



DO INNOVATIVE USERS GENERATE MORE USEFUL INSIGHTS? AN ANALYSIS OF CORPORATE VENTURE CAPITAL INVESTMENTS IN THE MEDICAL DEVICE INDUSTRY

SHERYL WINSTON SMITH^{1*} and SONALI K. SHAH^{2*}

¹*Fox School of Business, Temple University, Philadelphia, Pennsylvania, U.S.A.*

²*Foster School of Business, University of Washington, Seattle, Washington, U.S.A.*

Users are an important source of innovation. Scholars have suggested that established firms will gain valuable innovative insights by working with user innovators. However, no study compares the extent to which knowledge sourced from innovative users, as compared to other external sources of knowledge, triggers the creation of new technologies and commercial products within established firms. This leaves established firms with little guidance when it comes to choosing where to search for external knowledge that ignites innovation. Based on existing empirical work in the literature on user innovation, we build a theoretical framework that explains why user knowledge will provide established firms with more 'useful' innovative insights than will other sources of knowledge. We test this claim in the context of corporate venture capital investment in the medical device industry. We find that established firms incorporate more knowledge from user innovators than from other sources of external knowledge into their patents and highly innovative products. Accessing the knowledge contained in user-generated innovations enriches the product development outcomes of established firms. We trace the flow of knowledge from start-ups to established firms using both an established method based on backward patent citation data and a novel algorithmic method that compares the content of regulatory documents. Copyright © 2013 John Wiley & Sons, Ltd.

INTRODUCTION

Innovation is essential to the growth and survival of technology-based firms, but innovation is also a complex, challenging, and knowledge-intensive activity. The process by which firms develop innova-

tive new products involves novel insights and considerable learning about technologies and markets (Clark and Fujimoto, 1991; Brown and Eisenhardt, 1995; Taylor, 2010). Research and development activities seed new technological capabilities and innovative products through the combination of knowledge that is known to a firm with knowledge that is new to it (Nelson and Winter, 1982; March, 1991; Leiponen and Helfat, 2010). New knowledge can be generated internally or sourced from the external environment. From a practical perspective, however, many established firms find it difficult to generate radical innovations solely through internal processes (Tushman and Anderson, 1986; Henderson, 1993). Therefore, many established firms look to acquire and exploit knowledge developed

Keywords: innovation; user innovation; corporate venture capital; technology entrepreneurship; new product development

*Correspondence to: Sheryl Winston Smith, Fox School of Business, Temple University, 1801 Liacouras Walk, Philadelphia, PA 19122, U.S.A. E-mail: sheryl.winston.smith@temple.edu; Sonali K. Shah, Foster School of Business, University of Washington, Box 353200, Seattle, WA 98195-3200, U.S.A. E-mail: skshah@u.washington.edu. Both authors are equal contributors. Author names appear in reverse alphabetical order.

externally (Arora and Gambardella, 1990; Cohen and Levinthal, 1990; Helfat, 1997; Chesbrough, 2003).

As firms search the external landscape, they seek knowledge that contains valuable innovation-related insights. User innovators are one possible source of such insights. The term user innovator describes innovators who expect to directly benefit from developing a product or service by *using* it (von Hippel, 1988; Kline and Pinch, 1996). An established literature shows that users are the source of many important innovations in diverse industries, ranging from automobiles to scientific equipment to library software systems (Oudshoorn and Pinch, 2003; von Hippel, 2005). The user innovation process is unique in that users identify a variety of unmet needs, immerse themselves in the context within which an innovation is used, and garner resources from other users through participation in user innovation communities (von Hippel, 1988, 1994; Franke and Shah, 2003). These differences may lead to the creation of knowledge that is distinct from that developed by nonusers and more valuable to established firms.

While the user innovation literature has documented the prevalence of user innovation and its importance to technological progress, it has focused less on the implications of user innovation for established firms. The implicit assumption in the literature to date has been that established firms will benefit from incorporating user knowledge into their R&D processes. Hence, the focus of the literature has been on suggesting methods by which established firms can work with innovative users, such as the lead user method (Urban and von Hippel, 1988; Lilien *et al.*, 2002), toolkits for user innovation (von Hippel and Katz, 2002), and working with user innovation communities (Jeppesen and Molin, 2003; Jeppesen and Frederiksen, 2006; Shah, 2006). However, a critical issue remains unexplored: does acquiring knowledge from innovative users do more to benefit the invention and innovation outcomes of established firms than acquiring knowledge from other sources? Put another way, given that established firms can source knowledge from a variety of external sources, will knowledge sourced from innovative users be more 'useful' in developing new inventions and highly innovative commercial products? This article seeks to address this issue.

We investigate this question by measuring the extent to which established firms incorporate knowledge gained through relationships with innovative users, as compared to other external sources of knowledge, into their own inventions and innova-

tions. Specifically, we compare the knowledge that flows from start-ups rooted in different knowledge sources to established firms through the process of corporate venture capital (CVC) investment. Corporate venture capital investments are equity investments by established firms into private entrepreneurial ventures (Gompers and Lerner, 2000). CVC investing is believed to offer corporations a way to outsource, supplement, and/or accelerate their internal R&D processes by providing access to novel knowledge that fuels innovation (Dushnitsky and Lenox, 2005; Wadhwa, Phelps, and Kotha, 2010). Established firms engage in CVC investing to gain access to the technological knowledge start-ups possess (Siegel, Siegel, and MacMillan, 1988; Katila, Rosenberger, and Eisenhardt, 2008; Dushnitsky and Shaver, 2009).

We choose to conduct our study in the context of CVC investing because it provides a relatively controlled situation within which to compare the effects of knowledge from different sources on the technological outcomes of established firms. We use a founder's background in the industry to indicate the source of external knowledge being exploited by the start-up (Boeker, 1988; Eisenhardt and Schoonhoven, 1990; Beckman and Burton, 2008). Within our sample, founders have backgrounds as users of medical devices (i.e., practicing physicians), academic scientists, and individuals previously employed by established medical device firms. While the founders in our sample draw upon distinct knowledge sources, they also share common goals in that they are looking to commercialize their idea(s), have actively protected their intellectual property, and are backed by independent and corporate venture capital investors seeking a return on their investments. Within this context, differences in the integration of external knowledge from different sources into patents and products are likely to reflect inherent differences in the knowledge itself.

We construct a novel dataset of CVC relationships in the medical device industry ($n = 128$ start-up-investor dyads). The medical device industry is a particularly active arena of CVC activity (Ernst and Young, 2010); established medical device firms actively search for external knowledge to combat short product life cycles and intense competition, while start-ups seek resources to develop their innovative ideas. We examine two key indicators of an incumbent's innovative output—patents and innovative new products. This allows us to document the effects of knowledge sourced from user innovators

on both an important intermediate inventive outcome (patents) and the desired commercial outcome (innovative new products) of the corporate R&D process.

We find that incumbents engaged in CVC investment more frequently incorporate knowledge from innovative users into their subsequent inventions and innovations than they do from other external knowledge sources. Our contribution lies in showing the valuable product development benefits that established firms can derive from accessing knowledge from innovative users. In doing so, we establish the importance of user innovation on an outcome—the technological performance of established firms—that strategy scholars care deeply about. From a practical perspective, we provide insights that established firms can use to guide their search for external knowledge by illuminating the differential value of knowledge from distinct sources. In addition, this article makes a methodological contribution by introducing an algorithmic method for tracing knowledge flows between written documents to the strategy literature.

