knowledge components used when designing new inventions had been already extensively used in global technological history as sourced through USPTO-filed inventions.

Third, we reviewed the absorption knowledge search mode, which consists in designing new inventions by integrating new knowledge components without major changes in the previous knowledge structure (apart from the new knowledge components themselves). As this implies, this knowledge search mode requires the usage of distant-searched knowledge. Either of two assumptions, or one assumption encompassing two alternatives, can be made regarding the source of this knowledge: either the new knowledge component has been developed through completely new greenfield research projects within the company, or it has been sourced externally by firm acquisitions or partnerships (academia, inter-industry or intra-industry). It is highly probable that the latter approach is more extensively used by firms, as developing new greenfield projects with almost no relations with an established firm's previous expertise would be a rare event. The absorption mode was widely diffused among the 16 firms in our sample (at 37.9%). It echoes the concept introduced by Cohen and Levinthal (1990) of absorptive capacity, which assumes that firms (1) recognize and understand new external knowledge, (2) assimilate valuable external knowledge, and (3) apply assimilated external knowledge (Lane, Koka, Pathak, & Lane, 2006). Among the top 5 oil & gas firms, Exxon Mobil and BP have particular expertise in this domain, as respectively 58.9% and 47.3% of their inventions are designed through this knowledge search mode (making it their most used design mode). Furthermore, as knowledge components in this mode are mainly sourced externally, it refers to open innovation practices(Chesbrough, 2003), practices that firms in the oil & gas industry in the sample have also extensively embraced, as highlighted by Perrons (2014). Nevertheless, on average, technological originality score is relatively low for this mode (i.e., 0.279), in particular compared to the recombination mode. This means that, on average, the knowledge components sourced and combined by firms using this mode have also previously been extensively combined in the global technical landscape. Firms that extensively rely on this mode could face difficulties recognizing radically new industry knowledge, as supported by Perrons' (2014) finding that major firms in this sector prefer to be "*fast followers*" regarding new technology advancements due to high cost implications in case of failure. However, these firms could also face difficulties in the transformative learning stage when mirroring new knowledge components using more familiar firm components.

Fourth, the knowledge search mode that conduces to the highest technological originality score is the recomposition mode, which implies new inventions based on major changes regarding firm's knowledge base structure, potentially associated with new knowledge components. This mode was the least represented on average in the sample, as it only accounted for 3.5% of the designed inventions. In fact, 7 firms out of 16 in the sample filed fewer than 10

INPADOC family patent applications through that knowledge search mode. Furthermore, it was consistently the least developed knowledge search mode among the 16 firms in the sample, and SK Innovation Corp., BP, and Total Group were only three firms with more than 5% of inventions designed through it in their portfolios. The low occurrence of this mode in our sample seems to show that designing new inventions through this mode is particularly challenging for firms. Indeed, even in the top 5 companies, historical silos were strong (for example, between oil & gas exploration business units or refining & chemical business units). Mirroring knowledge components that the firm is familiar with but that have to be considered distant from the firm's inventor would be a complex task, but it has been shown in the literature that combining knowledge the firm is familiar with and new knowledge components is generally associated with high innovation performances (Ahuja & Lampert, 2001; Katila & Ahuja, 2002). Our results for inventions designed through the recomposition mode when specifically including new knowledge components, support these findings. While combining knowledge components with which the firm is familiar has in the past been mainly associated with learning traps (Ahuja & Lampert, 2001), focusing on knowledge couplings and the distance between combined knowledge components is a promising new areas of research. The findings also highlight the fact that the distinction between local and distant knowledge at firm level should also be envisioned at firm's inventor level.

Limitation and further research

The first limitation of the study concerns the generalizability of the results. The knowledge search taxonomy developed was applied only to one industry and to a restricted sample of firms. Nevertheless, we were able to perform the analysis on a large dataset of inventions (19,863). Extending our methodology through a larger sample of firms and industries will be a worthwhile avenue for future research. In particular, in-depth qualitative analysis of organizational implications of each knowledge search mode could help the understanding of knowledge search practices.

Second, we only focused here on the technological originality of the studied inventions, whereas other techniques could be applied to measure technological originality and examine whether similar results are found when using, for example, semantic analysis or citation data. In addition, other innovation performance indicators could be reviewed at firm level, in particular financial data or numbers of new launched products. We also see room for further research exploring mitigating variables, such as knowledge complexity, R&D intensity, spending, or reliance on basic science. Further research could also look at how mastering a combination of the four knowledge search modes at different levels could favor R&D efficiency.

Third, the findings contribute to the absorptive capacity and open innovation fields, as the various knowledge search modes encompass new knowledge components or combinations of knowledge components in the firm's knowledge base, some of which may have been sourced externally. We did not control for partnership effects or acquisitions. Looking at the extent to which these practices help the design of original inventions and how this relates to knowledge search modes could also be fruitful areas of research. Furthermore, heeding the call of Yayavaram and Ahuja (2008) to look closely at both the combination of knowledge and knowledge components, it would be interesting to look at whether open innovation practices favor new knowledge component sourcing or new combinations.

