1-2024

Innovation in dynamic knowledge landscapes: using topic modelling to map inventive activity and its implications for financial performance

Simon J.D. SCHILLEBEECKX
Singapore Management University, simon@smu.edu.sg

Yimin LIN
Singapore Management University, ymlin@smu.edu.sg

Gerard GEORGE
Georgetown University

INNOVATION: ORGANIZATION & MANAGEMENT
https://doi.org/10.1080/14479338.2022.2062365

Routledge
Taylor & Francis Group

ARTICLE

OPEN ACCESS Check for updates

Innovation in dynamic knowledge landscapes: using topic modelling to map inventive activity and its implications for financial performance

Simon J. D. Schillebeeckx[a], Yimin Lin[a] and Gerard George[b]

[a]Lee Kong Chian School of Business, Singapore Management University, Singapore; [b]McDonough School of Business, Georgetown University, United States

ABSTRACT

We examine the relationship between innovation and performance in the agricultural industry by studying how a firm's patent portfolio position in the knowledge landscape moderates the relationship between three firm search dimensions (scope, specialisation, and commitment) and the firm's financial performance. To represent dynamism in the knowledge landscape, we apply topic modelling to 67,120 patent texts of 571 firms and introduce two complementary perspectives on how knowledge can be dynamically categorised. This allows us to create two representations of the knowledge structure to identify contested (scarce) clusters of high (low) recent inventive activity and emergent clusters where knowledge ambiguity has reduced. Our findings suggest that search scope and commitment prove more valuable near scarce clusters, whereas search specialisation is most valuable in contested clusters. In addition, we find that all three search dimensions are positively moderated by proximity to emergent clusters, but vary in their effectiveness. Our results explain an additional 1% in within-firm financial performance (EBITDA). We discuss implications for innovation and strategy research as well as for practice.

ARTICLE HISTORY
Received 28 Febuary 2022
Accepted 31 March 2022

KEYWORDS
Innovation; search; patents; agriculture; topic modelling; Latent Dirichlet Allocation

The importance of discovery and innovation cannot be overstated in the knowledge economy. Innovation is key to corporate longevity and financial performance (Hall & Harhoff, 2012) and knowledge creation influences both the scale and the scope of organisations (Nickerson & Zenger, 2004). Indeed, the knowledge-based view posits that firms exist because they are superior at knowledge creation than markets (Grant, 1996; Kogut & Zander, 1992) and invest in a variety of strategies to create value from innovation that exploits the firm's and industry's knowledge base (Alexy et al., 2013; Cohen et al., 2000; Leppänen et al., 2021). An effective and commonly used mechanism to capture value and protect innovative activity is through patents, which confer legal rights to exclude others from delineated innovation areas (Hall & Harhoff, 2012). While the exact value of patents is difficult to determine, researchers have argued that the market value of patents correlates with forward citations as well as patent renewals (Bessen, 2009;

CONTACT Gerard George gerard.george@georgetown.edu

Griliches, 1998; Hall et al., 2005; Hegde & Sampat, 2009). Research has also established relationships between a firm's patent portfolio, its technological diversity, and financial performance (Deng et al., 1999; Lin et al., 2006; Miller, 2006).

We draw from the literature on technological landscapes (Fleming, 2001; Fleming & Sorenson, 2001; Knudsen & Srikanth, 2014; March, 1991) to establish contingent relationships between a firm's search dimensions, the firm's portfolio position in the knowledge landscape, and firm performance. Like Fleming and Sorenson (2001, p. 1019), 'we see invention as a process of recombinant search over technology landscapes'. Because firms have imperfect information and diverging perspectives about the searchable environment (Knudsen & Srikanth, 2014), the effectiveness of search is context-dependent (Purcell & McGrath, 2013). Therefore, we investigate how firm performance is influenced by the contingencies between three search dimensions (*scope, specialisation*, and *commitment*) and the firm's innovation portfolio position relative to specifically identified clusters in the technology landscape (*sparse, contested*, or *emergent clusters*).

The emergence of data science techniques provides opportunities to reframe established theories with granular measures that better encapsulate a phenomenon (George, Osinga et al., 2016). We delve deeper into patent text data to garner insight at a semantic level on the association between innovations rather than commonly used static classification or backward citation data. We apply Latent Dirichlet Allocation (LDA) to the 1995–2011 universe of agricultural patent abstracts[1] to identify the technological landscape (c.f. Hackett et al., 2021). This natural language processing technique relies on a generative statistical model that seeks to uncover unobserved groups in sets of observations and establishes similarity and difference in group membership between patent texts (Jung & Lee, 2016). We combine LDA, a form of topic modelling, with network analysis to create multi-dimensional positioning of vectorised texts. This permits the identification of various clusters within the innovation landscape, based on the semantic distance between inventions.

Our unit of analysis is the firm's recent invention portfolio, which by default occupies space in the evolving knowledge landscape. While looking at the patent portfolio as the unit of analysis is not that common in knowledge recombination research, a recent review highlighted the need to study the portfolio of knowledge components in reference to the context, i.e., the landscape, in which those components are used (Schillebeeckx et al., 2021; Xiao et al., 2021). Moreover, the distance between a firm's invention portfolio and that of its partners remains a topic of continued interest (Nan et al., 2018; Zhang et al., 2019). We therefore contend that a firm's invention portfolio's distance to industry knowledge is also worthy of closer attention. We do so by focusing on sparse, contested, and emergent clusters in the knowledge space.

A *sparse* cluster is an area within the wider knowledge landscape in which innovations are noticeably distinct from one another. Distance is associated with a low concentration of inventive activity and significant differences between the nearest neighbouring inventions. A *contested* cluster is an area in which innovations are similar, thereby making inventive processes contested activities in which many actors seek to uncover valuable peaks in proximity to one another. By re-evaluating and reconstructing the knowledge landscape every calendar quarter using topic modelling, our focus remains squarely on the relationship between the firm's portfolio and to what topics the wider industry is recently paying attention. This method changes static representations, which rely on a specific instance of a patent classification scheme, to *progressively dynamic* representations of the landscape that captures industry activity and innovation focus.

A complementary perspective on the knowledge landscape focuses not on focal industry attention, but on areas in which the knowledge structure has shifted. Specifically, we identify areas that are characterised by increasing clarity regarding the semantic connections between different patents. In such areas, ideas that previously seemed distinct, or even disjointed, become more coherent and cohesive following new discoveries. In other words, new knowledge has exposed previously unknown connections, making the relationships between prior innovations more transparent, i.e., *retrospectively dynamic*, and the topography of the technological landscape more predictable. We refer to areas characterised by significant shifts as *emergent* clusters.

Against this backdrop, we focus on three search dimensions: scope, specialisation, and commitment. *Search scope* (i.e., knowledge breadth) allows firms to draw connections between disjointed technological domains and helps counter myopia and knowledge insularity (Knudsen & Srikanth, 2014; Paruchuri & Awate, 2017). *Search specialisation* (i.e., focused knowledge or technological depth) allows firms to benefit from knowledge spillovers in technological landscapes (Audretsch & Feldman, 1996; Iammarino & McCann, 2006). Specialised firms are less at risk of leaking valuable knowledge and more easily absorb spillovers from other firms (Alnuaimi & George, 2016). Firms could also adopt scale-driven strategies by investing significantly in innovation, possibly without a specific scope or specialisation. Finally, *search commitment* reflects the firm's financial investment in developing its absorptive capacity (Cohen & Levinthal, 1990). Firms require significant tacit knowledge and committed resources to turn acquired knowledge into unique value through transformation (Zahra & George, 2002; Zou et al., 2018). Similar to Nag and Gioia's (2012) micro-processes where managers turn knowledge commonly held by rivals into uncommon knowledge by using idiosyncratic knowledge practices and unique scanning orientations, we envisage a meso-process that enables firms with high search commitment to create value in dynamic landscapes.

Our empirical context is the agricultural technological landscape, which is of vital importance to economic growth, well-being, climate change, and other grand challenges, but has received scarce attention in management (George, Howard-Grenville et al., 2016). As agricultural incumbents like Bayer and Monsanto make headlines by stating their planned merger will increase innovation (Gullickson, 2017), directing this innovation towards regenerative, sustainable resource management is essential in the fight against climate change and to halt the ongoing biodiversity collapse (George et al., 2021; George & Schillebeeckx, 2022; Pigford et al., 2018).

We contribute to the literature in three ways. First, we establish a positive, contingent relationship between search dimensions, portfolio position, and financial performance. By looking at the inventive portfolio position in dynamic technological landscapes, we enrich debates on ambidexterity (synchronous exploration and exploitation) versus punctuated equilibrium (sequential cycles of exploration and exploitation). Our findings give credence to the contingency perspective as the position within a technological landscape proves to be an important contingency for research to have beneficial effects (Clauss et al., 2021; Gupta et al., 2006).

Second, we extend insights into the role of knowledge landscapes by recognising that dynamic landscapes are redrawn iteratively by incorporating new knowledge. Specifically, our conceptualisation of 'emergent clusters' creates new insights into the broader role of environmental dynamism and firm performance. While patenting agencies regularly update their classification schemes to capture technological evolution [USPTO interviews, 2016], most studies have considered classification as a static artefact that allows researchers to proxy technological component recombination. We introduce real dynamism by recognising that new innovations can change the meaning and relevance of, as well as relationships or couplings between, prior inventions and can alter the semantic components between and within innovations, thereby transforming the technological landscape (Fleming & Sorenson, 2001). Finally, the use of topic modelling to capture dynamic properties of the technological landscape opens up promising areas for future research in which the same methodology can be used to investigate related as well as very different questions (Jung & Lee, 2016). Next, we develop hypotheses for the firm's search strategies in different technological landscapes.

Innovation in dynamic knowledge landscapes

Prior research has established a contingent relationship between search, knowledge, the innovation landscape, and innovation success (Fleming & Sorenson, 2004; Kneeland et al., 2020). We build on this work to study how key dimensions of a firm's search (scope, commitment, and specialisation) affect the firm's financial performance and how this relationship is affected by the firm's patent portfolio position in the knowledge landscape. Because knowledge landscapes are essentially dynamic, we introduce two complementary representations of the knowledge landscape to expose two types of dynamism: changes in industry attention and changes in industry understanding. We capture these respectively as topical clusters in which we see a strong (or weak) inventive density, which is suggestive of high (low) recent industry activity and as topical clusters in which recent industry innovations have been positioning in the vicinity of previously distant innovations, thereby filling gaps in the knowledge structure, and reducing ambiguity of the topography of the knowledge landscape.

Capturing environmental dynamism, or the rate and unpredictability of change in the external environment (Dess & Beard, 1984) is a challenge for innovation researchers. While scholars have connected dynamism measures both directly to firm performance (e.g., Garg et al., 2003) as well as indirectly as moderators of firm-level constructs that influence performance (Burns & Stalker, 1961; McArthur & Nystrom, 1991; Priem et al., 1995), the empirical operationalisation of dynamism has always remained challenging.

Our approach seeks to address a limitation in most empirical patent research that relies on a specific instance of the USPTO classification system. Because the USPTO regularly and retrospectively updates how it classifies patents, research that relies on USPTO classifications uses information about the knowledge structure that was not necessarily available at the time an invention was submitted. By using advanced textual analysis, we can create two complementary perspectives on how the knowledge structure evolves in dynamic knowledge landscapes without relying on this ex post-structural knowledge. Topic modelling is gaining in popularity in management research

(Hannigan et al., 2019) with researchers using the approach to better understand amongst others firm diversification (Choi et al., 2021) and breakthrough innovation (Kaplan & Vakili, 2015).