THEORY AND HYPOTHESES DEVELOPMENT: UNIQUE CHARACTERISTICS OF THE USER INNOVATION PROCESS

Users are distinguished from other sources of innovation by the primary motive that fuels their innovation activity: users innovate because they expect to derive benefit by *using* the innovations they create (von Hippel, 1988; Kline and Pinch, 1996). In contrast, many firms innovate because they expect to derive pecuniary benefits by selling the innovation to others (Schumpeter, 1942; von Hippel, 1988; Chandler, 1994). And, many academics innovate because they expect to derive status and reputation-enhancing benefits by publishing and promoting their results (e.g., Merton, 1979). These differences in motives produce differences in the process by which users, as compared to academic scientists or employees of firms, develop their innovations. We build theory to explain how the unique characteristics of the user innovation process—identifying a variety of unrecognized needs, immersion in the problem context, and community-based problem solving—allow users to expose qualitatively different insights than those exposed by other sources of innovation. These insights can be used as an input to the corporate innovation process. Later we will

describe each of these processes in detail and explain how established firms who access this knowledge will benefit from the unique insights this knowledge contains.

Users conceive of a variety of previously unrecognized needs

Users possess a deep understanding of the needs created by the absence of a product designed for a particular purpose, the absence of a product feature, and/or by the failures and shortcomings of existing products (von Hippel, 1988). They utilize this knowledge to identify the problems for which they will find innovative solutions. As a result, user innovations provide value by identifying and satisfying needs that are inadequately addressed by existing products. Nonusers may not be able to recognize these needs as quickly or at all.

Users develop innovations that established firms are unlikely to conceive on their own (von Hippel, 1988). Whereas innovations made by firms are more likely to improve product performance along established performance parameters (e.g., make a laptop lighter, faster, or more durable), user innovations are more likely to uncover altogether new features and/or product functionality (Riggs and von Hippel, 1994). One effect of this is that users developed early, field-defining innovations in a variety of industries (Franz, 2005; Shah, 2005; Mody, 2006). For example, users developed the first atomic force microscope as a tool to inspect thin-film superconducting materials (Mody, 2006). It would have been difficult for an existing scientific instruments manufacturer to realize that such a tool was needed by a small group of engineers seeking to build the first superconducting computer, hence these engineers had to create and build their own microscopes in order to use them. More generally, established firms find it difficult to identify the types of needs that users encounter and create on a routine basis.

User innovations often signal areas where consumers require a new element of functionality. By working with innovative users, established firms will gain valuable insights into previously unrecognized needs and their significance to consumers that they cannot gain from other sources.

Users immerse themselves in the problem context

Different users may use a product in different ways or in different environmental conditions. Because

individual users are situated in the environment in which an innovation will be used, they are often able to build a deep and accurate understanding of the environment. This is important because environmental factors often affect the functioning of an innovation in unexpected ways (Tyre and von Hippel, 1997). User innovators utilize this knowledge as they troubleshoot and correct problems with existing products and as they design and construct new innovations (Ogawa, 1998).

In contrast, nonusers' perceptions of the circumstances under which the product will be utilized can diverge from the realities of actual use. Such mismatches can lead products designed by nonusers to fail in practice. When a product fails, substantial problem-solving effort goes into uncovering and understanding the context in which a problem resides (Tyre and von Hippel, 1997). However, knowledge pertaining to the problem context can be costly to acquire, transfer, and use in a new location, making it difficult for scientists and engineers at established firms or academic institutions to correct the product's shortcomings (von Hippel, 1994; Ogawa, 1998). Understanding the nuances of the environments in which products are used would enable an established firm to design more effective products. By working with innovative users, established firms will gain a more nuanced understanding of the context(s) in which a product is utilized, and they will be able to incorporate this knowledge into their innovation process. Knowledge sourced from innovative users is, therefore, likely to provide valuable insights to established firms in the industry, enabling them to develop more robust products.

Users interact with innovation communities

Many users choose to work collectively in communities, sharing resources, knowledge, ideas, and innovative prototypes (Franke and Shah, 2003; von Hippel and von Krogh, 2003). These voluntary associations are composed of loosely affiliated users with common interests and provide a forum for relatively free and open information exchange. Individual users come from a wide variety of backgrounds and possess distinct and heterogeneous knowledge bases (Lüthje, Herstatt, and von Hippel, 2005).

By working within communities, users are able to tap a pool of heterogeneous knowledge and bring that knowledge to bear on a particular problem (Wenger, 1998; Brown and Duguid, 2001; Franke and Shah, 2003). Increasing the variety and diversity

of solutions considered leads to the creation of products that are more innovative (March, 1991). Bringing together individuals from *outside* the core discipline of a given field improves the chances of solving a problem and often results in the creation of highly innovative solutions because outsiders frame problems differently (Guimerà *et al.*, 2005; Tapscott, 2006).

Working within user communities also allows innovators to gauge interest in the innovation (Franke and Shah, 2003). As the innovator reaches out to others in the community, others may express interest in the innovation and choose whether to adopt the innovation for their own use, allowing the innovator to ascertain whether or not the innovation has broader appeal (Shah, 2005; Mody, 2006). Community participation, therefore, allows users to gauge demand for their innovation *prior* to starting a firm, reducing the innovation's commercial risk (Shah, 2005; Shah and Tripsas, 2007).

Interacting with a large number of individuals who possess distinct knowledge bases, are willing to provide feedback to improve the innovation, and indicate their personal interest in utilizing the innovation, allows users to generate effective—and sometimes even ingenious—solutions to their problems. By working with innovative users, established firms gain valuable insights that broaden their knowledge base in unexpected directions, expand the solution spaces they consider, and gauge the potential for consumer interest in the innovation needs. Knowledge sourced from innovative users is, therefore, likely to provide valuable insights to established medical device firms.

Three characteristics can differentiate the process of user innovation from the innovation processes of most firms and academic institutions: users (1) experience the shortcomings of existing products, (2) immerse themselves in the problem context, and (3) work collaboratively within innovation communities. These characteristics allow users to expose qualitatively different insights from those exposed by other sources of innovation. These insights can be used to inform an established firm's technological- and market-related decisions. Possessing user knowledge will allow established firms to create innovations that are more likely to address altogether novel needs, embody a deep and nuanced understanding of the environmental context in which the innovation must function, and integrate heterogeneous solution knowledge. Therefore, we expect that established firms will be more likely to use this

knowledge to fuel their R&D processes. We hypothesize that established firms will incorporate more knowledge from innovative users—as compared to nonusers—into their patents and innovative new products.

Hypothesis 1: Established firms will more frequently incorporate knowledge from users, relative to nonusers, into their new technologies.

Hypothesis 2: Established firms will more frequently incorporate knowledge from users, relative to nonusers, into their new products.

RESEARCH SETTING: CORPORATE VENTURE CAPITAL INVESTING IN THE MEDICAL DEVICE INDUSTRY

Several characteristics of CVC investing in the medical device industry make it a theoretically and methodologically attractive setting in which to examine the value of knowledge sourced from innovative users to established firms. First, individuals from a variety of backgrounds generate knowledge and found start-ups in the medical device industry and appear in our sample: users (physicians) generate knowledge as they seek to improve patient outcomes, academic scientists generate knowledge as part of their research endeavors, and former employees of established firms generate knowledge as they design, produce, and market commercial products. Because start-ups are highly influenced by their founders (Boeker, 1988; Eisenhardt and Schoonhoven, 1990; Delmar and Shane, 2006; Beckman and Burton, 2008), we can use a founder's background as an indicator of the source of external knowledge being exploited by the start-up.