Fourth, our model is dependent on the usage of patents as a proxy for firm's knowledge base. Patenting activities, however, are subject to several biases, in particular related to competitive landscape and firm's intellectual property business units' practices. Other research could benefit from the usage of a different proxy for firm's knowledge base, such as academic publications, or from more fine-grained qualitative analysis. It could be promising in particular to explore the case of small firms with no or limited patent portfolios.

Fifth, the knowledge clustering mode received a high ADOI score, an intriguing research result. In particular, knowledge components combined were familiar from the firm's standpoint and the knowledge was considered local also from the firm's inventor's standpoint. The low share of the clustering knowledge mode in the firm's inventions portfolio could indicate that firms are highly fragmented and silos effects are impacting the design of new inventions at a sub-business-unit level. Further research could benefit from exploring this mode in depth to determine what causes this effect.

In summary, this study has yielded several findings that can be useful for both theory and practice. We have argued that an in-depth taxonomy of how firms are relying on their knowledge base and new knowledge acquisition to develop new inventions, focusing both on knowledge components and knowledge structures and their effects, will help in the understanding of the innovation process. We also showed that designing inventions that combine knowledge components that the firm is familiar with but that are considered distant from a firm's inventor's perspective leads to high technological originality. A greater understanding of the organizational learning practices associated with this finding may offer great potential to understand firms' innovation performance.

REFERENCES

Agogué, M., Kazakçi, A., Hatchuel, A., Le Masson, P., Weil, B., et al. 2014. The impact of type of examples on originality: Explaining fixation and stimulation effects. ***Journal of Creative Behavior***, 48(1): 1–12.

Agogué, M., & Le Masson, P. 2014. Rethinking ideation: a cognitive approach of innovation lock-ins. ***The International Society for Professional Innovation Management***, 1–11.

Ahuja, G., & Lampert, C. M. 2001. Entrepreneurship in the large corporation: A longitudinal study of how established firms create breakthrough inventions. ***Strategic Management Journal***, 22(6–7): 521–543.

Alstott, J., Triulzi, G., Yan, B., & Luo, J. 2017a. Inventors' Explorations Across Technology Domains. ***Cambridge University Press***.

Alstott, J., Triulzi, G., Yan, B., & Luo, J. 2017b. Mapping technology space by normalizing patent networks. ***Scientometrics***, 110(1): 443–479.

Andriopoulos, C., & Lewis, M. W. 2010. Managing Innovation Paradoxes: Ambidexterity Lessons from Leading Product Design Companies. ***Long Range Planning***, 43(1): 104–122.

Arthur, W. B. 2007. The structure of invention. ***Research Policy***, 36(2): 274–287.

Arts, S., & Fleming, L. 2016. Paradise of Novelty—Or Loss of Human Capital? Exploring New Fields and Inventive Output. ***Organization Science***, 29(6): 1074–1092.

Chesbrough, H. 2003. ***Open Innovation, The New imperative for Creating and Profiting from Technology***. Harvard Business Review Press.

Choi, H., Shin, J., & Hwang, W. S. 2018. Two faces of scientific knowledge in the external technology search process. ***Technological Forecasting and Social Change***, 133(March): 41–50.

Cohen, W. M., & Levinthal, D. A. 1989. Innovation & Learning: The Two Faces of R&D. ***The Economic Journal***, 99 (397): 569–596.

Cohen, W. M., & Levinthal, D. A. 1990. Absorptive Capacity: A New Perspective on Learning and Innovation. ***Administrative Science Quarterly***.

Cooper, A. C., & Schendel, D. 1976. Strategic responses to technological threats. ***Business Horizons***, 19(1): 61–69.

Dahlin, K. B., & Behrens, D. M. 2005. When is an invention really radical?: Defining and measuring technological radicalness. ***Research Policy***, 34(5): 717–737.

Dosi, G. 1982. Technological paradigms and technological trajectories. ***Research Policy***, 11(1): 147–162.

Fleming, L. 2001. Recombinant Uncertainty in Technological Search. ***Management Science***, 47(1): 117–132.

Fleming, L., & Sorenson, O. 2001. Technology as a complex adaptive system: evidence from patent data. ***Research Policy***, 30: 1019–1039.

Fleming, L., & Sorenson, O. 2004. Science as a map in technological search. ***Strategic Management Journal***, 25(89): 909–928.

Grant, R. M. 1996. Prospering in Dynamically-competitive Environments: Organizational Capability as Knowledge Integration. ***Organization Science***, 7(4): 375–387.

Gruber, M., Harhoff, D., & Hoisl, K. 2013. Knowledge recombination across technological boundaries: Scientists vs. Engineers. ***Management Science***, 59(4): 837–851.

Hall, B., Jaffe, A., & Trajtenberg, M. 2000. ***Market Value and Patent Citations: A First Look***.

