

# Knowledge Bridging by Biotechnology Start-ups\*

*by*

David H. Hsu and Kwanghui Lim

This draft: May 2005 (original draft: August 2004)

## ABSTRACT

We investigate a type of technological boundary spanning-oriented search in which firms apply knowledge from one technical domain to innovate in another, a phenomenon we term knowledge bridging. In an analysis of all the biotechnology firms founded to commercialize recombinant DNA technology, we examine the consequences of knowledge bridging at the patent level of analysis, as well as the determinants of this search behavior at the firm-year level of analysis. The empirical setting allows us to establish a common technological starting point and examine technological search behavior over time. At the patent level of analysis, we find that knowledge bridging is significantly associated with measures of patent value. At the firm-year level, we find that firms vary in knowledge bridging use, which in turn is correlated with firms' initial search direction, funding environment munificence, and product development success (or lack thereof). The paper contributes to both the conceptual and empirical literature on technological search.

*Keywords:* knowledge exploration, technological boundary spanning, innovation, entrepreneurship, start-ups, biotechnology, patents.

---

\* The Wharton School, 2000 Steinberg Hall-Dietrich Hall, University of Pennsylvania, Philadelphia, PA 19104, [dhsu@wharton.upenn.edu](mailto:dhsu@wharton.upenn.edu) and NUS Business School, National University of Singapore, BIZ1-04-27, Singapore 117592, Republic of Singapore, [k@kwanghui.com](mailto:k@kwanghui.com). We thank Kathy Ku of the Stanford University Office of Technology Licensing for allowing us to access the Cohen Boyer patent records. We thank Scott Shane, Scott Stern, and participants of a 2004 Academy of Management session for useful comments and discussions. Josh Lerner generously provided access to his biotechnology index. Tong Zhao, Zhuang Wenyue and Mike Gonsalves provided valuable research assistance. We acknowledge use of the NUS patent database. We thank the Mack Center for Technological Innovation at Wharton and the National University of Singapore Business School for funding this project.

## **I. Introduction**

The notion that organizations engage in local search is well established in the management literature. Firms have a propensity to explore knowledge that is familiar and within easy reach from their existing geographic and technological positions. The underpinnings for this behavior have been explored at multiple levels of analysis, ranging from individual-level explanations of bounded rationality (March and Simon, 1958) to firm-level capabilities, routines, and learning myopia (Nelson and Winter, 1982; Levinthal and March, 1993). New venture “imprinting” by their founders (Stinchcombe, 1965) and the long-lasting impact of firms’ initial conditions (Cockburn, Henderson and Stern, 2000) also suggest mechanisms by which firms’ search behavior is perpetuated.

In environments in which innovation is important as the basis for competition, firms and their managers may be particularly concerned about the long term competitive effects of local search (March, 1990). Not surprisingly, then, there has been considerable interest in mechanisms associated with overcoming the constraints of local search. For example, Mowery, Oxley, and Silverman (1996) examine strategic alliances as a mechanism for overcoming local search. The mobility of engineers and scientists in the labor market is another means by which firms can overcome local search (Rosenkopf and Almeida, 2003). A third mechanism for overcoming local search has received much less empirical attention: applying knowledge and ideas from one technical domain to solve problems in another area.

In this paper, we assess the empirical importance of using ideas from one technical domain to create innovation in another, a phenomenon we term knowledge bridging. We ask two related research questions: (1) Is knowledge bridging behavior at the patent level related to innovation outcomes? and (2) what are the determinants of knowledge bridging at the firm level? We therefore plan to analyze both the consequences of knowledge bridging behavior and its antecedents. Economists, historians of technical change, and management scholars alike have analyzed knowledge recombination extensively (e.g., Schumpeter, 1934; Usher, 1954; Hargadon and Sutton, 1997; Fleming, 2001). However, this literature has been more conceptual than empirical<sup>1</sup>, and more concerned with the process of invention than commercialization. Our paper contributes to the literature by conceptually identifying knowledge bridging as a distinct form of

---

<sup>1</sup> See Fleming (2001) and Fleming and Sorenson (2004) for exceptions. These papers, while empirical, are primarily concerned with technological recombinations in the process of invention.

technical boundary spanning that incorporates knowledge recombination (and other mechanisms), as well as showing that it is important empirically both at the firm and invention levels of analysis. Specifically, we show that knowledge bridging has an important effect on the quality and generality of patented innovations. We also show that firm-level variables (such as founding characteristics and feedback in the product development process) and industry-level variables (munificence in the funding environment) affect a firm's decision to engage in knowledge bridging.