First, every quarter, we progressively update the knowledge landscape by investigating the topics in patent documents that were approved in the most recent 36 months. Within these quarterly snapshots of the knowledge landscape, clusters of high and low density can be identified, which are indicative of strong or low inventive attention to a latent topic. A firm's innovation portfolio is positioned relative to clusters of different density in these quarterly changing landscapes, which enables the firm to exploit distinct search capabilities.

Second, we retrospectively update the knowledge landscape, by looking at how the density has shifted through the introduction of new inventions. This allows us to expose how new information has changed the connections between prior inventions and thus altered the knowledge structure. While codified technological knowledge embedded in patents can be understood as a public good (Arrow & Universities-National Bureau, 1962), much of the knowledge required for innovative activity remains tacit so that important knowledge components remain private even after disclosure (Cowan et al., 2000). Yet over time, new patented innovations are being disclosed that may expose previously tacit knowledge in patents and their interlinkages.[2] Thus, the perspective of retrospective dynamism recognises that the knowledge structure, can shift when new knowledge is added to the environment. This structural shift is difficult to discern when relying on a singular instance of the USPTO classification scheme. This perspective thus recognises that every innovation not only becomes embedded in that landscape but can in fundamental ways change the topography of the knowledge landscape itself.

Dynamic landscapes with sparse and contested clusters

Search scope

We posit that search scope will benefit firm performance, especially in sparse clusters. Having a broad knowledge base can have positive outcomes for organisations (Felin & Zenger, 2014), because it increases the number of new ideas that firms introduce, and helps reduce organisational myopia as the firm can approach problems from distinct angles (Knudsen & Srikanth, 2014; Laursen & Salter, 2006; Von Hippel, 1988). In addition, having the knowledge and capabilities to search broadly facilitates the establishment of connections between domains and increases the variety of available knowledge to solve problems (March, 1991), which positively relates to innovation impact (Fleming & Sorenson, 2001; Kotha et al., 2011). It is therefore no surprise that the breadth of a team's knowledge increases the team's ability to engage in distant recombination and increases the likelihood of breakthrough inventions (Kneeland et al., 2020).

The positive effects of search scope are heightened in sparse clusters, i.e., those characterised by disjointed and seemingly scarce innovation activity across rugged landscapes where the connections between the individual innovations are weak or poorly understood (Yayavaram & Chen, 2015). First, firms with a broad knowledge base are better at knowledge brokerage (Carnabuci & Bruggeman, 2009; Paruchuri & Awate, 2017). While technological and organisational integration costs limit unchecked exploration (Katila & Ahuja, 2002), search scope will provide the firm with an important diversity-based advantage in sparse clusters. This knowledge diversity is less useful in contested clusters in which the valuable knowledge elements are well-known. Second, broad search reduces risks due to unforeseen events because a broader knowledge base is less likely to lose its value entirely due to a gale of competence-destroying creative destruction (Schumpeter, 1942). A more diverse knowledge base has potential for redeployment, providing the firm with real options to approach innovation challenges in distinct ways. Relatedly, diverging knowledge sources can serve as starting positions from which to initiate an innovation process. Thus, scope does not only allow brokerage between distant knowledge elements but it also enables the firm to start innovation processes from distinct vantage points. This latter capacity is especially relevant in sparse clusters in which no predictable path to a valuable peak exists. Therefore, we hypothesise that:

Hypothesis 1: Search scope's positive effect on firm performance is strengthened when a firm's innovation portfolio's position is proximal to sparse clusters.

Search commitment

For firms that operate in the knowledge economy, i.e., in industries in which R&D correlates with performance, commitment of significant resources to search is likely to reap rewards. Capon, Farley, and Hoenig's meta-analysis (1990) found about twice as many studies established a positive relationship between R&D investment and financial performance than a negative relationship. High search commitment raises the bar for satisficing search (Simon, 1978), which likely leads to better outcomes. Specifically for complex projects with uncertain outcomes, Pich et al. (2002) argued that it pays to search along multiple alternative problem-solution paths independently and select afterwards. Relatedly, Knudsen and Srikanth (2014) suggest that early coordination in the search space tends to heighten organisational myopia, which has negative performance implications. These arguments suggest that search commitment should contribute to improved performance because without significant commitment pursuing multiple pathways at once is impossible. Yet search commitment requires investment in time, human capital, and equipment and thus comes with opportunity costs. Therefore, even firms with high search commitment are likely to experience different performance effects depending on their portfolio's position in the evolving knowledge landscape.

We proffer that search commitment enhances performance especially near sparse clusters. Sparse clusters lack a transparent knowledge structure so that the landscape remains rugged and surprising, and significant search commitment can be beneficial as the firm's continued investment will be in areas it considers promising given its own expertise (Courtney et al., 1997). Because sparse clusters are characterised by unexploited and unexplored innovation potential, committing financial resources in the vicinity of sparse clusters is likely to pay off and is unlikely to risk competitive crowding. High search commitment combined with a portfolio in the vicinity of sparse clusters can benefit the firm financially because there are few competitors active in this zone of the innovation landscape (Hannan & Freeman, 1977; Swaminathan, 2001). High commitment near contested clusters is likely to be associated with ineffective slack search. When too many resources are poured into a well-established area, the crowding effect is likely to significantly reduce the potential upside. We therefore hypothesise;

Hypothesis 2: Search commitment's effect on firm performance is stronger when a firm's innovation portfolio's position is proximal to sparse clusters.

Search specialisation

The knowledge-based view, with its roots in information processing theory and Simon's (1978) principle of bounded rationality, suggests that efficiency in knowledge production – including its creation, acquisition and storage – calls for specialisation rather than scope (Grant, 1996). Specialisation can be understood as the outcome of a consistent local search strategy through which the firm develops domain-specific absorptive capacity and shapes both its organisational capabilities as well as their limits (Cohen & Levinthal, 1990; Cyert & March, 1963; Simon, 1982; Zou et al., 2018). The limits of local search have been extensively debated. They relate to myopic behaviour, reinforcement of cognitive biases, figurative blindness for distant domains and, more generally, inertia which can lead to organisational death (Laursen, 2012; Levinthal & March, 1981; March, 1991; Zajac & Bazerman, 1991). Therefore, there is 'near consensus on the need for balance' between local and more distant search, or between exploitation and exploration (Gupta et al., 2006, p. 697). However, this proclaimed 'need for balance' is agnostic about the characteristics of the knowledge structure. We proffer that search specialisation, rather than balance or exploration, will be beneficial to performance when the firm's innovation portfolio's position is near contested clusters because of opportunity crowding, knowledge spillovers, and appropriation.

Overall, specialist organisations tend to be niche players that compete at what for many other players is the periphery but what for them is a core market. Ecological theory postulates that specialists are more likely to appear when the environment consists of a narrow range of resources, while munificence within that range enhances their chances of survival (Hannan & Freeman, 1977; Swaminathan, 2001). Contested clusters rely on such a narrow subset of knowledge components and specialists devote most (if not all) of their attention to in-depth exploitation of a single domain (Swaminathan, 2001). Unlike knowledge brokers that recombine heterogeneous input ideas creatively, specialists focus on recombining homogenous knowledge components (Carnabuci & Bruggeman, 2009). Because there are limited opportunities to create valuable breakthroughs with such homogeneous components, only those firms that are able to move with focus and expertise are likely to find and exploit valuable opportunities, while less-specialised actors will be crowded out or will only be able to engage in marginal work that lacks significant value-creating potential (George et al., 2008). Having access to in-depth knowledge due to specialisation in contested clusters is a prerequisite for successful innovation because such local knowledge is instrumental to making cognitive breakthroughs associated with high firm value (Kaplan & Vakili, 2015). Specialisation is thus favoured.

Knowledge diffusion in contested clusters is not asset-independent because spillovers tend to result from heterogeneous and asymmetric firm knowledge bases that differentiate firms' absorptive capacity (Giuliani, 2007). Specialist firms with a strong knowledge base in the focal domain will be perceived as technological leaders and be targeted for advice, whereas others with weaker knowledge bases will lack the absorptive capacity to master the domain (Cohen & Levinthal, 1990; Giuliani, 2007). In a non-insular innovation domain to which multiple industries contribute (like agriculture), the presence of competing and divergent knowledge bases is therefore associated with knowledge spillover directionality favouring specialised firms that benefit proportionally more from patent-related disclosures (James et al., 2013). For firms that lack search specialisation, spillovers are likely to be accidental knowledge leaks or requisite disclosures when asking others for help, and could reduce the firm's ability to appropriate the financial rewards of its R&D investments (Alnuaimi & George, 2016). This leads us to conclude that search specialisation results in higher firm performance when the firm's innovation portfolio's position is proximal to contested clusters. Close to sparse clusters, search specialists will be lost in a rugged landscape in which the connections and couplings are not yet well established, in which they lack brokerage skills, and in which their local search strategy does not fare well (Fleming & Sorenson, 2001, 2004; Yayavaram & Chen, 2015). This leads to the third hypothesis:

Hypothesis 3: Search specialisation's effect on firm performance is strengthened when a firm's innovation portfolio's position is proximal to contested clusters.

Retrospectively dynamic landscapes and emergent clusters

We operationalise retrospectively dynamic landscapes as landscapes in which the interpretation of public knowledge and/or knowledge connections changes when new knowledge, in the form of newly issued patents, is disclosed. Within those landscapes, we focus on clusters that are characterised by changes in the knowledge structure that lead towards higher clarity of interconnections between elements in the innovation space, i.e., emergent innovation clusters in which new knowledge is increasing industry understanding and reducing knowledge ambiguity. Thus, retrospectively dynamic landscapes suggest a knowledge structure shift where the knowledge landscape dynamically evolves as new inventions take place.

Search scope provides the firm with a broad arsenal of knowledge components with which it can approach its innovation agenda. Near emergent clusters, characterised by an increasing density and industry awareness of how successful recombination happens (e.g., an emergent dominant design), search scope will only be useful if the firm's positioning near emergent clusters is a deliberate strategic choice. This is possible, but hard to ensure. Given the firm's broad array of search and knowledge capacities, its focal innovation portfolio will straddle multiple knowledge domains and hence proximity to emergent clusters is not necessarily the outcome of a deliberate focus of attention to a new and nascent technological niche. Moreover, even if the positioning is deliberate, the emerging stability of knowledge recombination favours local search strategies that fine-tune the requisite combinations: high search scope generalists may lack the capability to switch from their explorative mindset to a more exploitative mindset, which is required to develop novel and valuable ideas (Yayavaram & Chen, 2014).

Search specialists have a focused portfolio of innovations and have developed the absorptive capacity to learn the intricacies of interrelated knowledge in more predictable landscapes. When specialists position their innovation portfolio near an emergent cluster, the alignment between this portfolio and their historical capabilities and skills is likely to be high. Due to their narrow search focus, they are after all less likely to position their portfolio by happenstance close to an emergent cluster. In addition, search specialists lean towards exploitation of familiar landscapes and the recombination of a small number of homogeneous components (Carnabuci & Bruggeman, 2009), which will benefit them in emerging clusters in which density and homogeneity are increasing. Thus, in comparison to firms with broad search scope, the probability of deliberate positioning near emergent clusters is higher for search specialists and they are also more likely to have the skills to engage in value-creating innovation in those clusters.