Hargadon, A., & Sutton, R. I. 1997. Technology Brokering and Innovation in a Product Development Firm. ***Administrative Science Quaterly***, 42(4): 716–749.

Harhoff, D., Narin, F., Scherer, F., & Vopel, K. 1999. Citation frequency and the value of patented innovation. ***The Review of Economics and Statistics***, 81(3): 511–515.

Hatchuel, A., Reich, Y., Le Masson, P., Weil, B., & Kazakçi, A. O. 2013. Beyond Models and Decisions : Situating Design Through Generative Functions. ***ICED13: 19th International Conference on Engineering Design***, (August): 1–10.

Hatchuel, A., & Weil, B. 2003. A new approach of innovative design: an introduction to C-K theory. ***International Conference on Engineering Design ICED 03***, 1–15.

Hatchuel, A., & Weil, B. 2009. CK Design Theory: An Advanced Formulation. ***Research in Engineering Design***, 19(4): 181–192.

Henderson, R., & Clark, K. 1990. Architectural Innovation : The Reconfiguration of Existing Product Technologies and the Failure of Established Firms. ***Administrative Science Quarterly***, 35(1): 9–30.

Jaffe, A. 1986. Technological Opportunity and Spillovers of R&D: Evidence from Firm's Patents, Profits, and Market Value. ***The American Economic Revie***, 76(5): 984–1001.

Jaffe, A., & de Rassenfosse, G. 2010. Patent Citation Data in Social Science Research: Overview and Best Practices. ***Journal of the American Society for Information Science***, 1(6): 2581–2583.

Kaplan, S., & Tripsas, M. 2008. Thinking about technology: Applying a cognitive lens to technical change. ***Research Policy***, 37(5): 790–805.

Katila, R. 2002. New Product Search over Time: Past Ideas in Their Prime? ***The Academy of Management Journal***, 45(5): 995–1010.

Katila, R., & Ahuja, G. 2002. Something Old, Something New: A Longitudinal Study of Search Behavior and New Product Introduction. ***Academy of Management Journal***, 45(6): 1183–1194.

Kline, S. J., & Rosenberg, N. 1986. An overview of innovation. ***European Journal Innovation Management*** 38: 275–305.

Kogut, B., & Zander, U. 1992. Knowledge of the Firm , Combinative Capabilities , and the Replication of Technology. ***Organization Science***, 3(3): 383–397.

Korotayev, A., Bilyuga, S., Belalov, I., & Goldstone, J. 2018. Oil prices, socio-political destabilization risks, and future energy technologies. ***Technological Forecasting and Social Change***, 128(May 2017): 304–310.

Lane, P., Koka, B. R., Pathak, S., & Lane, P. I. 2006. The Reification of Absorptive Capacity: A Critical Review and Rejuvenation of the Construct. ***The Academy of Management Review***, 31(4): 833–863.

Lawson, B., & Samson, D. 2001. Developing Innovation Capability in Organisations: a Dynamic Capabilities Approach. ***International Journal of Innovation Management***, 05(03): 377–400.

Le Masson, P., Weil, B., & Hatchuel, A. 2010a. The design activity and innovation capability. ***Strategic Management of Innovation and Design***: 51–66. Cambridge University Press.

Le Masson, P., Weil, B., & Hatchuel, A. 2010b. ***Strategic Management of Innovation and Design***. Cambridge University Press.

Levinthal, D. A., & March, J. G. 1993. The Myopia of Learning. ***Strategic Management Journal***, 14: 95–112.

Leydesdorff, L., Kushnir, D., & Rafols, I. 2014. Interactive overlay maps for US patent (USPTO) data based on International Patent Classification (IPC). ***Scientometrics***, 98(3): 1583–1599.

Lobo, J., & Strumsky, D. 2019. Sources of inventive novelty: two patent classification schemas, same story. ***Scientometrics***, 120(1): 19–37.

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

Martínez, C. 2011. Patent families: When do different definitions really matter? ***Scientometrics***. https://doi.org/10.1007/s11192-010-0251-3.

Moncada-Paternò-Castello, P., Ciupagea, C., Smith, K., Tübke, A., & Tubbs, M. 2010. Does Europe perform too little corporate R&D? A comparison of EU and non-EU corporate R&D performance. ***Research Policy***, 39(4): 523–536.

Nakamura, H., Suzuki, S., Sakata, I., & Kajikawa, Y. 2015. Knowledge combination modeling: The measurement of knowledge similarity between different technological domains. ***Technological Forecasting and Social Change***, 94: 187–201.

Nelson, R., & Winter, G. S. 1982. ***An Evolutionary Theory of Economic Change***. Harvard University Press (MA).

Nerkar, A. 2003. Old is gold? The value of temporal exploration in the creation of new knowledge. ***Management Science***, 49(2): 211–229.

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

Nooteboom, B., Van Haverbeke, W., Duysters, G., Gilsing, V., & van den Oord, A. 2007. Optimal cognitive distance and absorptive capacity. ***Research Policy***, 36(7): 1016–1034.