Second, CVC investing is a well-established practice in the medical device industry (Pricewaterhouse Coopers, 2006). Established firms view CVC investing as a mechanism for gaining access to valuable technological insights (Siegel *et al.*, 1988; Katila *et al.*, 2008; Dushnitsky and Shaver, 2009). Four well-regarded established firms have established formal CVC programs through which they systematically identify, assess, and potentially invest in start-ups (Winston Smith, 2009). In preliminary field interviews, managers described how they identify promising start-ups in which to invest: they search widely for ideas, looking to identify the start-ups with the best ideas. They do not appear to differen-

tiate between start-ups based on the source of a founder's knowledge and, in fact, they sought to understand the value of each start-up's ideas individually. Their investment approach is illustrated by the following quote: 'the sieve is broad, but inside the sieve the competition is fierce' (pers. comm., 2006). Because these sophisticated established firms run systematic programs with the goal of identifying valuable knowledge and utilizing that knowledge in their internal product development process, we expect technological knowledge to flow from start-ups (i.e., the external knowledge source) to investors. This provides an ideal context to examine whether established firms are more likely to incorporate knowledge from innovative users than from other sources.

Third, formal intellectual property rights are utilized heavily in the industry to protect innovations (Levin, Cohen, and Mowery, 1985). The near universal use of patents by medical device firms and the cumulative nature of innovation in this industry provide us with a 'paper trail' through which to trace knowledge flows. As one manager explained: 'By its very nature, the device industry "stacks" patents on successful prior ones. . . . This is distinctly different from the pharmaceutical industry where the drug patent usually is the end product; [in contrast] there are literally thousands of patents that relate to a pacemaker.'

Fourth, regulatory requirements within the medical device industry also create a 'paper trail' through which we can trace the flow of knowledge from start-ups to CVC investors' highly novel products. A pre-market approval (PMA) must be submitted to the U.S. Food and Drug Administration for all novel devices (i.e., devices without a 'predicate') and 'supplemental' applications must be filed for all changes that will substantially impact the safety and effectiveness of the device (United States Food and Drug Administration, 2012). The PMA process involves extensive laboratory and clinical testing and external scientific review to ensure the safety and effectiveness of devices that sustain or support human life, prevent impairment of human health, or present an unreasonable risk of injury, (i.e., Class III medical devices). Collecting clinical trials data for a single, highly novel new product can cost upwards of \$100 million (Ernst and Young, 2009). Due to the regulatory oversight involved, medical devices for which PMA applications are submitted are highly novel and believed by the submitting firm to be commercially valuable.

RESEARCH METHOD

We test our hypotheses using a novel dataset on CVC investing in the medical device industry. We create this dataset by combining data from a number of sources, including CVC data, patent data, and PMA data. Our focal unit of analysis is the pair ('dyad') formed by a CVC investor and the start-up company in which it invests ($n = 128$).

Sample selection and data sources

CVC investments are equity investments made by established firms in start-up companies (Gompers and Lerner, 2000; Chesbrough and Tucci, 2003; Dushnitsky and Lenox, 2005). We are interested in the differential effects of user knowledge on established firms' innovative output and, therefore, we analyze the behaviors of corporate investors running systematic investment programs through which they invest in multiple start-ups. Our sample includes all CVC investments made by the four established medical device firms running formal CVC investing programs during the 1978 to 2007 time period. These four firms are Boston Scientific, Medtronic, Guidant, and Johnson & Johnson. We exclude firms with one-time or sporadic CVC investments, as the literature shows that such firms rarely reap strategic benefits from their investments (Gompers, 2002; Benson and Ziedonis, 2009).

We identify pairs of CVC investors and the start-ups in which they invest (i.e., dyads) using the VentureXpert database. VentureXpert aggregates data from the National Venture Capital Association and other sources and is used widely in studies of CVC activity (Dushnitsky and Lenox, 2005; Dushnitsky and Lenox, 2006; Benson and Ziedonis, 2009). Our unit of analysis is the dyad formed by an established firm and a start-up. If a start-up company received investment from more than one established firm, we count each CVC investor-start-up relationship as a unique dyad. Our data are cross-sectional in nature, as we follow each dyad over time. The original sample included 134 unique dyads, of which six were dropped due to incomplete data, leaving 128 unique dyads ($n = 128$). Over the sample period, the four established firms we studied invested approximately \$589 million in these start-ups through the process of CVC investment.

We create a novel dataset by matching each dyad with important innovation outcomes, namely patents and regulatory approval documents for highly novel

commercial products. We obtain the full text of all patents associated with every established firm and start-up company in our sample from the U.S. Patent and Trademark Office (USPTO) database. We verify corporate hierarchies and assignee names using Delphion's CorporateTree database to account for acquisitions and name changes. We then identify all citations to start-up patents in all CVC investor patents using software code written for this purpose. We also link our dyads with data documenting commercial products introduced by the established firms. We obtain the full text records of all PMA applications filed by Medtronic, Boston Scientific, Guidant, and Johnson & Johnson over the sample period from the PMA database of the U.S. Food and Drug Administration.

We assemble detailed career history data for the founder(s) of all start-ups in our sample. There is no one source for such data, hence career history data were aggregated from a variety of sources, including: personal, corporate, and institutional Web sites; SEC documents; professional directories; *Forbes* and *Businessweek* databases of executives and boards of directors; patent applications; social networking Web sites such as LinkedIn; and phone calls with founders.

Dependent variables

We wish to analyze the extent to which a start-up's knowledge is integrated into an established firm's patents and novel products, as a function of the source of a start-up's knowledge. Therefore, we construct two dependent variables in our analysis: the number of times an established firm cites the patents of a start-up in their own patents (*Backward Cites*), and the number of times an established firm draws on knowledge contained in the patents of a start-up in their own PMAs (*Product Generation*). Both patents and new product introductions are important technological outcomes for firms in the medical device industry (Ernst and Young, 2010).

Knowledge incorporated into new technologies (backward cites)

Our first dependent variable is *Backward Cites*, a count variable of the number of times a corporate investor's patents cite a given start-up's patents. 'Backward citations' to prior patents represent a legally enforceable boundary between existing prior art and the new invention (Trajtenberg, 1990).

Existing theory and a substantial empirical literature establishes backward patent citations as a measure of knowledge transfer between individuals or organizations (Mowery, Oxley, and Silverman, 1996; Griliches, 1998; Hall, Jaffe, and Trajtenberg, 2005). The literature suggests that the rate of backward citation peaks approximately three years after the original patent application is filed (Jaffe and Trajtenberg, 2002; Puranam and Srikanth, 2007). We collect backward patent citation data through 2010 to account for a minimum three-year lag (note that CVC investment data is collected through 2007).

Our measure of backward citations includes only citations made after the first round of CVC investment by the focal investor in the focal start-up. This allows us to truly compare the knowledge transferred from each start-up as part of the CVC investment process. We take any prior use of the start-up's knowledge by the established firm into account by controlling for backward citations made *prior* to the first round of CVC investment.

Knowledge incorporated into new products (product generation)

Our second dependent variable is *Product Generation*. This variable is a count of the number of PMA applications filed by an established firm that incorporate knowledge from a start-up company in which an established firm invests. Conceptually, we use this measure to indicate the extent to which innovative insights from a start-up company contribute to the creation of innovative products by the established firm who invests in the start-up. This variable is created by comparing the text contained in *all* of the start-up's previously filed patents with the text contained in each of the investor's subsequent successful PMA filings using a text-matching algorithm and calculating the 'knowledge overlap' between the two sets of documents. We then count the number of PMA applications that include a 'knowledge overlap' between the start-up's patents and the established firm's PMA applications and use this value as our dependent variable. We construct this variable through a multistep process.

The text-matching algorithm used to construct the *Product Generation* variable is based on the vector space model. The vector space model is a well-known technique in computer science for measuring the similarity between two documents (Salton, Wong, and Yang, 1975; Salton, 1988; Kwon and Lee, 2003; Manning, Raghavan, and Schutze, 2008). A related technique has been applied in the finance

literature to identify product market synergies in mergers and acquisitions (Hoberg and Phillips, 2010) and to analyze the relationship between IPO prospectuses and pricing (Hanley and Hoberg, 2010). To our knowledge, this is the first use of such an algorithm in the strategy literature.