Finally, when interdependencies among search elements are well understood, experiential search is likely to lead problem solvers to the highest peaks (Caner et al., 2017; Levinthal, 1997). In emergent clusters, interdependencies are becoming clearer but still require further investment before they turn into contested clusters (in which specialists excel). It is therefore not surprising that incumbents are able to enter new and emerging niches at the right moment in order to pursue valuable opportunities (King & Tucci, 2002). Felin and Zenger (2014) postulate that when hidden knowledge is low and problem complexity is high, as is the case when innovating in emergent clusters, the optimal search solution is centralised and theory-guided. Firms are more likely to engage in theory-guided research when they are able to commit significant resources to search and employ scientists with advanced degrees who are more likely to rely on theory and also are more expensive to employ. High search commitment allows for the necessary experimentation which could help firms whose portfolios are positioned near emergent clusters. Given that all three forms of search strategies may benefit, but with varying degrees of effectiveness, we develop a comparative hypothesis:

Hypothesis 4: Firms with strong search commitment benefit more from proximity to emergent clusters in terms of firm performance than firms with high search specialisation and high search scope.

Dynamic portfolio position in semantic landscapes

To determine the strategic network position of a firm's innovation portfolio, prior literature has often relied on prior art citations that reflect explicit ties between patents that enable the construction of citation networks that structure the knowledge space (Yayavaram & Ahuja, 2008). While this approach is useful, it suffers from important shortcomings: First, a relation between two patents is established only when there is a citation and sometimes examiners and/or inventors can opt to simply use a single citation where alternatives were possible, so that citation networks can be incomplete. Second, even if all valid citations were added, one is still not able to derive the degree of the relationship between the patents. A third issue is that there could be bias in prior art citations. Inventors can have a tendency to cite their own patents even if the claims in the current patent do not warrant such citations, and examiners tend to have some favourite patents in their directly accessible memory (often those patents with a very broad and well written body) that they regularly cite because of its broad descriptions [interviews with USPTO examiners, 2016].

Technological classifications provide an alternative tool to categorise the formation of knowledge networks because these classifications are used by examiners to facilitate their own search activity when reviewing new patent applications. In terms of process, a patent expert with broad knowledge about the relevant field of the invention assigns a specialised examiner to a patent based on an initial crude classification into a primary class and subclass after which the examiner receives the patent in her stack for review. The examiner accepts to further examine the application for novelty and usefulness or rejects it, in which case it gets sent back up the chain and sent down to someone with more relevant expertise. Over the course of the review, multiple classifications are added when the examiner considers the patent to be a potentially valuable source for related areas as well [interviews with USPTO examiners, 2016]. While the likelihood of not including specific classifications is lowered by the private incentive examiners have to be exhaustive (to facilitate future search), the inability to discern degree of relationship between two classifications persists here as well. Moreover, while classification schemes evolve regularly,[3] researchers tend to use a specific instance of the classification scheme, thereby ignoring the dynamic aspects of an ever-evolving classification structure. We address these measurement limitations by using novel and advanced textual analysis tools described next.

Topic modelling via Latent Dirichlet Allocation (LDA)

Textual analysis of patent documents can be used to gain new insights into innovation practices. Kaplan and Vakili (2015) studied fullerene patents using topic modelling – an analysis technique that 'uses the co-location of words in a collection of documents to infer the underlying (or latent) topics in those texts and weight of each topic in each individual document' (p. 6 ev) – to identify those patents that deliver a cognitive breakthrough by being the first document to be linked heavily to a newly emerging topic. We use LDA-based topical modelling to 'uncover automatically themes that are latent in a collection of documents and to identify which composition of themes best accounts for each document' (Kaplan & Vakili, 2015, p. 12 ev). LDA thus assigns a numerical vector, based on a predetermined number of topics, to each individual text (patent title + abstract), using the information of all texts in the selected environment, which consists of all the agricultural patents in a specific period. We then calculate the Hellinger distance between each pair of vectorised texts to capture the interlinkage among patents (Blei et al., 2003). We follow Nikita's (2016) tuning approach to determine the optimal number of topics 'm' depending on sample size. Unsurprisingly, this method leads us to use more topics to adequately capture the topical variety if the focal environment consists of innovations during a 3-year [t-3, t-1][4] period (m = 195) than a 1 year period (m = 100).

Given a specific period, quantisation of all patent text data creates a topic probability distribution vector for each patent. These vectors serve as building blocks for the formation of semantic spaces. We are interested in three types of reference locations: sparse, contested, and emergent locations that serve as epicentres for their respective clusters (see Figure 1). To identify the progressively dynamic locations, we calculate for each patent in period [t-3], the average distance to its 100 Nearest Neighbours in period [t-3, t-1] (kNN with k = 100), which gives us a density measure. For each period, we isolate the top 5% in terms of high density (low distance) and call these contested locations, and low density (high distance) and call these sparse locations. To identify the retrospectively dynamic or emergent locations, we compare the average distance between each patent and its 100 Nearest Neighbours in period t-3 with the average

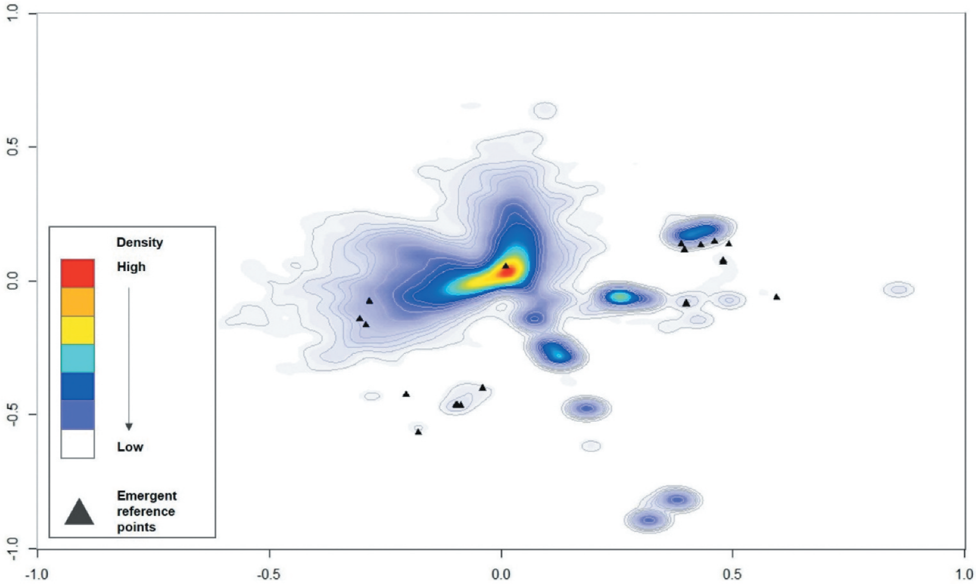

Figure 1. Contested and sparse clusters and emergent cluster reference points in two-dimensional space.

distance between the patent and the same 100 Nearest Neighbours in period [t-3, t-1]. We identify the top 5% of patents whose 100 Nearest Neighbours are closer (density increase) in the longer period than in the 1-year period as emergent locations.

The Hellinger distance captures the distance between probability distribution vectors (Blei & Lafferty, 2009; Nikulin, 2001). For two vectors $x_p = (x_{p1}, \ldots, x_{pi}, \ldots, x_{pm}) and x_q = (x_{q1}, \ldots, x_{qi}, \ldots, x_{qm})$ with m distinct topics in X, the Hellinger distance is constrained between 0 and 1 and defined as per the formula below. The focal density (or sparseness) measure takes the average of this distance over the k = 100 nearest neighbours (kNN):

$$d_H(x_p, x_q, X) = \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^{m} \left(\sqrt{x_{pi}} - \sqrt{x_{qi}}\right)^2} \text{ [Hellinger Distance between two patents]}$$

$$\rho(p, S) = \frac{\sum_{q \in S_k} d_H(p,q,S)}{k} \text{ [Density } \rho \text{ in semantic space S of size k = 100 NN]}$$

Our study period ranges from 1995–01-01 to 2011–12-31 (17 years, 68 quarters). Each patent is allocated, based on its application date, to the quarter Qi_t with I = 1, 2, 3, or 4 and t = `95 ... `11 in which the patent was filed [Q1_95, Q4_11]. We define a rolling window of length 12 quarters – a period [t-3, t-1] – before the focal quarter. Given a patent p, we first identify its quarter Qi_t and the start date of that quarter after which we trace back all patents for 12 preceding quarters (3 years) applied for before Qi_t's start date. For our progressively dynamic clusters, we thus compare the firm's quarterly patent portfolio with the sparse and contested reference locations from period t-3, based on the information available in the semantic space during period [t-3, t-1]. For example, for a patent applied for in Q2_00, we look for the reference sparse and dense locations of patents applied for between [Q2_97, Q1_98], while the topic probability distribution vectors use all information from period [Q2_97, Q1_00]. We assess for each patent from Q2_00 the proximity to all reference locations and take the average of the 'q'[5] smallest values (i.e., the 'q' most similar reference points to each firm patent). We then aggregate these patent average distances per quarter and per fiscal year to create our moderating variables.

Data and methods

To test our theory, we rely on three data sources. First, we combine quantitative variables from the USPTO patent dataverse (Hall et al., 2001; Kogan et al., 2012), made available via the PatentsView data portal, with financial reporting information available from WRDS' Compustat. Management, finance, and economics scholars have used similar sources to study innovation and its relation to desirable outcomes such as impact (Fleming, 2001; Fleming & Sorenson, 2004; Nerkar, 2003; Trajtenberg, 1990), new product introductions (Katila, 2002), licencing likelihood (Mowery et al., 2001) as well as financial performance (Cockburn & Griliches, 1987; Lin et al., 2006). Deng et al. (1999) argued that patent quantity and impact are associated with future performance of R&D-intensive companies. Similarly, Bloom and Van Reenen (2002) established a positive relation between patents and market value. Also, Hall et al. (2005) found that patent output relative to R&D or average number of citations per patent relates positively to firm performance.

We extract all 67,120 utility patents from IPC class A01 (agriculture, forestry, animal husbandry, hunting, trapping, fishing), applied for between 1995 and 2011 inclusive, and develop aggregate measures at the firm-year level. Agriculture has scarcely been used as a context within which to study innovation, despite its importance for sustainability (Pigford et al., 2018). We add financial information from both the North American and the Global Compustat database and convert all data to USD by using the foreign exchange rate at the final day of each firm's fiscal year. In order to merge the financial data with the patent data, we rely on the reported company name (Compustat) and the assignee organisation (PatentsView). Following the n-gram similarity method described in Kondrak (2005), we compare the two lists and use the following heuristic: every match with n-gram similarity above 0.70 is manually checked for confirmation while the others are considered as different companies. This approach thus excludes subsidiaries with very different names than the parent company. Following this method, we end up with 560 firms for which we have 1,797 firm-year observations. For these firms, we also extract all the non-agriculture patents from the 1995–2011 period to be able to determine search scope and search specialisation.

Firm performance

As dependent variable (DV), we use the natural logarithm of the firm's annual EBITDA (Earnings Before Interest, Taxes, Depreciation and Amortisation), derived from Compustat in time t +1.