O'Connor, G. C. 2008. Major innovation as a dynamic capability: A systems approach. ***Journal of Product Innovation Management***, 25(4): 313–330.

O'Reilly, C. A., & Tushman, M. L. 2008. Ambidexterity as a dynamic capability: Resolving the innovator's dilemma. ***Research in Organizational Behavior***, 28: 185–206.

Park, G., Shin, J., & Park, Y. 2006. Measurement of depreciation rate of technological knowledge: Technology cycle time approach. ***Journal of Scientific and Industrial Research***, 65(2): 121–127.

Perrons, R. K. 2014. How innovation and R&D happen in the upstream oil & gas industry: Insights from a global survey. ***Journal of Petroleum Science and Engineering***, 124: 301–312.

Raisch, S., Birkinshaw, J., Probst, G., & Tushman, M. L. 2009. Organizational ambidexterity: Balancing exploitation and exploration for sustained performance. ***Organization Science***, 20(4): 685–695.

Sarica, S., Yan, B., Bulato, G., Jaipurkar, P., & Luo, J. 2019. Data-Driven Network Visualization for Innovation and Competitive Intelligence. ***Proceedings of the 52nd Hawaii International Conference on System Sciences***, 127–135.

Schoen, A., Villard, L., Laurens, P., Cointet, J.-P., Heimeriks, G., et al. 2012. The Network Structure of Technological Developments ; Technological Distance as a Walk on the Technology Map. ***STI Conference***, 734–742.

Schumpeter, J. A. 1934. ***Theory of Economic Development*** (New Edition). Routledge, London.

Squicciarini, M., Denis, H., & Criscuolo, C. 2013. Supporting investment in knowledge capital, growth and innovation. ***OECD***.

Strumsky, D., & Lobo, J. 2015. Identifying the sources of technological novelty in the process of invention. ***Research Policy***, 44(8): 1445–1461.

Sydow, J., Schreyögg, G., & Koch, J. 2009. Organizational Path Dependence: opening the black box. ***Academy of Management Review***, 34(4): 689–709.

Teece, D. J. 1996. Firm organization, industrial structure, and technological innovation. ***Journal of Economic Behavior and Organization***, 31: 193–225.

Thomson-Reuters. 2017. ***Top-100 Global Energy Leaders***. http://top100innovators.com/.

Trajtenberg, M. 1990. A Penny for Your Quotes: Patent Citations and the Value of Innovations. ***The RAND Journal of Economics***, 21(1): 172.

Verhoeven, D., Bakker, J., & Veugelers, R. 2016. Measuring technological novelty with patent-based indicators. ***Research Policy***, 45(3): 707–723.

Veugelers, R., & Wang, J. 2019. Scientific novelty and technological impact. ***Research Policy***, 48(6): 1362–1372.

West, J., & Bogers, M. 2014. Leveraging external sources of innovation: A review of research on open innovation. ***Journal of Product Innovation Management***, 31(4): 814–831.

Yayavaram, S., & Ahuja, G. 2008. Decomposability in Knowledge Structures and Its Impact on the Usefulness of Inventions and Knowledge-base Malleability. ***Administrative Science Quarterly***, 53(2): 333–362.

Yayavaram, S., & Chen, W.-R. 2015. Changes in firm knowledge couplings and firm innovation performance: the moderating role of technological complexity. ***Strategic Management Journal***, 36: 377–396.