We begin by identifying the documents we want to compare. For each dyad, we wish to compare each PMA application filed by an established firm with the knowledge contained in all the patents previously filed by the start-up. For each PMA application, we create a document that contains the text of all of the patents previously filed by a start-up. Differences in document length and word frequency are accounted for through the use of standard normalization techniques employed in the computer science literature (Manning *et al.*, 2008).

We then identify all the words present in each document (i.e., the established firm's PMA application and the document containing the corresponding set of patents filed by the start-up). We remove common 'stop words' (e.g., a, an, the, of), proper nouns (e.g., names, companies, cities, and countries), and generic words (e.g., application, manufacturing, facility) (Manning *et al.*, 2008). We are left with key words only. Through this process, we create a list of key words present in each PMA application by the established firm and a list of key words present in the set of patents previously filed by the start-up.

Next, we measure the knowledge overlap between the two documents. We do this by measuring the similarity between the words contained in the two sets of documents. This is accomplished mathematically by representing the two documents as vectors in a multidimensional vector space, with each dimension representing a particular key word, and then measuring the angle between the vectors. The intuition behind this approach is roughly as follows: two documents that overlap completely are represented by identical vectors, whereas two documents with no overlap might be thought of as vectors that are positioned at right angles to one another. Mathematically, we can then describe the similarity between the two documents by calculating the cosine of the angle between the two vectors (recall that each vector represents the content of a particular document). The cosine of the angle between the two documents that overlap completely is 1 (i.e., the cosine of 0 degrees), representing a high degree of knowledge overlap; whereas the cosine of the angle between the two documents with no overlap is 0 (i.e.,

the cosine of 90 degrees). The resulting value indicates the similarity between a particular PMA application and the set of patents previously filed by a start-up, and it ranges from 0 to 1; this value is our knowledge overlap score (and is also referred to as a cosine similarity score in the literature). A simplified example and graphical illustration of this process is provided in the Appendix.

We repeat this process for each PMA application filed by an established firm, thereby measuring the extent of the knowledge overlap between every PMA filed by the established firm and the corresponding set of a given start-up's patents. For each dyad, we then count the number of PMA applications filed by the established firm that have a knowledge overlap with the start-up's patents. In order to ensure that we are capturing a meaningful degree of knowledge transfer, we set a minimum threshold value at a knowledge overlap score of 0.3. This is in keeping with approaches in the computer science literature, which suggest taking into account the context in determining a threshold for ranking information (Manning *et al.*, 2008). We also explore sensitivity to higher and lower thresholds.¹ Thereby, the *Product Generation* variable indicates the extent to which knowledge from a particular start-up is incorporated throughout the established firm's product portfolio.

Independent variable

Physician-founded start-up (physician founded)

Our focal explanatory variable is a dummy variable equal to '1' if a start-up was founded by a *practicing* physician and '0' otherwise (*Physician Founded*). Because we wish to identify benefits of knowledge derived through use, we ensure this variable included only founders who were practicing physicians immediately prior to founding their start-up. Practicing physicians are the *users* of medical devices: they use devices to treat patients. Individuals who possess MD degrees but chose not to practice medicine are not considered users by our definition: these individuals possess medical knowledge derived through education, but they do not possess knowledge of medicine derived through use.

¹ Knowledge overlap scores in our sample range from 0 to 0.62. In unreported regressions, we experiment with thresholds set at 0.15, 0.2, 0.4, and 0.45. The effect of having a physician founder remains positive and significant at all of these levels, with the magnitude of the effect generally increasing as the threshold becomes more stringent.

Control variables

Start-up controls

The likelihood of citations to a given start-up's patents should increase with the number of patents the start-up company has and with the start-up's age (Sorensen and Stuart, 2000). Therefore, we control for the number of patents filed by the start-up over the lifetime of the company (*LN-Start-up Patents*) and for the age of the start-up (in years) based on the founding date in the VentureXpert database (*Age of Start-up*).

Dyadic controls

We include controls for dyad-level characteristics that might influence innovation outcomes. We include a dummy variable in our patent citation regressions for whether the established firm cited a start-up's patents prior to making the first round of CVC investment (*Cited Prior*). Existing backward citations from the established firm to the start-up's patents indicate the relevance of start-up knowledge *prior* to the formation of a CVC relationship and might influence the likelihood of subsequent citation. Rosenkopf and Almeida (2003) and Agarwal, Ganco, and Ziedonis (2009) used similar measures to account for the relevance of a given firm's patent portfolio to the citing firm.

We also control for the total amount of money invested by the established firm in the start-up company (*LN-CVC Investment*). The total amount of CVC funding has been shown to increase start-up performance (Park and Steensma, 2012) and investor patent output (Dushnitsky and Lenox, 2005). We control for the year in which the investment was made (*Year*) to account for any exogenous events related to the specific year in which the dyad was initiated.

CVC investor control variables

We include dummy variables for three of the four established firms in our sample—Boston Scientific, Medtronic, and Guidant—to account for heterogeneity among the established firms. To avoid multicollinearity, the fourth firm—Johnson & Johnson—is the benchmark.

Model and econometric approach

We test our first hypothesis by using backward patent citations to measure the knowledge transferred from a

start-up's patents to an established firm's patents (*Backward Cites*). These patent citation counts are bounded by zero and assume integer values. Given the nature of count data, we model the relationship between CVC investment and incorporation of knowledge from the start-up as a negative binomial distribution (Greene, 2008).

Our data exhibit overdispersion around the mean, suggesting that the negative binomial distribution is a better fit for the data than the Poisson distribution. The negative binomial specification adjusts for overdispersion in variance. It can be used to allow for observation-specific effects (Hausman, Hall, and Griliches, 1984) and has been widely used to analyze non-negative count data, such as patent citations, in which the assumption of mean equal to variance in a Poisson distribution is violated.

The expected number of backward citations from established firm i to start-up j 's patents is assumed to be an exponential function of whether the start-up founder is a physician or not, and X_{ij} , a vector of dyad, start-up, and established firm controls. We model the transfer of knowledge from a start-up's patents to an established firm's patents as:

$$E[\textit{backward cites}] = \exp(\beta_1 * \textit{Physician Founder} + \beta' X_{ij} + \varepsilon) \quad (1)$$

We account for differences in duration of the relationships by econometrically treating the age of the investment (*CVC-Investment Age*) as the period of exposure.

We test our second hypothesis by measuring the number of new products generated by the established firm that incorporate knowledge from start-up patents (*Product Generation*). Again, this variable is a count variable that exhibits overdispersion around the mean, therefore we estimate the relationship using negative binomial regression (Greene, 2008). We model *Product Generation* as a function of whether the start-up founder is a physician, and X_{ij} , a vector of dyad, start-up, and established firm controls:

$$E[\textit{Product Generation}] = \exp(\beta_1 * \textit{Physician Founder} + \beta' X_{ij} + \varepsilon) \quad (2)$$

All regressions are carried out with heteroskedasticity-robust standard errors. We control

for unobserved heterogeneity due to external shocks with year dummies. The specifications also include firm-specific dummies, which capture the time-invariant effects specific to each of the established firms.

FINDINGS

We present summary statistics for our main variables in Table 1 and a correlation matrix in Table 2. The descriptive statistics show that physicians founded 51 percent of the start-ups in our sample. Of the remaining 49 percent of start-ups in our sample, 13 percent were founded by academic scientists and 36 percent by former employees of medical device companies. Physician-founded start-ups produce similar numbers of patents as non-physician-founded start-ups.

The results of the negative-binomial maximum-likelihood models in Table 3 consistently support Hypothesis 1: established firms more frequently cite the patents of user-founded start-ups than they cite the patents of other start-ups. The dependent variable is *Backward Cites*. The coefficient on our focal explanatory variable, *Physician Founded*, is positive and highly significant ($p < 0.01$) (Table 3, Column 1). The effect is large in magnitude.