Search Scope is an explanatory variable that captures the breadth of a firm's knowledge base, which is the result of the firm's historical research activities. It is measured as 1 – HHI with HHI = the Herfindahl index of all the firm's patents' 3-digit classifications applied for in period [t-3, t-1]. *Search Commitment* is reflected in the built-up knowledge base of the firm over time. We proxy the firm's commitment to engage in innovative search by calculating the natural logarithm of the sum of its R&D expenses (i.e., commitment of financial resources) in year t-1, 80% of R&D expenses in year t-2, and 65% of R&D expenses in year t-3 (see, Hall (2007); Sandner and Block (2011) for a similar approach). We add these weights to account for knowledge devaluation and leakage, imperfect knowledge retention, and the tightening of the opportunity space for older knowledge (Dosi, 1982; Huber, 1991; Nerkar, 2003; Walsh & Ungson, 1991). We divide this value by its sample maximum to ensure the search commitment variable moves between 0 and 1 which facilitates comparison across the predictors. *Search Specialisation* captures the extent to which the firm's innovative effort is focused on the agricultural industry. It is operationalised as the number of firm patents in period [t-3, t-1] that belong to the agricultural class A01, divided by the total number of patents of the firm in the same period.

Proximity to progressively dynamic clusters

This aggregated measure captures for the firm's entire patent portfolio in a given fiscal year, the average distance between each firm patent and its q = 3 nearest contested (sparse) reference locations. Those q = 3 locations are a subset of the top (bottom) 5% of patents applied for during t-3, created with topical distribution vectors using information from [t-3, t-1]. This results in two measures, i.e., *distance to contested clusters* (reference location = top 5% highest density clusters) and *distance to sparse clusters* (reference location = top 5% lowest density clusters). Note that the q = 3 selected reference locations can differ for each patent and that the resulting measure is a network average of the distance between a focal patent P's quantised vector and the 3 most relevant (in terms of Hellinger distance) identified reference locations. Eventually, we include 1 – the resulting value in our regression as the proximity to either sparse or contested clusters (see Table 1).

Proximity to retrospectively dynamic, emergent clusters

First, we determine for every patent the density with its 100 Nearest Neighbours in the year t-3. Then, we calculate the density shift as the density of the same 100 Nearest Neighbours in a longer period [t-3, t-1],[6] minus the density value determined before (for year t-3). While this value can positive, we only focus on negative density shifts that are characterised by patents becoming more similar or interlinkages between patents becoming clearer following the addition of new information. Next, we isolate the top 5% of patents that are characterised by the most negative density shifts and call those the emergent locations. Then, we calculate for each patent in the focal quarter Qi_t, the distance to each of all the emergent locations. We then take the q = 3 lowest values (lowest distance to emergent locations) per patent and average those over the quarterly portfolio. We do this for each quarter in each calendar year (note that for each quarter the starting point of period t-3 shifts with 3 months). Finally, we average the values for four quarters based on the firm's fiscal year. We again use 1 – the resulting value to capture the proximity to emergent clusters.

Table 1. Stepwise explanation of Strategic Positioning in Semantic Technological Landscapes

Objective	Steps
Identify reference locations for P_{it} to determine static (sparse and contested) clusters	(1) For quarter Q in period t_s, identify all patents in period t-3 (2) Code all patent vectors from period t-3 with information from period [t-3, t-1] as semantic vectors of length M = 100 (3) Determine for each patent from period t-3 the average Hellinger distance from its 100 nearest neighbours from period [t-3, t-1] (4) Select the top 5% of patents from period t-3 with highest (lowest) average density and call these contested (sparse) locations
Identify reference patents for P_{it} to determine dynamic emergent clusters	(1) For quarter Q in period t_s, identify all patents in period t-3 (2) Code all patents in period t-3 as semantic vectors of length M = 100 (3) Determine for each patent from period t-3 the average distance from its 100 nearest neighbours in the same period t-3 (4) Determine for each patent from period t-3 the average distance from the same 100 nearest neighbours as identified in step 2. This time take into account all information (and possible new topics) available in period [t-3, t-1] (5) Select 5% with strongest reduction in average distance and call these emergent locations
Determine portfolio position	(1) Calculate Hellinger distance between every firm patent and each reference location (sparse / contested / emergent) in a specific period (2) Take the minimal distance of each firm patent portfolio to its three nearest reference (sparse / contested / emergent) and average this number (3) Aggregate the averages across all firm patents per quarter Q (4) Aggregate the quarterly values to fiscal years (5) The resulting value is the portfolio's average Hellinger distance from the reference (sparse / contested / emergent) locations and is constrained by [0, 1]. We use 1 – this value to capture the proximity of the firm's portfolio to sparse / contested / emergent locations.

We add controls that have been shown to absorb variance in firm performance such as firm size (log number of employees), log of firm sales, and the value of a firm's total assets. In addition, we control for the log of the maximum number of patents the firm has in any IPC class outside of the agricultural field (global depth) and add a control for the firm's agricultural knowledge diversity by calculating a Herfindahl index of the number of patents the firm has within each of the IPC WIPO A01 subclasses. We add unreported dummy controls for the fiscal year, the two-digit SIC codes between 21 and 39 (i.e., group D Manufacturing), and for five regions (North-America (USA and Canada), European Continent, Asia (Taiwan, South Korea, Japan, and India), Anglo-Saxon (i.e., Great-Britain, Ireland, and Australia), and rest of the world (Bahamas, Brazil, and Israel).

Analysis and results

Table 2 reports descriptive statistics and correlations. Given the high correlation between the log for sales and employees, it is no surprise that the variance inflation factors, after running simple OLS, of those two measures exceed 10. Including only one of them had no impact on our results so we decided to leave them both in as they merely serve as controls. Note that the VIFs for the interaction terms are larger than 10, as is expected, but that this does not affect the obtained p-values (Allison, 2012). In addition, it is

Table 2. Descriptive statistics and correlations.

No.	Variables	Mean	s.d.	Min	Max	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1)	Performance (log EBITDA)	6.80	1.05	0.69	10.20											
(2)	Proximity to Sparse Cluster	0.47	0.04	0.31	0.73	0.12										
(3)	Proximity to Contested Cluster	0.58	0.10	0.31	0.86	−0.02	0.09									
(4)	Proximity to Emergent Cluster	0.63	0.07	0.38	0.85	0.02	0.34	0.61								
(5)	Search Scope	0.68	0.18	0.00	0.96	0.34	0.07	−0.10	−0.12							
(6)	Search Specialisation	0.17	0.22	0.00	1.00	−0.36	−0.04	−0.06	−0.02	−0.46						
(7)	Search Commitment	0.53	0.20	−0.14	1.00	0.78	0.09	0.02	0.12	0.36	−0.54					
(8)	Global Depth	2.38	1.30	0.00	6.67	0.62	0.09	0.02	0.03	0.37	−0.60	0.70				
(9)	Local HHI	0.78	0.31	0.00	1.00	−0.26	−0.02	−0.03	0.02	−0.19	0.04	−0.21	−0.25			
(10)	Sales (log)	6.37	3.22	−5.52	12.13	0.78	0.15	−0.10	−0.07	0.42	−0.42	0.73	0.57	−0.25		
(11)	Number of employees (log)	0.97	2.53	−6.21	5.95	0.81	0.15	−0.06	−0.08	0.43	−0.47	0.77	0.62	−0.25	0.95	
(12)	Total Assets	9,355	20,394	0.08	237,218	0.75	0.10	−0.08	−0.01	0.26	−0.23	0.61	0.45	−0.17	0.52	0.56

reassuring to find that despite high correlations between other variables, the VIFs remain below the critical value of 10 and the standard errors are reasonable, suggesting that the used regression accurately isolates the contributions of the different variables.

We run our analysis in Stata 12 on an unbalanced panel of 1,797 firm-year observations, the results of which are presented in Table 3. We use Swamy–Arora estimators and robust firm-clustered standard errors to account for unobserved heterogeneity and autocorrelation and add firm-specific random effects to control for the differential capacity of firms to appropriate the returns of patents (Cockburn & Griliches, 1987). As a robustness check, we run the same models using Wallace and Hussain estimators and find consistent results. We also run a fixed effect regression with robust clustered standard errors. The results are consistent except for the insignificance of the interaction between emergent clusters and search commitment in model 7. We do not include a lagged response variable because that risks making the estimators of the variance components inconsistent (Mohammadi, 2012).

Hypothesis 1 stipulated that search scope's effect on firm performance would be more positive when the firm's invention portfolio is located near sparse rather than contested clusters. In Table 3, Model 3 indeed exhibits a positive and significant interaction between search scope and proximity to sparse clusters ($b = 2.391$, $p < 0.1$) and no significant effect for interaction with proximity to contested clusters. This aligns with the argument that broad enables knowledge brokerage, which is especially valuable when firms innovate in sparse clusters in which many of the interlinkages between knowledge topics are still unknown or weak at best. Figure 2 exhibits the complex effects of the scope – sparse clusters interaction. Firms perform better when further away from sparse clusters, except if they have very high search scope. For average search scope (0.68), it is still preferred to be distant from sparse clusters ($\mu - 2\sigma$) but this effect reverses when search scope crosses the 0.83 threshold which happens for about 21% of sample observations. Thus, only those organisations with a very broad knowledge base perform better close to sparse clusters when their knowledge brokerage skills give them an advantage over the competition.

Hypothesis 2 predicted that the effect of search commitment on firm performance would be stronger when a firm's invention portfolio is located near sparse clusters. Model 4 provides support for this hypothesis. While Model 2 exhibits an overall positive effect of search commitment on performance ($b1 = 0.425$, $p < 0.05$), model 4 shows this effect is driven by proximity to sparse clusters as the main effect turns negative ($b1 = -1.873$, $p < 0.01$) and the interaction effect is positive ($b2 = 5.321$, $p < 0.001$). When looking at Figure 4, we can see that for a search commitment value of 0.64 ($\mu = 0.52$, $\sigma = 0.2$) a pivot happens. When commitment is below average, staying away from sparse areas is preferred. However, firms with strong search commitment to ($> \mu + \sigma/2$) perform better near sparse areas. Hypothesis 2 is supported.

Hypothesis 3 suggested that search specialisation would have more positive performance effects when the firm' inventive portfolio is located near contested clusters. Model 5 indeed exhibits the predicted positive interaction between search specialisation and proximity to contested clusters, while no such interaction is established for proximity to sparse clusters. Figure 6 exhibits a weak negative relationship between specialisation and firm performance for a portfolio far away ($\mu - 2\sigma$) from contested clusters. Once the firm's research activities are closer to contested clusters, the positive slope appears and increases significantly. For those firms that are only active in the agricultural domain