Empirically studying knowledge bridging requires constructing novel measures of the concept that will allow for across-patent and across-firm comparisons. We start with the premise that the extent to which patents and firms cite patents in different technical areas indicates the degree of knowledge bridging from those areas. We construct knowledge bridging measures at the patent and firm levels based on the technical areas of innovation in relation to technical areas referenced to arrive at those innovations. To construct the firm level measure of knowledge bridging, we compare the primary patent classes of a firm's patents in relation to the primary patent classes the firm cites on an annual basis to construct a longitudinal dataset.

To investigate the temporal aspects of knowledge bridging-based technological search from firms' inception, we require a sample of firms that were founded to exploit a given technological opportunity. While firms will often start from different initial conditions, constructing a sample that is relatively uniform in the basic technology upon which they are capitalizing is important. This allows us to observe both differences in initial conditions, together with evolutionary development patterns for firms that are homogenous with respect to initial technology. This study design is similar in spirit as that used by Shane (2000) in his study of various recognized entrepreneurial opportunities arising from a common three-dimensional printing technology. Our study also employs an empirical strategy similar to Cockburn, Henderson and Stern (2000) in looking for residual effects net of initial conditions.

The historical circumstances surrounding the commercialization of recombinant DNA technology via non-exclusive licensing of the Cohen-Boyer patent by Stanford University is a fortuitous one from a research standpoint in meeting the above criteria. We are able to examine all de novo start-ups founded to commercialize this technology due to generous access to the Stanford licensing program records by the Stanford Office of Technology Licensing. The recombinant DNA innovation allowed DNA from two or more sources to be combined into a

single source. The commercialization of this innovation is also of considerable interest since it essentially initiated the biotechnology industry.<sup>2</sup> Using this empirical setting, we are able to study both the impact of start-ups' knowledge bridging search behavior at the patent level, as well as factors that affect the use of knowledge bridging at the firm-year level of analysis.

The empirical results are presented in two stages. In the first stage, we find that patents employing a high degree of knowledge bridging have more impact than those with a low degree of knowledge bridging, as measured by forward citation counts. They are also more general-purpose in nature as they are cited by patents from a broader range of technological areas than patents employing a low degree of knowledge bridging. In the second stage of the analysis, we document the substantial variation in firms' use of knowledge bridging in their business development. We then investigate covariates of the degree to which firms change their technological search patterns, taking as the unit of analysis a firm-year. We find that firms differ in their use of knowledge bridging, which is correlated with firms' initial technological position; FDA drug approvals/terminations; funding environment munificence; and their entry order.

In the next section, we review the relevant literature and derive empirical predictions. Section III discusses the data and method employed, while section IV presents the empirical results. A final section concludes and discusses future directions.

## **II. Literature and Hypothesis Development**

### *A. "Imprinting" of Firms' Initial Technological Positions*

Founder imprinting and search routinization are powerful reasons for organizational inertness. Founder imprinting can be manifested in firm philosophies, policies and procedures as they relate to organizational culture, human resource management, research and development practices (Stinchcombe, 1965). Search has its roots in individual cognitive patterns, suggesting that search patterns tend to be subject to routinization (March and Simon, 1958). Such behavior is also part and parcel of organizational life (Nelson and Winter, 1982).

---

<sup>2</sup> The biotechnology industry is quite a technologically dynamic one, and therefore presents an interesting empirical setting for this study. As of 2003, for example, biotechnology innovations have accounted for 155 U.S. Federal Drug Administration (FDA) approved drugs, with over 370 biotechnology clinical trials and vaccines in development (BIO website, accessed May 24, 2004). Furthermore, biotechnology firms are a significant source of upstream innovation for pharmaceutical firms (Gans, Hsu and Stern, 2002): of the 691 new chemical entities approved by the FDA between 1963 and 1999, 38 percent were licensed by pharmaceutical firms, primarily from biotechnology firms (DiMasi, 2000).

Indeed, Helfat (1994) and Henderson and Cockburn (1994) find substantial (and varied) fixed firm effects in research and development across two different industries: petroleum and pharmaceuticals, suggesting substantial within-industry heterogeneity in R&D investment strategy. Furthermore, Boeker (1989) found that semiconductor start-ups typically maintained the corporate strategies they had at the time of founding. Cockburn, Henderson and Stern (2000) suggest that initial conditions are important in explaining the level of adoption of a firm's R&D policies in a later period (in their case, the initial extent of science-driven drug discovery by pharmaceutical firms persisted over long periods). Their results also suggest that while initial conditions are important, they are not all-encompassing in explaining variation in the adoption of strategies that may affect organizational performance.