FIGURE 1

Analytical framework and research questions

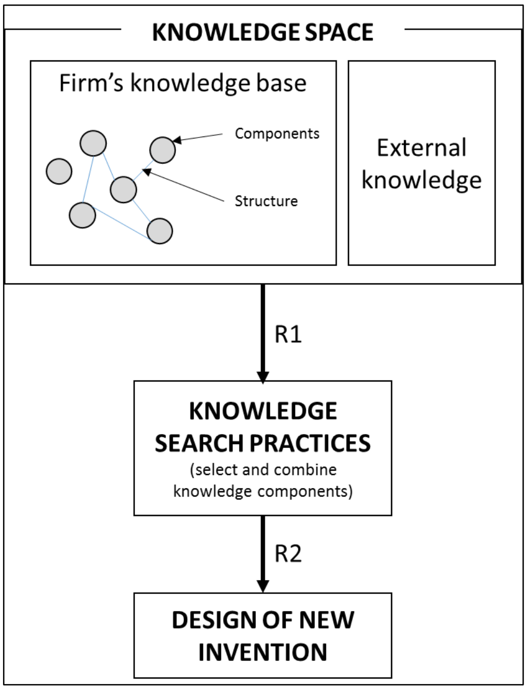


FIGURE 2

Rationales of the knowledge search taxonomy

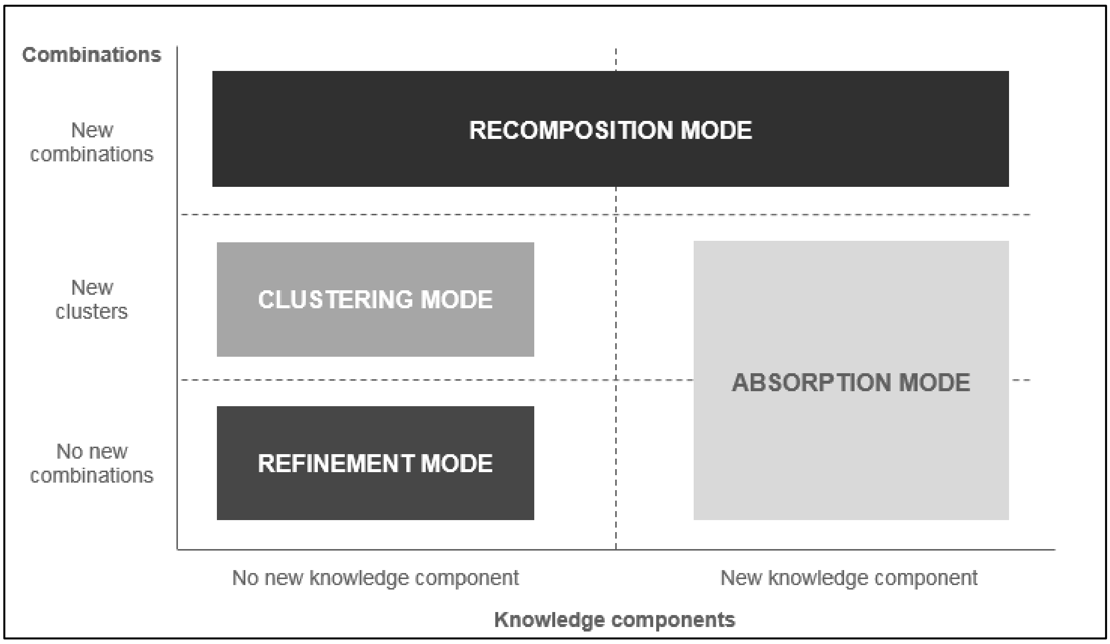


FIGURE 3

Stages of model implementation

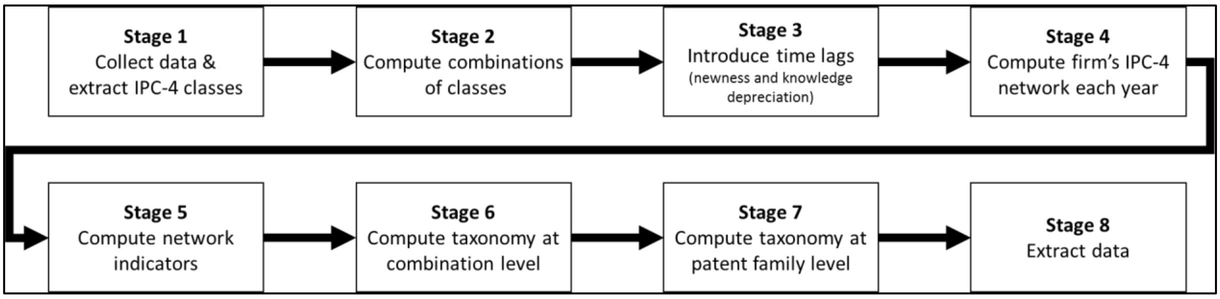


FIGURE 4

Knowledge search modes in the sample

	No. patent families	Repartition
Refinement mode	9,336	47.0%
Absorption mode	7,534	37.9%
Clustering mode	2,304	11.6%
Recomposition mode	689	3.5%
Total	**19,863**	**100%**