Table 3. Financial performance implications of search and portfolio position.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Global depth	0.043*	0.039†	0.039†	0.038†	0.039†	0.046*	0.044*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Agriculture HHI	−0.040	−0.037	−0.034	−0.039	−0.040	−0.035	−0.037
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Log Sales	0.036**	0.035**	0.035**	0.035**	0.036**	0.033**	0.035**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Log # Employees	0.181***	0.157***	0.157***	0.159***	0.155***	0.158***	0.158***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Total Assets	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Proximity to		−0.352	−1.999*	−3.357***	−0.271	−0.352	−3.034**
Sparse Cluster		(0.26)	(0.84)	(0.79)	(0.30)	(0.26)	(1.00)
Proximity to		0.159	−0.193	0.329	0.080	0.249	0.119
Contested Cluster		(0.19)	(0.34)	(0.31)	(0.20)	(0.20)	(0.22)
Proximity to		0.021	0.025	0.092	0.021	−2.722***	−2.025**
Emergent Cluster		(0.18)	(0.18)	(0.18)	(0.18)	(0.74)	(0.77)
Search Scope		0.051	−1.402*	0.047	0.054	−0.808*	−1.040†
		(0.07)	(0.64)	(0.07)	(0.07)	(0.37)	(0.58)
Search Specialization		0.094	0.096	0.101	0.052	−1.145**	−1.168**
		(0.07)	(0.07)	(0.07)	(0.44)	(0.38)	(0.40)
Search Commitment		0.425*	0.432*	−1.873**	0.421*	−1.254*	−2.574***
		(0.20)	(0.20)	(0.70)	(0.20)	(0.56)	(0.74)
Sparse	Industry activity –		2.391†				0.407
X Scope	progressively		(1.23)				(1.35)
Contested	dynamic		0.556				
X Scope			(0.52)				
Sparse				5.321***			4.221**
X Commitment				(1.33)			(1.55)
Contested				−0.324			
X Commitment				(0.55)			
Sparse					−0.687		
X Specialisation					(0.84)		
Contested					0.630†		0.711†
X Specialisation					(0.35)		(0.40)
Emergent	Knowledge					1.382*	1.452*
X Scope	structure shifts –					(0.61)	(0.69)
Emergent	retrospectively					2.000**	1.381*
X Specialisation	dynamic					(0.61)	(0.69)
Emergent						2.699**	1.646†
X Commitment						(0.85)	(0.99)
Constant	6.070***	5.934***	6.909***	7.191***	5.936***	7.570***	8.459***
	(0.11)	(0.15)	(0.44)	(0.38)	(0.16)	(0.43)	(0.55)
N	1797	1797	1797	1797	1797	1797	1797
N_g	571	571	571	571	571	571	571
df_m	40	46	48	48	48	49	52
r2_w	0.338	0.337	0.340	0.347	0.339	0.342	0.350
r2_b	0.802	0.807	0.807	0.806	0.807	0.810	0.808
r2_o	0.821	0.826	0.826	0.826	0.826	0.829	0.828

p-values in parentheses, †$p < 0.10$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

(specialisation = 1), the economic difference between positioning far away or close to contested clusters (4σ) is over 260,000 USD in EBITDA. This value is significant is as it is 50% above the median EBITDA in the sample (174,036 USD) and about 19% of the mean EBITDA (1,361,000 USD).

For hypothesis 4, the results in model 6 exhibit significant and positive interactions for all search dimensions and proximity to emergent clusters. The interaction effect with search commitment has the most positive coefficient, which seems to support our

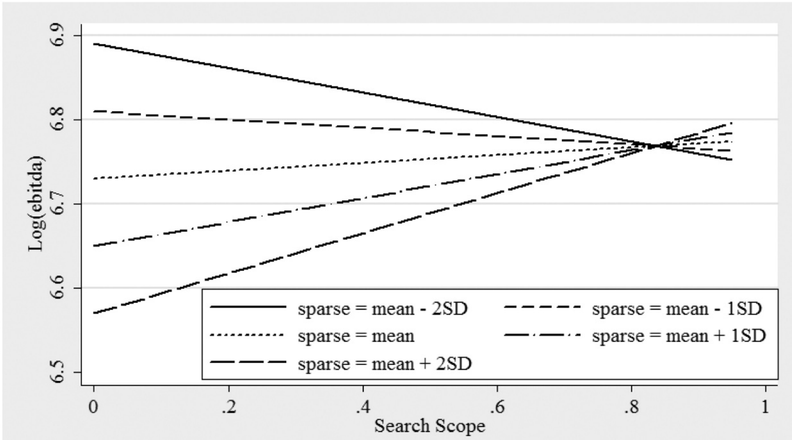

Figure 2. Search scope and proximity to sparse clusters' effect on firm performance.

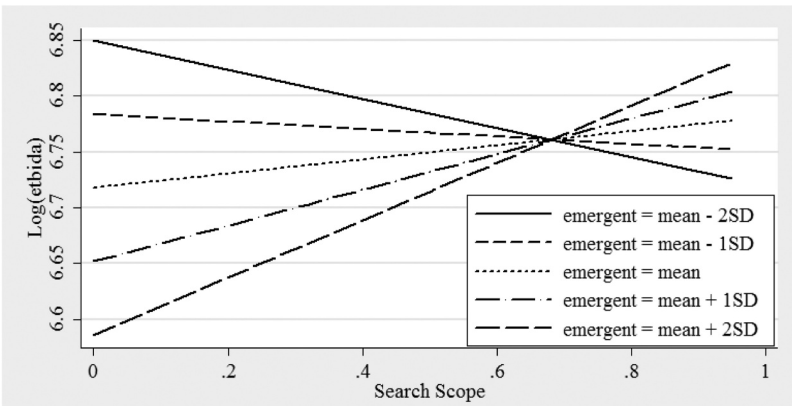

Figure 3. Search scope and proximity to emergent clusters' effect on firm performance.

argumentation. To better assess the hypothesis, we rely on both a graphical interpretation and a calculation of effect sizes. We argued that in emergent clusters, the need for deep commitment to search stems from the increasing clarity of the interconnections and interdependence, which makes it easier to predict viable avenues for future growth. Yet additional resource investments and more exploration are required to fine-tune the emerging dominant design in the emergent cluster. Because of this, it still takes search commitment to be able to drive performance, rather than only specialisation.

Figures 3, 5, and 7 exhibit the marginal effects of the three discussed search dimensions on performance, when moderated by proximity to emergent clusters. The slopes overall become more positive as proximity to emergent clusters increases. For search scope and search specialisation, the slopes even turn negative if the proximity is below average. Firms with low search commitment, for instance, should steer clear of emergent clusters and should direct their inventive attention to areas where the

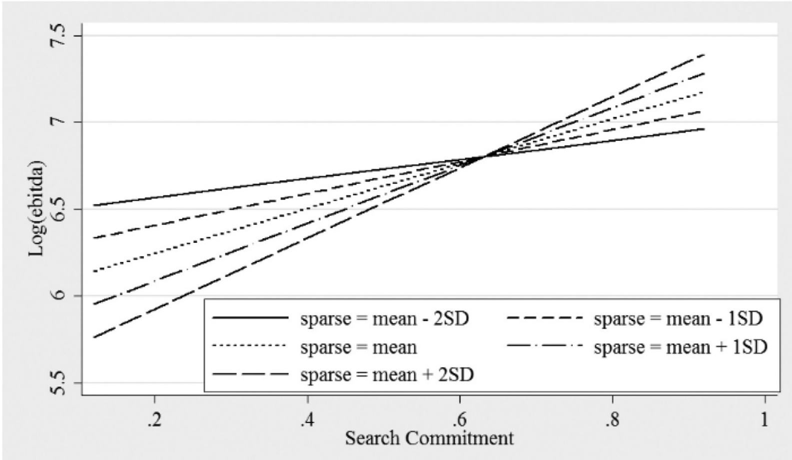

Figure 4. Search commitment and proximity to sparse clusters' effect on firm performance.

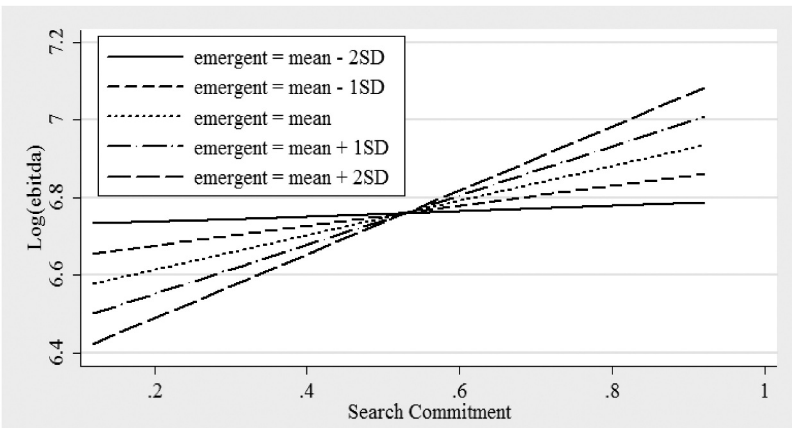

Figure 5. Search commitment and proximity to emergent clusters' effect on firm performance.

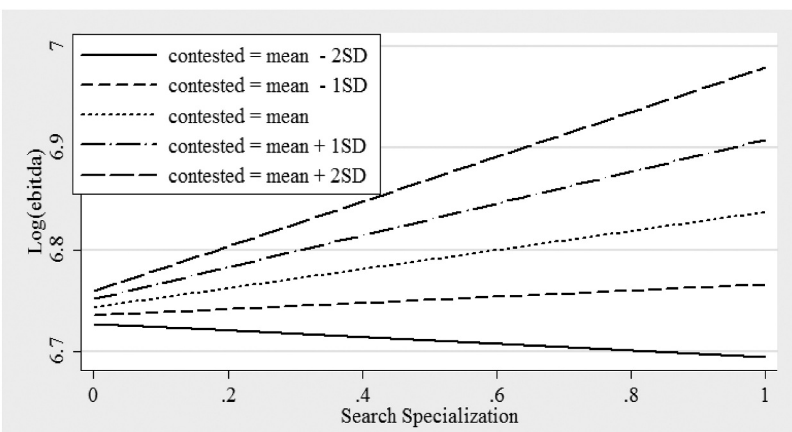

Figure 6. Search specialisation and proximity to contested clusters' effect on firm performance.

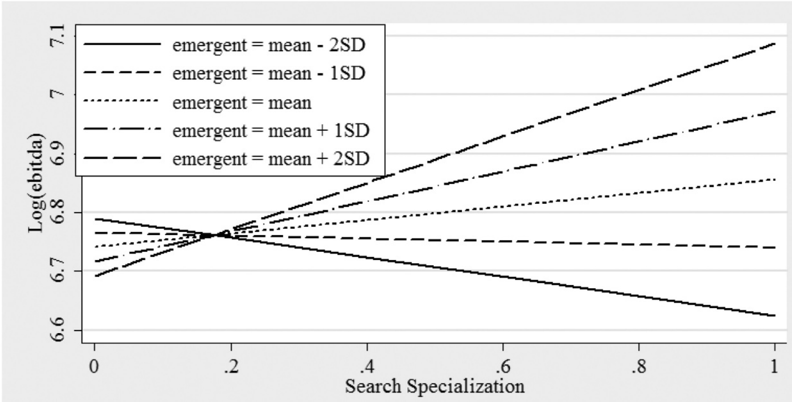

Figure 7. Search specialisation and proximity to emergent clusters' effect on firm performance.

intricate connections between the various topics of the knowledge infrastructure are still more ambiguous. In those areas, lower search commitment pays off as it is associated with lower costs and higher performance. It is only when search commitment is around its mean value of 0.52 that proximity to emergent clusters starts to result in a performance advantage.

We see a weak positive effect of search scope – proximity to emergent clusters gets weakly rewarded once search scope exceeds its mean value (0.68). Similarly, we see the same effect for search specialisation that also reaches the critical juncture around its mean value (0.18). Figure 8 gives a comparison between the three search dimensions at their respective means ± 1 or 2 standard deviations at average distance from emergent clusters and at mean values for all other variables and with the dummy variables set to zero. That image shows clearly the negligible difference between the moderated effects of search scope and search specialisation while search commitment's effect grows. At a higher distance from emergent clusters, the effects of both search scope and specialisation become negative while close to emergent clusters the effects angle up a little more.