To illustrate the magnitude of the effect, we calculated the incidence rate ratio by exponentiating the coefficient of interest, holding all other coefficients constant (Table 3, Column 2). Established firms are expected to cite the patents of physician-founded start-ups 2.14 times more often than they cite the patents of non-physician-founded start-ups, all else equal. Holding all other variables at their means, the marginal impact of investing in a physician-founded start-up is an additional 1.3 citations.

Approximately 9 percent of the start-ups in our sample had no backwards citations. Start-ups may lack citations for two distinct reasons: a start-up may not have any patents or a start-up may possess patents that received no backward citations. We took this into account using zero-inflated negative binomial (ZINB) estimation, which allows us to first estimate the likelihood of falling into either category and then, taking this into account, estimate the expected number of citations (Greene, 2008). We allowed for zero inflation around the number of patents (Table 3, Column 3). A Vuong test confirmed the appropriateness of the zero-inflated model. The results of the zero-inflated negative binomial

Table 1. Summary statistics

	Full sample (<i>All start-ups</i>)			Physician-founded			Non-physician-founded		
	N = 128			N = 66			N = 62		
	<i>Mean</i>	<i>Min</i>	<i>Max</i>	<i>Mean</i>	<i>Min</i>	<i>Max</i>	<i>Mean</i>	<i>Min</i>	<i>Max</i>
Dependent variables									
Backward cites	7.859	0	196	9.394	0	196	6.226	0	121
Product generation	0.766	0	10	1.167	0	10	0.339	0	5
Independent variable									
Physician founded	0.516	0	1						
Start-up controls									
LN-start-up patents	2.148	0	4.4	2.115	0	4.3	2.182	0	4.4
Age of start-up	12.80	4	34	12.67	6	34	12.95	4	29
Dyadic controls									
Cited prior	0.203	0	1	0.258	0	1	0.145	0	1
LN-CVC investment	7.864	3.049	10.5	7.993	3.555	10.5	7.726	3.049	10.0
CVC investor controls									
<i>CVC investor is</i>									
Boston Scientific	0.156	0	1	0.197	0	1	0.113	0	1
Medtronic	0.281	0	1	0.227	0	1	0.339	0	1
Guidant	0.125	0	1	0.136	0	1	0.113	0	1
J&J	0.430	0	1	0.439	0	1	0.419	0	1

Table 2. Correlation coefficients of major variables

<i>Variable</i>	Backward cites	Product generation	Physician-founded start-up	LN-start-up patents	Age of start-up	Cited prior	LN-CVC investment
Backward cites	1.000						
Product generation	0.084	1.000					
Physician-founded	0.063	0.219	1.000				
LN-start-up patents	0.273	0.186	-0.014	1.000			
Age of start-up	0.177	-0.006	-0.025	0.253	1.000		
Cited prior	0.260	0.063	0.140	0.262	-0.030	1.000	
LN-CVC invest.	-0.119	-0.015	0.108	0.316	-0.275	0.066	1.000

regressions provide similarly strong support for Hypothesis 1: established firms are expected to cite the patents of user-founded start-ups 1.96 times more often than they cite the patents of non-physician-founded start-ups, all else equal.

The results in Table 4 provide consistent support for Hypothesis 2: established firms more frequently incorporate knowledge from user-founded start-ups than from other start-ups into their PMA applications. The dependent variable is *Product Generation*. We see that the coefficient on *Physician Founded* is strongly positive and significant based on negative binomial regression (Table 4, Column 1). The magnitude of the effect is large and strongly significant

($p < 0.05$). To calculate the magnitude of this effect, we exponentiate the coefficients (Table 4, Column 2). The expected number of PMAs an established firm will produce based on external knowledge sourced from a physician-founded start-up is 0.7, as compared to only 0.2 for knowledge sourced from a non-physician-founded start-up. Holding all else equal, an established firm is expected to introduce a new PMA incorporating knowledge from a start-up 3.08 times more frequently if the founder of the start-up is a physician. For robustness, we further employ zero-inflated negative binomial (ZINB) estimation (Table 4, Column 3). The results are nearly identical to the negative binomial estimation.

Table 3. Total number of backward citations by CVC investor patents to start-up patents

<i>Dependent variable:</i> Backward cites	(1) Negative binomial regression	(2) <i>Incidence ratio</i>	(3) Zero-inflated negative binomial regression	(4) <i>Incidence ratio</i>
Independent variable				
Physician founded	0.760*** [4.48]	2.14	0.671*** [3.29]	1.96
Start-up controls				
LN-start-up patents	1.244*** [12.53]	3.47	1.047*** [14.13]	2.85
Age of start-up	-0.143*** [-3.19]	0.87	-0.161*** [-4.43]	0.85
Dyadic controls				
Cited prior	1.362*** [2.77]	3.91	1.394*** [2.64]	4.03
LN-CVC investment	-0.351*** [-6.38]	0.70	-0.398*** [-7.36]	0.67
Year of investment	-0.121** [-1.88]	0.886	-0.154*** [-2.81]	0.86
CVC investor controls				
<i>CVC investor is:</i>				
Boston Scientific	1.087*** [6.68]	2.97	1.178*** [5.36]	3.25
Medtronic	0.844*** [5.04]	2.33	0.931*** [7.49]	2.54
Guidant	-0.489*** [-4.87]	0.61	-0.406*** [-8.87]	0.67
<i>Inflated (LN-start-up patents)</i>				
Observations	128		128	
Log pseudo-likelihood	-244.09		-242.85	
Wald chi-squared	87.39		80.03	

z-statistics in brackets, two-tailed tests *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Robust standard errors are clustered at the CVC-investor level. For all regressions, the omitted CVC investor is Johnson & Johnson.

Finally, it is worth noting that our results also show that established medical device firms differ in the extent to which they utilize knowledge from start-ups. Boston Scientific is the most likely to incorporate start-up knowledge into their patents, followed by Medtronic, Johnson & Johnson, and then Guidant (Table 3). Established firms also seem to differ in their ability to incorporate new knowledge into products. Johnson & Johnson is the most likely to incorporate knowledge into PMAs, followed by Medtronic, Guidant, and Boston Scientific (Table 4). These differences suggest that an established firm's ability to incorporate knowledge into its patents does not guarantee that knowledge will be incorporated into commercial products. This is perhaps not surprising, given the complexity of the innovation process as it goes from early develop-

ment through to commercialization. However, it does highlight the importance of examining multiple outcome measures when conducting studies of the corporate innovation process.

DISCUSSION

Our findings show that the source of innovative knowledge matters: accessing knowledge from different sources results in different levels of innovation for established firms. Established firms can gain significant and valuable product development benefits by accessing knowledge from innovative users. Moreover, in the context we study, these benefits are greater than the benefits derived by accessing knowledge from universities or other established firms.

Table 4. Total number of CVC investor PMA applications incorporating start-up knowledge

<i>Dependent variable:</i> Product generation	(1) Negative binomial regression	(2) <i>Incidence ratio</i>	(3) Zero-inflated negative binomial regression	(4) <i>Incidence ratio</i>
Independent variable				
Physician founded	1.124** [1.84]	3.08	1.231** [1.89]	3.43
Start-up controls				
LN-start-up patents	0.666*** [3.71]	1.95	0.056 [0.32]	1.06
Age of start-up	-0.057 [-0.91]	0.95	-0.045 [-0.71]	0.96
Dyadic controls				
LN-CVC investment	-0.451 [-1.60]	0.64	-0.306* [-1.67]	0.74
Year of investment	0.216*** [2.34]	1.24	0.217*** [2.19]	1.24
CVC investor controls				
<i>CVC investor is:</i>				
Boston Scientific	-1.066*** [-5.38]	0.35	-1.027*** [-3.19]	0.36
Medtronic	-0.001 [-0.00]	1.00	0.033 [0.16]	1.03
Guidant	-0.211 [-0.54]	0.81	-0.278 [-0.94]	0.76
<i>Inflated (LN-start-up patents)</i>				
			-0.917 [-1.45]	
Observations	128		128	
Log pseudo-likelihood	-116.87		-113.89	
Wald chi-squared	49.04		38.89	

z-statistics in brackets, two-tailed tests *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Robust standard errors are clustered at the CVC-investor level. For all regressions, the omitted CVC investor is Johnson & Johnson.