Table 1 - Knowledge search modes in the sample

FIGURE 5

Firm's INPADOC patent family applications per Knowledge Search Mode

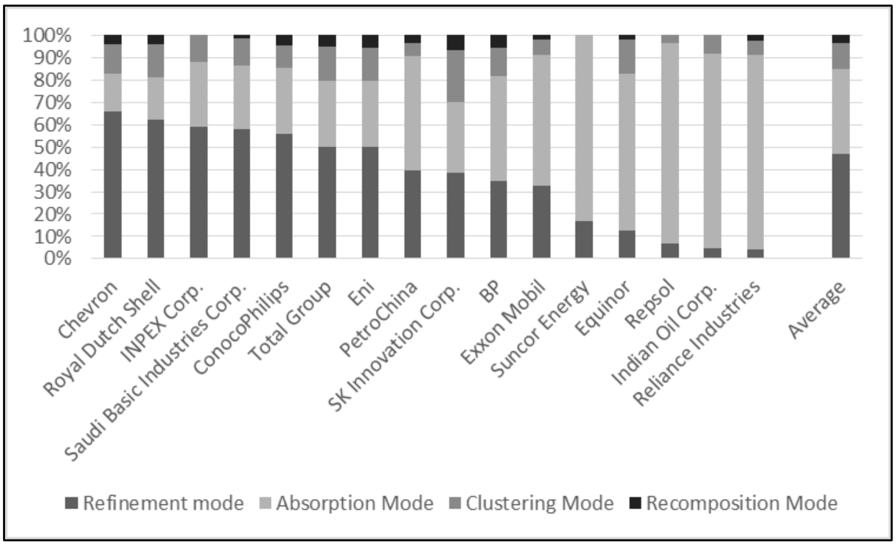


FIGURE 6

ADOI density in the sample

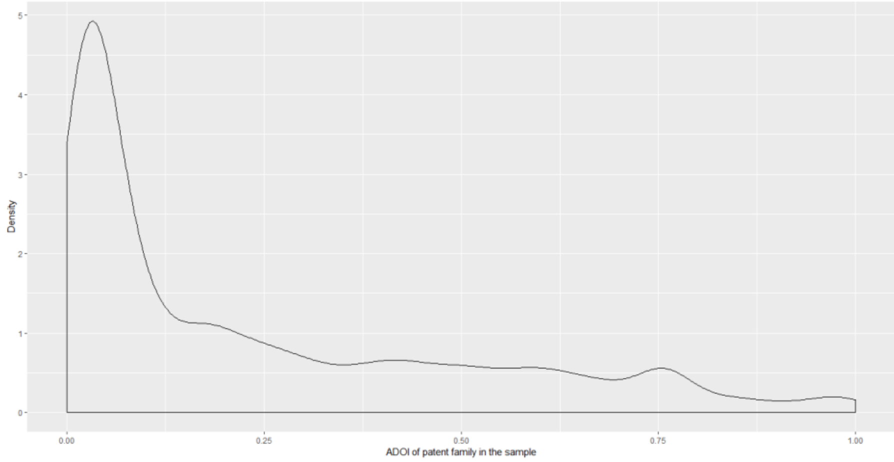


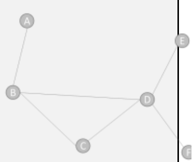
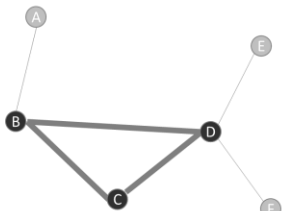
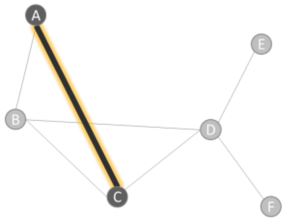
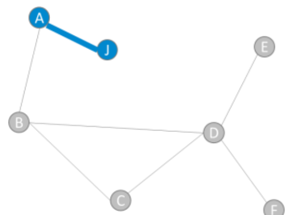
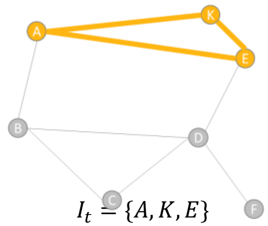
FIGURE 7

Analysis ADOI score and knowledge search mode

	ADOI average score per filed patent families	95% confidence interval		p-value
		min.	max.	
Refinement mode	0.160	0.156	0.165	$p < 0.001$
Absorption mode	0.279	0.273	0.285	$p < 0.001$
Clustering mode	0.387	0.376	0.397	$p < 0.001$
Recomposition mode	0.466	0.448	0.484	$p < 0.001$

APPENDIX A

Formal model of Knowledge Search Modes & Stylized Examples

	Initialization	Refinement mode	Custering mode	Absorption mode	Recomposition mode
Invention	*NA*	$I_{t+1} = \{1, \dots, n\}$	$I_{t+1} = \{1, \dots, n\}$	$I_{t+1} = \{[1, \dots, m], [n, \dots, o]\}$	$I_{t+1} = \{1, \dots, n\}$
Period	*t*	*t+1*	*t+1*	*t+1*	*t+1*
Vertex properties	NA	$V(G_{t+1}) - V(G_t) = 0$	$V(G_{t+1}) - V(G_t) = 0$	$V(G_{t+1}) - V(G_t) \geq 1$	$V(G_{t+1}) - V(G_t) \geq 0$
Edge properties	NA	$E(G_{t+1}) - E(G_t) = 0$	$E(G_{t+1}) - E(G_t) \geq 1$	$E(G_{t+1}) - E(G_t) \geq 1$	$(G_{t+1}) - E(G_t) \geq 1$
Geodesic distance properties	NA	$\max(d_t^{i-j}) = 1, \forall\, i,j \in [1,n]$	$d_t^{i-j} = 2, \exists\, i,j \in [1,n]$ $\max(d_t^{k-l}) \leq 2, \forall\, k,l \in [1,n]$	Case 1: $V(I_{t+1}) \cap V(G_t) = \emptyset$ Case 2: $d_t^{i-j} \begin{cases} = \emptyset, \exists\, i,j \in [1, \dots m] \\ \leq 2, \forall\, i,j \in [n, \dots, o] \end{cases}$	$d_t^{i-j} \in [3; +\infty[$ $\exists\, i,j \in [1,n]$
Stylized example		 $I_t = \{B, C, D\}$	 $I_t = \{A, C\}$	 $I_t = \{A, J\}$	 $I_t = \{A, K, E\}$