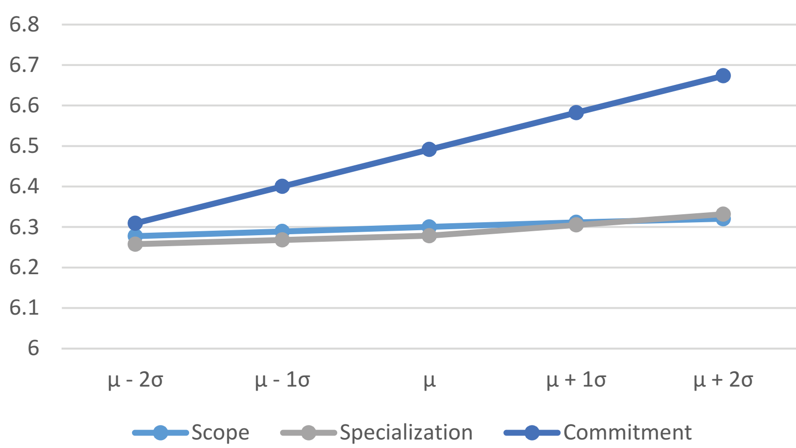

Figure 8. Performance impact of search at mean distance μ from emergent clusters.

This confirms our fourth hypothesis that search commitment is required to boost performance when innovating near emergent areas. The difference between the moderated effects of search scope and specialisation, compared at their respective means, is insignificant.

Discussion

Building on the literature on search in technological landscapes, we investigated how search dimensions and a firm's inventive portfolio's position in the evolving knowledge landscape interact to drive financial performance at the firm level. By adopting novel data science techniques to analyse textual data in combination with network analysis, we were able to create dynamic, semantic landscapes that illuminate both changes in recent industry activity as well as in industry understanding. Our approach extends prior literature on static landscapes by focusing on various areas of interest in a dynamically evolving knowledge landscape. We discuss two interrelated aspects of our findings that speak to ongoing theoretical debates about organisational ambidexterity (search scope v search specialisation) while acknowledging the importance of search commitment more generally. We also take a closer look at environmental dynamism and the function search has in changing technological landscapes. Finally, we explore some of the possibilities our methodology offers for other fields of enquiry.

Revisiting the exploration – exploitation debate

Almost 80 years ago, Schumpeter (1939) described innovation as a process of recombination of different building blocks, which initiated strands of research that sought to explain how this recombination succeeds. Search became a fundamental theoretical construct that explains how firms innovate and adapt to changing competitive landscapes (Levinthal, 1997). Recombinant search, as Fleming (2001) calls it, is on average easier and more successful when it combines similar elements (local search) than when it combines more divergent elements (distant search) which entails higher risks but potentially also reaps higher rewards. Yet, scholars have long stressed the need to balance both search dimensions to optimise organisational outcomes (Gupta et al., 2006; March, 1991). A recent article, for instance, shows that the firms that succeed at breakthrough innovation initially explore unfamiliar terrain and eventually exploit their recently accumulated knowledge (Kinga Randle & Pisano, 2021).

Our operationalisation recognises that search scope (exploration) and search specialisation (exploitation), need not exist on a continuum but can occur concurrently or sequentially. Recent studies have shown the contingent effects of these strategies for instance, finding that an exploitation strategy can only lead to competitive advantage if the firm also has high strategic agility (Clauss et al., 2021). Our findings reveal no significant independent effects of either scope or specialisation, suggesting that neither is on average performance-enhancing in the agricultural industry. The effects only materialise when considering the context of the evolving knowledge landscape.

Our findings suggest most firms do not benefit financially when their patent portfolio position is located near sparse clusters. Only those firms with very high search scope or commitment can do better in the vicinity of sparse clusters than at a distance. We also find that at high distance from contested clusters, increasing specialisation seems detrimental to performance. This implies that the benefits of search scope, specialisation, or commitment are contingent on the knowledge landscape and the firm's portfolio position within said landscape. This aligns with the generally accepted contingency perspective in that 'the appropriateness of each mechanism [is] a function of environmental and organizational context' (Gupta et al., 2006, p. 698). One should thus not absolutely favour exploration over exploitation or vice versa.

Our study of search dimensions and environmental contingencies provides new insights in questions raised by Gupta et al. (2006) regarding orthogonality and continuity as well as ambidexterity versus punctuated equilibrium. The question of whether exploration and exploitation are continuous or orthogonal constructs is sometimes answered by the creation of a single variable that ranges from more exploitative to more explorative search. This notion is increasingly debunked however, as scholars recently showed that exploration and exploitation invoke different brain regions and cognitive processes, suggesting they do not operate on a continuum (Laureiro-Martínez et al., 2015).

Katila and Ahuja (2002) on the contrary test orthogonality by creating distinct measures for exploitative search (reuse of existing knowledge) and explorative search (use of new knowledge). They find a curvilinear relation between exploitative search and new product introduction and a positive linear relation for explorative search. This suggests both processes can co-exist and have different effects, but it is unclear whether the found interaction effect truly suggests a non-zero dot product[7] (i.e., orthogonality, the two variables are not perpendicular in Euclidean space). Others suggest that both exploration and exploitation tend to exist intertemporally within organisations (Siggelkow & Levinthal, 2003), and that doing so can facilitate breakthrough success (Kinga Randle & Pisano, 2021), which resonates with the punctuated equilibrium perspective. At the inventor-firm level, Tzabbar and Kehoe (2014), for instance, show that firms tend to move from exploitation towards more exploration when a star scientist leaves the firm and that this effect can be moderated by the star scientist's collaborative involvement. Research on ambidexterity thus remains an important area for more research (Ehls et al., 2020).

We focus on search (scope) across all knowledge domains and search specialisation inside the agricultural domain. As such, we are 'analyzing exploration and exploitation in multiple, loosely connected domains' without assuming these variables to be on a one-dimensional continuum (Gupta et al., 2006, p. 698). We find a significant negative correlation between both variables (−0.46) suggesting that it is rather unlikely for a firm to be exploitative (specialised) in one domain and explorative (scope) in another. Yet 'unlikely' does not mean impossible. While solutions like internal structural and intertemporal ambidexterity implicitly presume continuity between exploitation and exploration, we recognise imperfect continuity and propose firms should jointly consider their search capabilities, including search and recombination, and the dependence between their own portfolio and the evolving knowledge landscape in order to drive value-creating performance. Near sparse clusters, exploration strategies with broad scope will prevail while near contested clusters, exploitation strategies of specialisation have the upper hand.

Topic modelling, dynamism, and knowledge landscapes

A core contribution of this study is to deploy novel methods and build novel theory to capture real dynamism in the environment. When discussing dynamic environments, Pfeffer and Salancik (2003, p. 69) noted that 'changes can come from anywhere without notice and produce consequences unanticipated by those initiating the changes and those experiencing the consequences'. While theoretical models have had a dynamic underpinning, empirical operationalisations have remained largely static although Kaplan and Vakili (2015) have begun to bring in some dynamism. The use of topic modelling in combination with semantic network analysis allows us to gain granular insight into how each invention reshapes the topography of the landscape, and how subsequent inventions create dynamic clusters.

The retrospectively dynamic perspective on the knowledge landscape reveals that emergent clusters are littered with opportunities: every search dimension positively correlates with performance in these clusters while high commitment is the most powerful. This gives the advantage in these areas to incumbent firms that have more slack resources. This provides an interesting avenue for future research as one would typically assume that start-ups and small nimble players are more likely to succeed in emerging innovation areas. Whether this finding is reflective of our specific context (agriculture) or whether this represents a broader trend requires closer attention (Murray, 2017). The theoretical construct of retrospective dynamism is, from our perspective, valuable. The notion that it sometimes takes a new insight to finally understand two old knowledge components, or that it takes finding a missing puzzle piece to eventually put two and two together, is familiar to everyone who ever studied complex problems. So far, we had not been able to study this at scale because we lacked the data and methods to do so. More research will be necessary to understand the implications of such emergence in technological landscapes.

Yet our current findings do not provide simple advice to managers because the relationships between search dimensions and the firm's patent portfolio position are complex. Distance from emergent clusters is generally better for those with low scope, low specialisation, or low commitment, suggesting that innovating in emergent clusters requires a deliberate strategy. However, our findings also show that whatever that strategy is, as long as the firm commits to either a broad search scope or to strong specialisation, success is possible. This suggests that when faced with emergence, firms have to choose between exploitative and explorative strategies to succeed, making both orthogonal to success.

LDA and semantic modelling

Recent editorials suggested that data science techniques could provide significant opportunities to refine and develop new theories on what may be considered mature domains (e.g., George, Osingal, Lavie & Scott, 2016). Our study seizes this opportunity by making innovative use of topic modelling to create semantic technological landscapes. To our knowledge, Kaplan and Vakili (2015) were the first to apply topic modelling to identify patents that originate a novel topic. They combined LDA with expert revision of 100 generated topics and looked for the first patents (within a 12 month period) that score above a specific threshold weighting. This enabled them to identify 'topic-originating' patents using semantic analysis and establish a significant relationship with the number of forward citation. While the authors use a static environment of 2,276 fullerene and nanotube patents granted by the USPTO between 1991 and 2005, we rely on a much larger sample of almost 62,000 patents and evolving environments. Their reliance on a single textual environment does not permit the types of dynamism we rely on for our moderating variables. While they trace back the origins of innovation using a full information environment, we seek out areas of temporary density increase and reduction using a partial information environment which better captures the actual uncertainty with regard to strategic choices companies must make when it comes to positioning their efforts in the innovation landscape as it unfolds over time.

The Kaplan and Vakili (2015) study generates convincing insight into the predictors of value (in terms of forward citations) of cognitively novel patents. However, their methodology does not allow actionable inference that managers or R&D departments can use in order to boost their innovation success. We build on their methodological sophistication to provide actionable insights. Specifically, by relying only on past, publicly available information to construe our semantic landscapes and exploring the impact of strategic positioning within those landscapes, given core search capabilities of different firms, our results provide cautious advice for R&D managers. We realise that simply knowing which areas may be the most promising to invest resources in, is by itself not a guarantee to successful innovation. But, managers could use our method to figure out which dense or emerging clusters are most closely linked to their extant knowledge base and hence provide plausibly fruitful ground for further inventive efforts, and identify sparse clusters to steer clear from, unless the firm has wide scope to facilitate the required knowledge brokerage.

Future research could apply similar methodologies to diverse questions in organisational discourse such as employee engagement, CEO narratives for strategic change, and studying customer or stakeholder relationships. For instance, one could analyse stock responses to corporate announcements and annual reports by using a similar unsupervised learning method as we did to discover topics in those reports. Researchers could determine the Hellinger distance between topics in the annual reports and corporate announcements and figure out whether stock markets respond differently to announcements that align closely with topics from annual reports or not. Our study provides an excellent test case for more complex analyses using data science techniques in semantic discourses.

Conclusion

Our study provides insight into how the constant gales of innovation not only change the individual firm's portfolio and its interconnections but also the technological space retrospectively. By strategically positioning the firm's innovation portfolio near specific clusters of inventive activity, the firm is able to participate, and shape, promising innovation areas. We find that the performance implications of such positioning are contingent upon the search strategy the firm adopts. Search specialists can maximise their performance by strategically positioning their innovation portfolio near carefully selected contested clusters, while search generalists can exploit their broad knowledge base by establishing profitable connections near sparse clusters. When the interconnections between inventions are less well known but are emerging rapidly, commitment outperforms scope as well as specialisation strategies of search. The study calls for future research that re-examines innovation in truly dynamic technological landscapes with new data analytic tools developed here. Even in traditional areas such as agriculture in which innovation has not been the most important strategic differentiator, we find that innovation-oriented search capabilities and positioning strategies of patent portfolios have a meaningful economic effect.