User innovators are a critical and important component of the innovation landscape. The importance of users as a source of innovation has been largely overlooked in the strategy, product development, and search literatures. This study establishes the distinct value that established firms can generate by utilizing knowledge sourced from innovative users.

To date, research on user innovation has focused largely on documenting the prevalence and importance of user innovation and identifying ways in which established firms can work with users (Urban and von Hippel, 1988; Lilien *et al.*, 2002; von Hippel and Katz, 2002; Jeppesen and Molin, 2003; Jeppesen and Frederiksen, 2006; Shah, 2006). The existing literature has not explained why the insights provided by innovative users might enrich the R&D processes of established firms nor illustrated the differential value that users might bring to established

firms. In this manuscript, we build theory to explain why insights from innovative users will be beneficial to established firms. Our theory is grounded in empirical research documenting the unique elements of the user innovation process. We then examine the effects of user, as compared to nonuser, knowledge on the invention and innovation outcomes of established firms. We find that user knowledge is more beneficial to established firms with respect to two important outcomes of the R&D process: patenting and new product development. These findings demonstrate the importance and relevance of user innovation to the technological performance of established firms, thereby connecting the phenomenon of user innovation to the field of strategic management.

These findings also contribute to the literature on product development. The process by which new

products are developed within firms generally begins with market research aimed at identifying the needs of customers and progresses through a number of internal processes by which a firm matches customer needs with its own capabilities to establish a product vision and then designs, produces, advertises, and distributes a product (Brown and Eisenhardt, 1995; Prandelli, Sawhney, and Verona, 2008; Taylor, 2010). However, from a practical perspective, most new product development attempts fail (Cooper and Kleinschmidt, 1995; Prandelli *et al.*, 2008). There are many causes for these problems, and the existing literature has examined issues related to the internal dynamics of the firm in detail (e.g., Brown and Eisenhardt, 1995; Taylor, 2010). Our theory and findings suggest another pathway for increasing returns from the new product development process: by working with innovative users, established firms may be able to access better market and technological insights that ultimately result in the creation of more products. That is to say, knowledge gained from innovative users appears to be particularly generative for the new product development process. To this end, it is important to note that there is a variety of methods through which firms might engage innovative users. CVC investment in user-founded firms is just one of several methods; other methods include, but are not limited to, participation in user innovation communities, consulting or licensing arrangements, and implementation of the lead user method.

This article also contributes to the literature on CVC investing. Scholars have posited that CVC investment benefits established firms by allowing them to access the valuable technological knowledge possessed by start-ups (Dushnitsky and Lenox, 2005; Schildt, Maula, and Keil, 2005; Wadhwa *et al.*, 2010). Correlational support for this claim exists, showing that engaging in CVC investment leads to the production of greater numbers of patents. However, this research has not been able to document the actual transfer of knowledge and, hence, the mechanisms underlying how established firms derive benefit from CVC investment are unclear.² We provide evidence supporting the idea that CVC investing provides a ‘window on technology’ for established firms through which they access innovative insights from start-up firms. We show that

these insights are used as inputs into the established firm’s R&D processes and integrated into new products.

This article also makes a methodological contribution to the management literature by introducing a method for tracing knowledge flows that is based on the vector space model drawn from the field of computer science. Knowledge is a prominent theoretical construct in the strategy and technology management literatures (Cohen and Levinthal, 1990; Nonaka and Takeuchi, 1995; Grant, 1996; von Krogh, Nonaka, and Nishigushi, 1999; Helfat and Raubitschek, 2000). However, our ability to examine knowledge flows and the effects of knowledge flows using quantitative methods is somewhat limited; most scholars studying knowledge use either qualitative methods or quantitative methods based on backwards patent citation measures. Utilizing the vector space model to directly measure knowledge flows—as we do through the use of the *Product Generation* measure—will improve the field’s ability to test and improve existing theories and build new theory.

Finally, our study provides practical guidance for established firms as they search externally for knowledge. The existing literature on search and organizational learning suggests that organizations seek knowledge that is different from that the firm already possesses and from partners whose organizational characteristics are similar to their own (see, for example Mowery *et al.*, 1996; Lane and Lubatkin, 1998; Katila and Ahuja, 2002). We suggest that organizations also consider the *source* of innovation-related knowledge by investigating where the roots of that knowledge lie: in use, academic scientific inquiry, or the industrial product development process. This insight might be particularly helpful to organizations limited in their ability to search for knowledge, due to resource constraints or other limitations.

We hope this manuscript will be the first in a series of articles that investigate the value of knowledge sourced from innovative users to established firms. Much work remains to be done in this area. Ours is a single industry study and future research might wish to replicate and expand these findings in other industries. Innovative users are present in many industries (von Hippel, 1988; Shah and Tripsas, 2007; von Hippel, de Jong, and Flowers, 2010; Shah, Winston Smith, and Reedy, 2012). In addition, while we investigate two outcomes that are of immense importance to product development within

² We thank an anonymous reviewer for pointing this out as a contribution of this study.

established firms, other outcome variables remain to be investigated, particularly in light of the fact that established firms do source knowledge from a variety of external knowledge sources. Future research might examine additional outcome variables to refine our understanding of the differential benefits derived from different knowledge sources—such as universities, competitors, and users. We encourage scholars to examine differences in the types of products produced using insights from different knowledge sources (e.g., products embodying incremental versus radical innovations), as well as the effects of differences in the characteristics of knowledge sourced from different knowledge sources (e.g., by working with academics, firms might access very early-stage and cutting-edge technological insights, whereas by working with other established firms or previous employees of those firms, firms might access insights pertaining to manufacturing processes or architectural innovations).

CONCLUSION

Innovative users matter. Scholars have long been aware that innovations come from a variety of sources, including employees of firms, scientists at academic institutions, and users. While much effort has gone into investigating each of these sources individually, little research has sought to compare the differential effects of knowledge from these sources on the innovation outcomes of established firms. We find that sourcing knowledge from users, as compared to nonusers, provides greater innovation-related benefits to established firms.

ACKNOWLEDGEMENTS

We wish to thank Christine Beckman, Emily Cox, Gary Dushnitsky, Kathy Eisenhardt, Suresh Kotha, Alan Meyer, Nandini Rajagopalan, seminar participants at the Georgia Institute of Technology, and attendees of the 2011 West Coast Research Symposium, the 2011 Academy of Management Conference, and the 2011 Strategic Management Society Conference for their thoughtful comments on earlier versions of this article. Snehal Awate and Rose Kim provided excellent research assistance.