APPENDIX B
Methodology Details

Stage 1: Collect data & extract IPC-4 classes
IPC-4 classes that were part of a given patent were aggregated at INPADOC patent family level; we computed IPC-4 data for the 48,662 INPADOC patent families. The date used for the INPADOC patent family is the earliest date for any of the patents in the family. Each INPADOC patent family including application year and IPC-4 n-tuples, was assigned to a company portfolio.

Stage 2: Compute combination of classes
Based on the list of IPC-4 n-tuples of each INPADOC patent family, a list of IPC-4 pairwise combinations was built. A total of 83,970 combinations and INPADOC patent family pairs were retrieved from the sample.

Stage 3: Introduce time lags
Time lags are introduced to take into account newness of knowledge components and knowledge stock depreciation. Each new IPC-4 or IPC-4 pairwise combination in a firm's patent portfolio is considered new to the firm for a duration of 3 years. We set a knowledge components and combinations lifetime of 8 years; if a given IPC-4 or pairwise combination of IPC-4s was not filed again through another INPADOC patent family in that time, it was deleted from the firm's knowledge base after 8 years. Following the application of time lags, 204,239 combinations of IPC codes were computed. It has to be noted that for a given company, if an IPC-4 is deleted after 8 years due to no further applications but is then filed again after 8 years, it is treated as a new one.

Stage 4: Compute firm's IPC-4 network each year
Based on IPC-4s extracted from mono-class INPADOC patent families and IPC-4 pairwise combinations extracted from multi-class INPADOC patent families, with the application of time lags, IPC-4 networks were computed to design firms' knowledge base. A network graph was generated each year from 1980 to 2019 for each firm in the sample. Because of variations regarding the first year of patenting due on different dates of firm establishment, 558 graphs were generated.

Stage 5: Compute network indicators
For each network graph, the geodesic distance between any pair of IPC-4s is computed, generating 4,777,396 observations. Furthermore, each vertex of each network graph was classified as belonging to the principal component or to a non-principal component.

Stage 6: Compute taxonomy at IPC-4 pairwise combination level
For each pairwise IPC-4 combination, a first level of analysis is computed to generate the complete taxonomy at INPADOC patent family level at the next stage. Hence, we distinguished between:

- IPC-4 pairwise combinations with only non-new IPC-4s that had been filed for more than 3 years—*refinement pairwise combinations*.
- IPC-4 pairwise combinations with only non-new IPC-4s and with a previous geodesic distance strictly equal to 2—*knowledge cluster enhancement pairwise combinations*.
- New combinations of IPC-4s not related to the principal component—*strict external absorption pairwise combinations*.
- New IPC-4 pairwise combinations with a geodesic distance strictly superior to 2, with interactions only with principal components and including both a new and a non-new IPC-4—*non-strict external absorption*.
- New IPC-4 pairwise combinations with a geodesic distance strictly superior to 2, with interactions only with principal components and including two non-new IPC-4s—*recomposition pairwise combination.*

Stage 7: Compute taxonomy at patent family level
As INPADOC patent families may comprise multiple combinations depending on the size of the IPC-4 n-tuples, we defined an algorithm to apply the complete taxonomy. The steps are applied successively to the stock of IPC-4 pairwise combinations per INPADOC patent family.

- Step 1: At least one strict external *absorption pairwise combination*—**absorption patent family**;
- Step 2: At least one *recomposition pairwise combination* (entailing no absorption pairwise combination)—**recomposition patent family**;
- Step 3: At least one *non-strict external absorption pairwise combination* (entailing no *strict external absorption pairwise combination* or *recomposition pairwise combination*)—**absorption patent family**;
- Step 4: At least one *knowledge cluster enhancement pairwise combination* (entailing no step 1 to 3 combinations)—**knowledge cluster enhancement patent family**;
- Step 5: other cases leading to at least one *refinement pairwise combination*—**refinement patent family**.