Notes

1. Abstracts contain a synoptic overview of the patented invention that is meant to capture the invention's most fundamental contributions. We use abstracts rather than full patent documents for computational reasons.
2. Note that our conceptualisation of dynamism in the environment differs from Yayavaram and Chen's (2014) 'change in couplings'. They are comparing how the connections between technological components evolve in between two mutually exclusive periods by relying on the 2005 USPTO classification system. Some couplings become more common while others less so. For us, the couplings themselves are subject to change, i.e., some couplings between two patents will disappear while new knowledge will emerge that allows us to add previously unknown couplings. This is akin to a quarterly revision of the UPSTO classification scheme itself.
3. In 2013 for instance, the USPTO agreed to switch from its final version of the US Patent Classification system to the Cooperative Patent Classification in conjunction with its European counterpart (EPO). As from January 2015, the USPTO completely stopped using the old structure (http://www.tprinternational.com/making-the-switch-from-uspc-to-cpc/)
4. 't' is defined as a year period consisting of four consecutive quarters.
5. We tried q = 1, 2, 3, 5, 10. Results in the paper rely on q = 3 reference locations.
6. This approach enables us to capture how the knowledge structure evolves as new knowledge. By comparing a patent to its 100 nearest neighbours in year t-3 and then to its 100 nearest neighbours in years [t-3, t-1] it becomes possible to see which patents have become 'surrounded' by other patents, implying shifts in the knowledge structure.
7. The authors do not include the product of the linear exploration and the quadratic term for exploitation which makes it difficult to assess the true interaction effect.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Simon J. D. Schillebeeckx http://orcid.org/0000-0003-0834-1208
Gerard George http://orcid.org/0000-0002-6963-5085

References

Alexy, O., George, G., & Salter, A. (2013). Cui Bono? The selective revealing of knowledge and its implications for innovative activity. *Academy of Management Review*, *38*(2), 270–291. https://doi.org/10.5465/amr.2011.0193

Allison, P. (2012). When can you safely ignore multicollinearity. *Statistical Horizons*, Accessed 01 02 2022. https://statisticalhorizons.com/multicollinearity/
Alnuaimi, T., & George, G. (2016). Appropriability and the retrieval of knowledge after spillovers. *Strategic Management Journal*, *37*(7), 1263–1279. https://doi.org/10.1002/smj.2383
Arrow, K. J. (1962). Economic welfare and the allocation of resources for invention. In Universities-National Bureau (Ed.), *The rate and direction of inventive activity: Economic and social factors* (pp. 609–626). Princeton University Press.
Audretsch, D. B., & Feldman, M. P. (1996). R&D spillovers and the geography of innovation and production. *The American Economic Review*, *86*(3), 630–640. http://www.jstor.org/stable/2118216
Bessen, J. (2009). Estimates of patent rents from firm market value. *Research Policy*, *38*(10), 1604–1616. https://doi.org/10.1016/j.respol.2009.09.014
Blei, D. M., & Lafferty, J. D. (2009). Topic models. In A. N. Srivastava & M. Sahami (Eds.), *Text mining: Classification, clustering, and applications* (Vol. 10, pp. 71–94). CRC Press.
Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003, January). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022. https://dl.acm.org/doi/10.5555/944919.944937
Bloom, N., & Van Reenen, J. (2002). Patents, real options and firm performance. *The Economic Journal*, *112*(478), C97–C116. https://doi.org/10.1111/1468-0297.00022
Burns, T., & Stalker, G. (1961). *The Management of Innovation*. London: Tavistock.
Caner, T., Cohen, S. K., & Pil, F. (2017). Firm heterogeneity in complex problem solving: A knowledge-based look at invention. *Strategic Management Journal*, *38*(9), 1791–1811. https://doi.org/10.1002/smj.2615
Capon, N., Farley, J. U., & Hoenig, S. (1990). Determinants of financial performance: A meta-analysis. *Management Science*, *36*(10), 1143–1159. https://doi.org/10.1287/mnsc.36.10.1143
Carnabuci, G., & Bruggeman, J. (2009). Knowledge specialization, knowledge brokerage and the uneven growth of technology domains. *Social Forces*, *88*(2), 607–641. https://doi.org/10.1353/sof.0.0257
Choi, J., Menon, A., & Tabakovic, H. (2021). Using machine learning to revisit the diversification–performance relationship. *Strategic Management Journal*, *42*(9), 1632–1661. https://doi.org/10.1002/smj.3317
Clauss, T., Kraus, S., Kallinger, F. L., Bican, P. M., Brem, A., & Kailer, N. (2021). Organizational ambidexterity and competitive advantage: The role of strategic agility in the exploration-exploitation paradox. *Journal of Innovation & Knowledge*, *6*(4), 203–213. https://doi.org/10.1016/j.jik.2020.07.003
Cockburn, I. M., & Griliches, Z. (1987). *Industry effects and appropriability measures in the stock markets valuation of R&D and patents*. National Bureau of Economic Research Cambridge.
Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, *35*(1), 128–152. https://doi.org/10.2307/2393553
Cohen, W. M., Nelson, R. R., & Walsh, J. P. (2000). Protecting their intellectual assets: Appropriability conditions and why US manufacturing firms patent (or not): 50. National Bureau of Economic Research.
Courtney, H., Kirkland, J., & Viguerie, P. (1997). Strategy under uncertainty. *Harvard Business Review*, *75*(6), 67–79. https://hbr.org/1997/11/strategy-under-uncertainty
Cowan, R., David, P. A., & Foray, D. (2000). The explicit economics of knowledge codification and tacitness. *Industrial and Corporate Change*, *9*(2), 211–253. https://doi.org/10.1093/icc/9.2.211
Cyert, R. M., & March, J. G. (1963). *A behavioral theory of the firm*. Prentice-Hall.
Deng, Z., Lev, B., & Narin, F. (1999). Science and technology as predictors of stock performance. *Financial Analysts Journal*, *55*(3), 20–32. https://doi.org/10.2469/faj.v55.n3.2269
Dess, G. G., & Beard, D. W. (1984). Dimensions of organizational task environments. *Administrative Science Quarterly*, *29*(1), 52–73. https://doi.org/10.2307/2393080

Dosi, G. (1982). Technological paradigms and technological trajectories: A suggested interpretation of the determinants and directions of technical change. *Research Policy*, *11*(3), 147–162. https://doi.org/10.1016/0048-7333(82)90016-6

Ehls, D., Polier, S., & Herstatt, C. (2020). Reviewing the field of external knowledge search for innovation: Theoretical underpinnings and future (re-) search directions. *Journal of Product Innovation Management*, *37*(5), 405–430. https://doi.org/10.1111/jpim.12549

Felin, T., & Zenger, T. R. (2014). Closed or open innovation? Problem solving and the governance choice. *Research Policy*, *43*(5), 914–925. https://doi.org/10.1016/j.respol.2013.09.006

Fleming, L. (2001). Recombinant uncertainty in technological search. *Management Science*, *47*(1), 117–132. https://doi.org/10.1287/mnsc.47.1.117.10671

Fleming, L., & Sorenson, O. (2001). Technology as a complex adaptive system: Evidence from patent data. *Research Policy*, *30*(7), 1019–1039. https://doi.org/10.1016/S0048-7333(00)00135-9

Fleming, L., & Sorenson, O. (2004). Science as a map in technological search. *Strategic Management Journal*, *25*(8-9), 909–928. https://doi.org/10.1002/smj.384

Garg, V. K., Walters, B. A., & Priem, R. L. (2003). Chief executive scanning emphases, environmental dynamism, and manufacturing firm performance. *Strategic Management Journal*, *24*(8), 725–744. https://doi.org/10.1002/smj.335

George, G., Howard-Grenville, J., Joshi, A., & Tihanyi, L. (2016). Understanding and tackling societal grand challenges through management research. *Academy of Management Journal*, *59* (6), 1880–1895. https://doi.org/10.5465/amj.2016.4007

George, G., Kotha, R., & Zheng, Y. (2008). Entry into insular domains: A longitudinal study of knowledge structuration and innovation in biotechnology firms. *Journal of Management Studies*, *45*(8), 1448–1474. https://doi.org/10.1111/j.1467-6486.2008.00805.x

George, G., Merrill, R., & Schillebeeckx, S. J. D. (2021). Digital sustainability and entrepreneurship: How digital innovations are helping tackle climate change and sustainable development. *Entrepreneurship Theory & Practice*, *45*(5), 999–1027. https://doi.org/10.1177/1042258719899425

George, G., Osinga, E., Lavie, D., & Scott, B. (2016). From the editors: Big data and data science methods for management research. *Academy of Management Journal*, *59*(5), 1–15. https://doi.org/10.5465/amj.2016.4001

George, G., & Schillebeeckx, S. J. D. (2022). Digital transformation, sustainability, and purpose in the multinational enterprise. *Journal of World Business*, *57*(2), 101236. https://doi.org/10.1016/j.jwb.2022.101326

Giuliani, E. (2007). The selective nature of knowledge networks in clusters: Evidence from the wine industry. *Journal of Economic Geography*, *7*(2), 139–168. https://doi.org/10.1093/jeg/lbl014

Grant, R. M. (1996). Toward a knowledge-based theory of the firm. *Strategic Management Journal*, *17*(S2), 109–122. https://doi.org/10.1002/smj.4250171110

Griliches, Z. (1998). Patent statistics as economic indicators: A survey. In Z. Griliches (Ed.), *R&D and productivity: The econometric evidence* (pp. 287–343). University of Chicago Press.

Gullickson, G. (2017). What Bayer is saying about the Bayer-Monsanto Merger. Accessed 01 02 2022. https://www.agriculture.com/news/crops/what-bayer-is-saying-about-the-bayer-monsanto-merger.

Gupta, A. K., Smith, K. G., & Shalley, C. E. (2006). The interplay between exploration and exploitation. *Academy of Management Journal*, *49*(4), 693–706. https://doi.org/10.5465/amj.2006.22083026

Hackett, E. J., Leahey, E., Parker, J. N., Rafols, I., Hampton, S. E., Corte, U., Chavarro, D., Drake, J. M., Penders, B., Sheble, L., Vermeulen, N., & Vision, T. J. (2021). Do synthesis centers synthesize? A semantic analysis of topical diversity in research. *Research Policy*, *50*(1), 104069. https://doi.org/10.1016/j.respol.2020.104069

Hall, B. H. (2007). *Measuring the returns to R&D: The depreciation problem: 1-54*. National Bureau of Economic Research.

Hall, B. H., & Harhoff, D. (2012). Recent research on the economics of patents. *Annual Review of Economics*, *4*(1), 541–565. https://doi.org/10.1146/annurev-economics-080511-111008

Hall, B. H., Jaffe, A. B., & Trajtenberg, M. (2001). *The NBER patent citation data file: Lessons, insights and methodological tools*. National Bureau of Economic Research.

Hall, B. H., Jaffe, A. B., & Trajtenberg, M. (2005). Market value and patent citations. *RAND Journal of Economics*, *36*(1), 16–38 https://www.jstor.org/stable/1593752.

Hannan, M. T., & Freeman, J. (1977). The population ecology of organizations. *American Journal of Sociology*, *82*(5), 929–964. https://doi.org/10.1086/226424

Hannigan, T., Haans, R. F. J., Vakili, K., Tchalian, H., Glaser, V., Wang, M., Kaplan, S., & Jennings, P. D. (2019). *Topic modeling in management research: Rendering new theory from textual data*. Academy of Management Annals(ja).