REFERENCES

- Agarwal R, Ganco M, Ziedonis R. 2009. Reputations for toughness in patent enforcement: implications for knowledge spillovers via inventor mobility. *Strategic Management Journal* **30**(13): 1349–1374.
- Arora A, Gambardella A. 1990. Complementarity and external linkages: the strategies of the large firms in biotechnology. *Journal of Industrial Economics* **38**: 361–379.
- Beckman CM, Burton MD. 2008. Founding the future: path dependence in the evolution of top management teams from founding to IPO. *Organization Science* **19**(1): 3–24.
- Benson D, Ziedonis RH. 2009. Corporate venture capital as a window on new technologies: implications for the performance of corporate investors when acquiring start-ups. *Organization Science* **20**: 329–351.
- Boeker W. 1988. Organizational origins: entrepreneurial and environmental imprinting at the time of founding. In *Ecological Models of Organizations*, Carroll G (ed). Ballinger Publishing: Cambridge, MA.
- Brown JS, Duguid P. 2001. Knowledge and organization: a social-practice perspective. *Organization Science* **12**(2): 198–213.
- Brown SL, Eisenhardt KM. 1995. Product development: past research, present findings, and future directions. *Academy of Management Review* **20**: 343–348.
- Chandler AD Jr. 1994. *Scale and Scope: The Dynamics of Industrial Capitalism*. Harvard University Press: Cambridge, MA.
- Chesbrough H. 2003. *Open Innovation: The New Imperative for Creating and Profiting from Technology*. Harvard Business School Press: Boston, MA.
- Chesbrough H, Tucci C. 2003. Corporate venture capital in the context of R&D. Working paper, University of California, Berkeley, CA.
- Clark K, Fujimoto T. 1991. *Product Development Performance*. Harvard Business School Press: Boston, MA.
- Cohen WM, Levinthal DA. 1990. Absorptive capacity: a new perspective on learning and innovation. *Administrative Science Quarterly* **35**(1): 128–152.
- Cooper RG, Kleinschmidt EJ. 1995. Benchmarking the firm's critical success factors in new product development. *Journal of Product Innovation Management* **12**: 374–391.
- Delmar F, Shane S. 2006. Does experience matter: the effect of founding team experience on the survival and sales of newly founded ventures. *Strategic Organization* **4**(3): 215–247.
- Dushnitsky G, Lenox M. 2005. When do firms undertake R&D by investing in new ventures? *Strategic Management Journal* **26**(10): 947–965.
- Dushnitsky G, Lenox M. 2006. When do incumbents learn from entrepreneurial ventures? Corporate venture capital and investing firm innovation rates. *Research Policy* **34**(5): 615–639.

- Dushnitsky G, Shaver M. 2009. Limitations to interorganizational knowledge acquisition: the paradox of corporate venture capital. *Strategic Management Journal* **30**(10): 1045–1064.
- Eisenhardt KM, Schoonhoven CB. 1990. Organizational growth: linking funding team, strategy, environment, and growth among U.S. semiconductor ventures, 1978–1988. *Administrative Science Quarterly* **35**: 504–529.
- Ernst and Young. 2009. Pulse of the industry: medical technology report 2009. Available at: [http://www.ey.com/Publication/vwLUAssets/Pulse_of_the_industry_2009:_medtech_review/\\$FILE/Pulse_Final.pdf](http://www.ey.com/Publication/vwLUAssets/Pulse_of_the_industry_2009:_medtech_review/$FILE/Pulse_Final.pdf) (accessed 20 December 2010).
- Ernst and Young. 2010. Pulse of the industry: medical technology report 2010. Available at: http://www2.eycom.ch/publications/items/med/2010_pulse_of_the_industry/201010_EY_Pulse_of_the_industry.pdf (accessed 20 December 2010).
- Franke N, Shah SK. 2003. How communities support innovative activities: an exploration of assistance and sharing among end-users. *Research Policy* **32**: 157–178.
- Franz K. 2005. *Tinkering: Consumers Reinvent the Early Automobile*. University of Pennsylvania Press: Philadelphia, PA.
- Gompers PA. 2002. Corporations and the financing of innovation: the corporate venturing experience. *Economic Review of the Federal Reserve Board of Atlanta* **87**(4): 1–17.
- Gompers PA, Lerner J. 2000. The determinants of corporate venture capital success: organizational structure, incentives, and complementarities. In *Concentrated Corporate Ownership*, Morck RK (ed). University of Chicago Press: Chicago, IL.
- Grant RM. 1996. Toward a knowledge-based theory of the firm. *Strategic Management Journal*, Winter Special Issue **17**: 109–122.
- Greene WH. 2008. *Econometric Analysis*. Pearson Prentice Hall: Upper Saddle River, NJ.
- Griliches Z. 1998. *R&D and Productivity: The Econometric Evidence*. University of Chicago Press: Chicago, IL.
- Guimerà R, Uzzi B, Spiro J, Amaral LAN. 2005. Team assembly mechanisms determine collaboration network structure and team performance. *Science* **308**(5722): 697–702.
- Hall BH, Jaffe A, Trajtenberg M. 2005. Market value and patent citations. *RAND Journal of Economics* **36**(1): 16–38.
- Hanley KW, Hoberg G. 2010. The information content of IPO prospectuses. *Review of Financial Studies* **23**(7): 2821–2864.
- Hausman J, Hall BH, Griliches Z. 1984. Econometric models for count data with an application to the patents-R&D relationship. *Econometrica* **52**(4): 909–938.
- Helfat CE. 1997. Know-how and asset complementarity and dynamic capability accumulation: the case of R&D. *Strategic Management Journal* **18**(5): 339–360.
- Helfat CE, Raubitschek RS. 2000. Product sequencing: co-evolution of knowledge, capabilities, and products. *Strategic Management Journal* **21**(10/11): 961–979.
- Henderson R. 1993. Underinvestment and incompetence as responses to radical innovation: evidence from the photolithographic alignment equipment industry. *RAND Journal of Economics* **24**(2): 248–270.
- Hoberg G, Phillips G. 2010. Product market synergies and competition in mergers and acquisitions: a text-based analysis. *Review of Financial Studies* **23**(10): 3773–3811.
- Jaffe AB, Trajtenberg M. 2002. *Patents, Citations, and Innovations: A Window on the Knowledge Economy*. MIT Press: Cambridge, MA.
- Jeppesen LB, Frederiksen L. 2006. Why do users contribute to firm-hosted user communities? The case of computer controlled music instruments. *Organization Science* **17**: 22–44.
- Jeppesen LB, Molin MJ. 2003. Consumers as co-developers: learning and innovation outside the firm. *Technology Analysis & Strategic Management* **15**(3): 363–384.
- Katila R, Ahuja G. 2002. Something old, something new: a longitudinal study of search behavior and new product introduction. *Academy of Management Journal* **45**: 1183–1194.
- Katila R, Rosenberger JD, Eisenhardt KM. 2008. Swimming with sharks: technology ventures, defense mechanisms, and corporate relationships. *Administrative Science Quarterly* **53**(2): 295–332.
- Kline R, Pinch T. 1996. Users as agents of technological change: the social construction of the automobile in the rural United States. *Technology and Culture* **37**: 763–795.
- Kwon OW, Lee JH. 2003. Text categorization based on k-nearest neighbor approach for Web site classification. *Information Processing & Management* **39**(1): 25–44.
- Lane PJ, Lubatkin M. 1998. Relative absorptive capacity and interorganizational learning. *Strategic Management Journal* **19**(5): 461–477.
- Leiponen A, Helfat CE. 2010. Innovation objectives, knowledge sources, and the benefits of breadth. *Strategic Management Journal* **31**(2): 224–236.
- Levin RC, Cohen WM, Mowery DC. 1985. R&D appropriability, opportunity, and market structure: new evidence on some Schumpeterian hypotheses. *American Economic Review* **75**(2): 20–24.
- Lilien GL, Morrison PD, Searls K, Sonnack M, von Hippel E. 2002. Performance assessment of the lead user idea generation process. *Management Science* **46**(12): 1513–1527.
- Lüthje C, Herstatt C, von Hippel E. 2005. The dominant role of ‘local’ information in user innovation: the case of mountain biking. *Research Policy* **34**(6): 951–965.
- Manning C, Raghavan P, Schütze H. 2008. *Introduction to Information Retrieval*. Cambridge University Press: Cambridge, U.K.