Stage 8: Data extraction
As we were focusing on knowledge bases that were dynamically changing over time, we needed to account for time to allow network stabilization. We thus used the sample of patent families only after one complete cycle of knowledge depreciation (i.e., 8 years); furthermore, INPADOC patent families filed between 2017 and 2019 were deleted due to the delay between the moment of patent publication and submission to ensure data consistency. After these restrictions, the sample contains 32,594 INPADOC patent families. Finally, in line with previous studies, we excluded INPADOC patent families with only a single class, as we were focusing on combinations. The final sample contained 19,825 INPADOC patent families

APPENDIX C

Repartition of the knowledge search modes in the sample (thousands of INPADOC patent families)

	Average no. patent families filed per year	*Standard deviation*	Average no. IPCs in the knowledge base	*Standard deviation*	Average no. IPCs pairwise combinations in the knowledge base	*Standard deviation*	Total number of patent families in the sample
Exxon Mobil	199.9	*47.1*	235.3	*26.2*	1,174.4	*191.5*	5,596
Royal Dutch Shell	139.0	*31.5*	228.2	*30.2*	982.6	*261.1*	3,891
Chevron	74.5	*32.0*	166.4	*24.5*	580.7	*132.8*	2,087
BP	58.9	*32.3*	185.1	*51.2*	598.0	*224.2*	1,648
Saudi Basic Industries Corp.	65.5	*78.2*	75.0	*74.0*	348.0	*417.4*	1,638
Total Group	57.5	*28.9*	129.2	*43.3*	435.2	*244.8*	1,609
ConocoPhillips	40.6	*24.7*	139.1	*38.1*	387.7	*132.3*	1,138
Eni	36.3	*14.3*	102.7	*34.3*	323.1	*146.2*	1,017
Equinor	14.1	*6.7*	53.7	*30.2*	95.9	*69.7*	380
SK Innovation Corp.	41.0	*13.3*	69.4	*42.5*	255.7	*161.5*	246
Reliance Industries	9.5	*11.7*	28.2	*34.7*	69.0	*101.4*	181
PetroChina	9.0	*6.2*	27.9	*19.3*	35.9	*33.3*	117
Repsol	4.9	*2.9*	33.1	*14.5*	49.9	*25.7*	88
Indian Oil Corp.	5.7	*5.5*	21.3	*16.1*	32.9	*31.9*	86
INPEX Corp.	6.3	*6.6*	12.2	*9.6*	30.6	*28.6*	76
Suncor Energy	3.1	*2.0*	22.0	*8.0*	42.7	*34.6*	65
Average / Total	**47.9**	**54.4**	**95.5**	**76.2**	**340.1**	**350.3**	**19,863.0**

APPENDIX D

Repartition of the knowledge search modes in the sample (thousands of INPADOC patent families)

	Refinement mode		Clustering mode		Absorption mode		Recomposition mode		Total no. patent families
	No. patent families	*% of firm's portfolio*	No. patent families	*% of firm's portfolio*	No. patent families	*% of firm's portfolio*	No. patent families	*% of firm's portfolio*	
Exxon Mobil	1,818	*32.5%*	387	*6.9%*	3,295	*58.9%*	96	*1.7%*	5,596
Royal Dutch Shell	2,422	*62.2%*	587	*15.1%*	726	*18.7%*	156	*4.0%*	3,891
Chevron	1,371	*65.7%*	270	*12.9%*	359	*17.2%*	87	*4.2%*	2,087
BP	570	*34.6%*	206	*12.5%*	779	*47.3%*	93	*5.6%*	1,648
Saudi Basic Industries Corp.	945	*57.7%*	196	*12.0%*	470	*28.7%*	27	*1.6%*	1,638
Total Group	804	*50.0%*	245	*15.2%*	476	*29.6%*	84	*5.2%*	1,609
ConocoPhillips	636	*55.9%*	114	*10.0%*	335	*29.4%*	53	*4.7%*	1,138
Eni	508	*50.0%*	148	*14.6%*	301	*29.6%*	60	*5.9%*	1,017
Equinor	48	*12.6%*	57	*15.0%*	267	*70.3%*	8	*2.1%*	380
SK Innovation Corp.	94	*38.2%*	57	*23.2%*	79	*32.1%*	16	*6.5%*	246
Reliance Industries	8	*4.4%*	11	*6.1%*	157	*86.7%*	5	*2.8%*	181
PetroChina	46	*39.3%*	7	*6.0%*	60	*51.3%*	4	*3.4%*	117
Repsol	6	*6.8%*	3	*3.4%*	79	*89.8%*	-	*0.0%*	88
Indian Oil Corp.	4	*4.7%*	7	*8.1%*	75	*87.2%*	-	*0.0%*	86
INPEX Corp.	45	*59.2%*	9	*11.8%*	22	*28.9%*	-	*0.0%*	76
Suncor Energy	11	*16.9%*	-	*0.0%*	54	*83.1%*	-	*0.0%*	65
Total	**9,336**	***47.0%***	**2,304**	***11.6%***	**7,534**	***37.9%***	**689**	***3.5%***	**19,863**

APPENDIX E

Analysis ADOI score and Recomposition mode—details

	ADOI of recomposition mode inventions	95% confidence interval		Occurrence	Share of recombination mode	p-value
		min.	max.			
Recomposition mode - without new knowledge components	0.448	0.427	0.469	528	76.6%	p < 0.001
Recomposition mode - with new knowledge components	0.525	0.494	0.555	161	23.4%	p < 0.001
Recomposition Mode - average	**0.466**	**0.448**	**0.484**	**689**	**100.0%**	**p < 0.001**