Hegde, D., & Sampat, B. (2009). Examiner citations, applicant citations, and the private value of patents. *Economics Letters*, *105*(3), 287–289. https://doi.org/10.1016/j.econlet.2009.08.019

Huber, G. P. (1991). Organizational learning: The contributing processes and the literatures. *Organization Science*, *2*(1), 88–115. https://doi.org/10.1287/orsc.2.1.88

Iammarino, S., & McCann, P. (2006). The structure and evolution of industrial clusters: Transactions, technology and knowledge spillovers. *Research Policy*, *35*(7), 1018–1036. https://doi.org/10.1016/j.respol.2006.05.004

James, S. D., Leiblein, M. J., & Lu, S. (2013). How firms capture value from their innovations. *Journal of Management*, *39*(5), 1123–1155. https://doi.org/10.1177/0149206313488211

Jung, H. J., & Lee, J. (2016). The quest for originality: A new typology of knowledge search and breakthrough inventions. *Academy of Management Journal*, *59*(5), 1725–1753 https://doi.org/10.5465/amj.2014.0756.

Kaplan, S., & Vakili, K. (2015). The double-edged sword of recombination in breakthrough innovation. *Strategic Management Journal*, *36*(10), 1435–1457. https://doi.org/10.1002/smj.2294

Katila, R. (2002). New product search over time: Past ideas in their prime? *Academy of Management Journal*, *45*(5), 995–1010. https://doi.org/10.2307/3069326

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. https://doi.org/10.5465/3069433.

King, A., & Tucci, C. L. (2002). Incumbent entry into new market niches: The role of experience and managerial choice in the creation of dynamic capabilities. Management Science, *48*(2), 171–186. doi: https://doi.org/10.1287/mnsc.48.2.171.253

Kinga Randle, D., & Pisano, G. P. (2021). The evolutionary nature of breakthrough innovation: An empirical investigation of firm search strategies. *Strategy Science*, *6*(4), 290–304. https://doi.org/10.1287/stsc.2021.0134

Kneeland, M. K., Schilling, M. A., & Aharonson, B. S. (2020). Exploring uncharted territory: Knowledge search processes in the origination of outlier innovation. *Organization Science*, *31*(3), 535–557. https://doi.org/10.1287/orsc.2019.1328

Knudsen, T., & Srikanth, K. (2014). Coordinated exploration organizing joint search by multiple specialists to overcome mutual confusion and joint myopia. *Administrative Science Quarterly*, *59*(3), 409–441. https://doi.org/10.1177/0001839214538021

Kogan, L., Papanikolaou, D., Seru, A., & Stoffman, N. (2012). Technological innovation, resource allocation, and growth, *Working Paper Series*, January ed.: 1–62. Cambridge, MA, USA: National Bureau of Economic Research.

Kogut, B., & Zander, U. (1992). Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science*, *3*(3), 383–397. https://doi.org/10.1287/orsc.3.3.383

Kondrak, G. (2005). N-gram similarity and distance. *Paper presented at the String processing and information retrieval*, Berlin/Heidelberg, Springer Link, 115–126. doi:10.1007/11575832_13.

Kotha, R., Zheng, Y., & George, G. (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. https://doi.org/10.1002/smj.915

Laureiro-Martínez, D., Brusoni, S., Canessa, N., & Zollo, M. (2015). Understanding the exploration–exploitation dilemma: An fMRI study of attention control and decision-making performance. *Strategic Management Journal*, *36*(3), 319–338. https://doi.org/10.1002/smj.2221

Laursen, K. (2012). Keep searching and you'll find: What do we know about variety creation through firms' search activities for innovation? *Industrial and Corporate Change*, *21*(5), 1181–1220. https://doi.org/10.1093/icc/dts025

Laursen, K., & Salter, A. (2006). Open for innovation: The role of openness in explaining innovation performance among UK manufacturing firms. *Strategic Management Journal*, *27* (2), 131–150. https://doi.org/10.1002/smj.507

Leppänen, P., George, G., & Alexy, O. (2021). When do novel business models lead to high performance? A configurational approach to value drivers, competitive strategy, and firm environment. *Academy of Management Journal*. https://doi.org/10.5465/amj.2020.0969

Levinthal, D. A. (1997). Adaptation on rugged landscapes. *Management Science*, *43*(7), 934–950. https://doi.org/10.1287/mnsc.43.7.934

Levinthal, D. A., & March, J. G. (1981). A model of adaptive organizational search. *Journal of Economic Behavior & Organization*, *2*(4), 307–333. https://doi.org/10.1016/0167-2681(81)90012-3

Lin, B.-W., Lee, Y., & Hung, S.-C. (2006). R&D intensity and commercialization orientation effects on financial performance. *Journal of Business Research*, *59*(6), 679–685. https://doi.org/10.1016/j.jbusres.2006.01.002

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, *2*(1), 71–87. https://doi.org/10.1287/orsc.2.1.71

McArthur, A. W., & Nystrom, P. C. (1991). Environmental dynamism, complexity, and munificence as moderators of strategy-performance relationships. *Journal of Business Research*, *23*(4), 349–361. https://doi.org/10.1016/0148-2963(91)90020-X

Miller, D. J. (2006). Technological diversity, related diversification, and firm performance. *Strategic Management Journal*, *27*(7), 601–619. https://doi.org/10.1002/smj.533

Mohammadi, T. (2012). Relative performance of components variance estimators in random effects models. *Iranian Journal of Economic Research*, *17*(50), 83–98. https://ijer.atu.ac.ir/article_2879.html?lang=en

Mowery, D. C., Nelson, R. R., Sampat, B. N., & Ziedonis, A. A. (2001). The growth of patenting and licensing by US universities: An assessment of the effects of the Bayh–Dole act of 1980. *Research Policy*, *30*(1), 99–119. https://doi.org/10.1016/S0048-7333(99)00100-6

Murray, A. (2017, November 13). Get ready for the new disrupters. *Time Magazine*, Warner Media.14.

Nag, R., & Gioia, D. A. (2012). From common to uncommon knowledge: Foundations of firm-specific use of knowledge as a resource. *Academy of Management Journal*, *55*(2), 421–457. https://doi.org/10.5465/amj.2008.0352

Nan, D., Liu, F., & Ma, R. (2018). Effect of proximity on recombination innovation in R&D collaboration: An empirical analysis. *Technology Analysis & Strategic Management*, *30*(8), 921–934. https://doi.org/10.1080/09537325.2018.1424327

Nerkar, A. (2003). Old is gold? The value of temporal exploration in the creation of new knowledge. *Management Science*, *49*(2), 211–229. https://doi.org/10.1287/mnsc.49.2.211.12747

Nickerson, J. A., & Zenger, T. R. (2004). A knowledge-based theory of the firm—The problem-solving perspective. *Organization Science*, *15*(6), 617–632. https://doi.org/10.1287/orsc.1040.0093

Nikita, M. (2016). Package ldatuning: 1-4. online: CRAN R-Project. Accessed 01 02 2022. https://cran.r-project.org/web/packages/ldatuning/ldatuning.pdf

Nikulin, M. (2001). Hellinger distance. In M. Hazewinkel (Ed.), *Encyclopedia of mathematics* (Vol. 10). Springer, 78.

Paruchuri, S., & Awate, S. (2017). Organizational knowledge networks and local search: The role of intra-organizational inventor networks. *Strategic Management Journal*, *38*(3), 657–675. https://doi.org/10.1002/smj.2516

Pfeffer, J., & Salancik, G. R. (2003). *The external control of organizations, a resource dependence perspective*. Harper & Row Publishers.

Pich, M. T., Loch, C. H., & De Meyer, A. (2002). On uncertainty, ambiguity, and complexity in project management. *Management Science*, *48*(8), 1008–1023. https://doi.org/10.1287/mnsc.48.8.1008.163

Pigford, -A.-A. E., Hickey, G. M., & Klerkx, L. (2018). Beyond agricultural innovation systems? Exploring an agricultural innovation ecosystems approach for niche design and development in sustainability transitions. *Agricultural Systems*, *164*(July), 116–121. https://doi.org/10.1016/j.agsy.2018.04.007
Priem, R. L., Rasheed, A. M., & Kotulic, A. G. (1995). Rationality in strategic decision processes, environmental dynamism and firm performance. *Journal of Management*, *21*(5), 913–929. https://doi.org/10.1177/014920639502100506
Purcell, R., & McGrath, F. (2013). The search for external knowledge. *Electronic Journal of Knowledge Management*, *11*(2), 158–167. https://academic-publishing.org/index.php/ejkm/article/view/987
Sandner, P. G., & Block, J. (2011). The market value of R&D, patents, and trademarks. *Research Policy*, *40*(7), 969–985. https://doi.org/10.1016/j.respol.2011.04.004
Schillebeeckx, S., Lin, Y., George, G., & Alnuaimi, T. (2021). Knowledge recombination and inventor networks: The asymmetric effects of embeddedness on knowledge reuse and impact. *Journal of Management*, *47*(4), 838–866. https://doi.org/10.1177/0149206320906865
Schumpeter, J. A. (1939). *Business Cycles*. Cambridge University Press.
Schumpeter, J. A. (1942). *Capitalism, socialism and democracy*. Routledge.
Siggelkow, N., & Levinthal, D. A. (2003). Temporarily divide to conquer: Centralized, decentralized, and reintegrated organizational approaches to exploration and adaptation. *Organization Science*, *14*(6), 650–669. https://doi.org/10.1287/orsc.14.6.650.24840
Simon, H. A. (1978). Rationality as process and as product of thought. *The American Economic Review*, *68*(2), 1–16. https://www.jstor.org/stable/1816653.
Simon, H. A. (1982). *Models of bounded rationality: Empirically grounded economic reason*. MIT press.
Swaminathan, A. (2001). Resource partitioning and the evolution of specialist organizations: The role of location and identity in the U.S. wine industry. *Academy of Management Journal*, *44*(6), 1169–1185. https://www.jstor.org/stable/3069395
Trajtenberg, M. (1990). A penny for your quotes: Patent citations and the value of innovations. *The Rand Journal of Economics*, *21*(1), 172–187. https://doi.org/10.2307/2555502
Tzabbar, D., & Kehoe, R. R. (2014). Can opportunity emerge from disarray? An examination of exploration and exploitation following star scientist turnover. *Journal of Management*, *40*(2), 449–482. https://journals.sagepub.com/doi/abs/10.1177/0149206313513613
Von Hippel, E. (1988). *The Sources of Innovation*. Oxford University Press.
Walsh, J. P., & Ungson, G. R. (1991). Organizational memory. *Academy of Management Review*, *16*(1), 57–91. https://doi.org/10.2307/258607
Xiao, T., Makhija, M., & Karim, S. (2021). A knowledge recombination perspective of innovation: Review and new research directions. *Journal of Management*, 014920632110559. https://doi.org/10.1177/01492063211055982
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. https://doi.org/10.2189/asqu.53.2.333
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*(3), 377–396. https://doi.org/10.1002/smj.2218
Zahra, S. A., & George, G. (2002). Absorptive capacity: A review, reconceptualization, and extension. *Academy of Management Review*, *27*(2), 185–203. https://doi.org/10.2307/4134351
Zajac, E. J., & Bazerman, M. H. (1991). Blind spots in industry and competitor analysis: Implications of interfirm (mis) perceptions for strategic decisions. *Academy of Management Review*, *16*(1), 37–56. https://doi.org/10.5465/amr.1991.4278990
Zhang, J., Yan, Y., & Guan, J. (2019). Recombinant distance, network governance and recombinant innovation. *Technological Forecasting and Social Change*, *143*(June), 260–272. https://doi.org/10.1016/j.techfore.2019.01.022
Zou, T., Ertug, G., & George, G. (2018). The capacity to innovate: A meta-analysis of absorptive capacity. *Innovation: Organization & Management*, *20*(2), 87–121. https://doi.org/10.1080/14479338.2018.1428105