- March J. 1991. Exploration and exploitation in organizational learning. *Organization Science* 2: 71–87.
- Merton RK. 1979. *Sociology of Science*. University of Chicago Press: Chicago, IL.
- Mody CCM. 2006. Universities, corporations, and instrumental communities: commercializing probe microscopy, 1981–1996. *Technology and Culture* 47: 56–80.
- Mowery DC, Oxley JE, Silverman BS. 1996. Strategic alliances and interfirm knowledge transfer. *Strategic Management Journal*, Winter Special Issue 17: 77–91.
- Nelson RR, Winter SG. 1982. *An Evolutionary Theory of Economic Change*. Harvard University Press: Cambridge, MA.
- Nonaka I, Takeuchi H. 1995. *The Knowledge-Creating Company*. Oxford University Press: New York.
- Ogawa S. 1998. Does sticky information affect the locus of innovation? Evidence from the Japanese convenience store industry. *Research Policy* 26: 777–790.
- Oudshoorn N, Pinch T (eds). 2003. *How Users Matter: The Co-Construction of Users and Technology*. MIT Press: Cambridge, MA.
- Park HD, Steensma HK. 2012. When does corporate venture capital add value for new ventures? *Strategic Management Journal* 33(1): 1–22.
- Prandelli E, Sawhney MS, Verona G. 2008. *Collaborating with Customers to Innovate: Conceiving and Marketing Products in the Networking Age*. Edward Elgar Publishing: Northampton, MA.
- PricewaterhouseCoopers. 2006. *Corporate Venture Capital Activity on the Rise in 2006*. Pricewaterhouse Coopers: Washington, DC.
- Puranam P, Srikanth K. 2007. What they know vs. what they do: how acquirers leverage technology acquisitions. *Strategic Management Journal* 28(8): 805–825.
- Riggs W, von Hippel E. 1994. Incentives to innovate and the sources of innovation: the case of scientific instruments. *Research Policy* 23: 459–469.
- Rosenkopf L, Almeida P. 2003. Overcoming local search through alliances and mobility. *Management Science* 49(6): 751–766.
- Salton G. 1988. Term-weighting approaches in automatic text retrieval. *Information Processing and Management: An International Journal* 24(5): 513–523.
- Salton G, Wong A, Yang CS. 1975. A vector space model for automatic indexing. *Communications of the ACM* 18(11): 613–620.
- Schildt HA, Maula MVJ, Keil T. 2005. Explorative and exploitative learning from external corporate ventures. *Entrepreneurship Theory and Practice* 29(4): 493–515.
- Schumpeter J. 1942. *Capitalism, Socialism, and Democracy* (2nd edn). George, Allen & Unwin, Ltd.: London, U.K.
- Shah SK. 2005. Open beyond software. In *Open Sources 2: The Continuing Evolution*, Dibona C, Cooper D, Stone M (eds). O'Reilly Media: Sebastopol, CA.
- Shah SK. 2006. Motivation, governance, and the viability of hybrid forms in open source software development. *Management Science* 52(7): 1000–1014.
- Shah SK, Tripsas M. 2007. The accidental entrepreneur: the emergent and collective process of user entrepreneurship. *Strategic Entrepreneurship Journal* 1(1): 123–140.
- Shah SK, Winston Smith S, Reedy EJ. 2012. Who are user entrepreneurs? Findings on innovation, founder characteristics, and firm characteristics. Kauffman Foundation: Kansas City, MO.
- Siegel R, Siegel E, MacMillan I. 1988. Corporate venture capitalists: autonomy, obstacles, and performance. *Journal of Business Venturing* 3: 233–247.
- Sorensen J, Stuart TE. 2000. Aging, obsolescence, and organizational innovation. *Administrative Science Quarterly* 45: 81–112.
- Tapscott D. 2006. *Wikinomics*. Penguin Group: New York.
- Taylor A. 2010. The next generation: technology adoption and integration through internal competition in new product development. *Organization Science* 21(1): 3–41.
- Trajtenberg M. 1990. A penny for your quotes: patent citations and the value of innovations. *RAND Journal of Economics* 21(1): 172–187.
- Tushman ML, Anderson P. 1986. Technological discontinuities and organizational environments. *Administrative Science Quarterly* 31(3): 439–465.
- Tyre MJ, von Hippel E. 1997. The situated nature of adaptive learning in organizations. *Organization Science* 8(1): 71–83.
- United States Food and Drug Administration. 2012. Premarket approval (PMA) information. Available at: <http://www.Fda.Gov/Medicaldevices/Deviceregulationandguidance/Howtomarketyourdevice/Premarket-submissions/Premarketapprovalpma/Default.Htm> (accessed 6 February 2012).
- Urban GL, von Hippel E. 1988. Lead user analyses for the development of new industrial products. *Management Science* 34(5): 569–582.
- von Hippel E. 1988. *The Sources of Innovation*. Oxford University Press: New York.
- von Hippel E. 1994. 'Sticky information' and the locus of problem solving: implications for innovation. *Management Science* 40(4): 429–439.
- von Hippel E. 2005. *Democratizing Innovation*. MIT Press: Cambridge, MA.
- von Hippel E, de Jong J, Flowers S. 2010. Comparing business and household sector innovation in consumer products: findings from a representative study in the U.K. *Management Science* 58: 1669–1681.
- von Hippel E, Katz R. 2002. Shifting innovation to users via toolkits. *Management Science* 48(7): 821–833.
- von Hippel E, von Krogh G. 2003. Open source software and the private-collective innovation model: issues for organization science. *Organization Science* 32(2): 209–233.

- von Krogh G, Nonaka I, Nishigushi T (eds). 1999. *Knowledge Creation: A Source of Value*. Macmillan: London, U.K.
- Wadhwa A, Phelps C, Kotha S. 2010. Creating exploratory innovations by learning from entrepreneurial ventures. In *New Frontiers in Entrepreneurship: Recognizing, Seizing, and Executing Opportunities*, Audretsch DB, Dagnino GB, Faraci R, Hoskisson R (eds). Springer: New York.
- Wenger E. 1998. *Communities of Practice*. Cambridge University Press: New York.
- Winston Smith S. 2009. The company they keep: innovation returns to corporate venture capital in the medical device industry, entrepreneurial clinicians, and competitive coinvestors. Available at: <http://ssrn.com/paper=1095651> (accessed 11 November 2010).

APPENDIX

We provide a hypothetical example to illustrate our text matching algorithm. We compare two patents to one PMA application. For simplicity, assume that the patents and PMA applications of interest contain only three words: 'bone,' 'graft,' and 'chamber.' Thus, we have a three-dimensional vector space. We

match the PMA document with Patent 1 and Patent 2. Each document is represented as a word vector. Every word forms one component of the vector and the weighted frequency of each word provides the magnitude. Graphically, we can represent this as a three-dimensional vector space as shown in Figure 1. Thus, in this diagram, the PMA document is more similar to Patent 2 than it is to Patent 1 and, thus, it will have a higher knowledge similarity score.

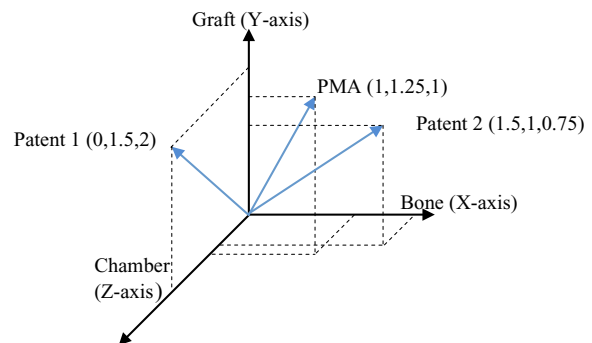


Figure 1. Illustration of knowledge matching technique