Because individuals have different backgrounds, knowledge, and skills, they will likely respond to entrepreneurial opportunity windows in different ways (Shane, 2000). This implies that individual beliefs about exploiting even a common technological innovation may very well be varied, which can account for differences in the initial position of entrepreneurial start-ups. Given the dual forces of technological search routinization and founder imprinting, we *expect firms' initial search conditions to persist over time.*

### *B. Mechanisms of Overcoming Local Search*

Given the constraints imposed by local search, it is not surprising that scholars have investigated ways to overcome such behavior, particularly in the context of fast-changing R&D environments. For example, Mowery, Oxley, and Silverman (1996) and Stuart and Podolny (1996) examine strategic alliances as a mechanism for overcoming local search. As well, engineers and scientists are mobile, and hiring on the scientific labor market is another means by which firms can access distant knowledge (Rosenkopf and Almeida, 2003). The efficacy of the latter mechanism is likely to be context-specific, however. For example, Zucker, Darby and Brewer (1998) find that in the early biotechnology industry, the scarce resource was specialized knowledge resident in highly accomplished university scientists. The fact that these scientists were for the most part not mobile helps explain the observed geographic concentration of the industry (large concentrations of firms located near academic centers of excellence in biology and chemistry).

Another possible mechanism for overcoming local search is to recombine existing organizational knowledge (Kogut and Zander, 1993). Doing so would inject valuable variation into the search process (by experimenting with new components or recombining old ones), perhaps in a manner akin to the act of invention itself (Fleming, 2001). This strand of the literature is highly relevant to our study. We therefore discuss this stream of research in some depth.<sup>3</sup>

Consider Figure 1, which illustrates two types of knowledge bridging. A first type involves borrowing ideas and knowledge from several areas and recombining them for innovation in another area. To illustrate this, consider the academic field of strategic management. It borrows knowledge from a number of disparate fields such as economics, sociology, history, and political science—and recombines insights and methods from these fields to create new knowledge about strategy. A second type of bridging involves taking knowledge from one domain and reapplying it to another. To illustrate this form of knowledge bridging “by analogy,” consider the birth and development of the academic field of evolutionary economics. Borrowing key ideas from evolutionary biology—such as principles of genetic variation and selection—evolutionary economists have advanced our knowledge of how organizations evolve in a way analogous to that of living organisms.

The phenomenon we wish to explore empirically, the idea that new knowledge comes from the recombination or hybridization of old ones, is not new. Schumpeter (1934: 65-66) conceived innovation as “carrying out new combinations.” Usher (1954: 21), in his classic work, argued: “There are other discontinuities that may be overcome, through some act of synthesis. The establishment of new organic relations among ideas, or among material agents, or in patterns of behavior is the essence of all invention and innovation.” More recently, Weitzman (1998) developed a model of economic growth that depends on idea recombinations, and cites supporting examples from a wide array of settings, including advances in mathematics and physics. Olsson and Frey (2002) examine the recombination of ideas for entrepreneurial exploitation, together with its limits.<sup>4</sup>

---

<sup>3</sup> The weighting of this discussion relative to the strategic alliance and scientific labor market search mechanisms does not necessarily reflect our expectations of the relative importance of mechanisms; rather this discussion reflects the comparative neglect of knowledge bridging in the empirical literature on technological search.

<sup>4</sup> A related strand of research has observed that knowledge spillovers exist as a result of imperfect appropriability (e.g., Griliches, 1992), and that geographic proximity may be needed to capture the spillovers (Jaffe, Trajtenberg and Henderson, 1993). The availability of knowledge spillovers is aided by U.S. institutions such as the patent

Management scholars have contributed to the knowledge recombination literature by examining cross-boundary knowledge flows at both the firm and individual levels of analysis. One early research stream emphasized the importance of technological “gatekeepers,” individuals that facilitate inter-organizational communication and cooperation by spanning organizational and sub-unit boundaries (e.g., Allen, 1977; Tushman and Scanlan, 1981). A strategic version of the boundary-spanning concept was put forth by Burt (1992). Individuals who connect structural holes in a network will be in an economically privileged position because the unconnected parties would have few or no alternative routes to link themselves outside of the boundary spanner. In addition, stored knowledge may reside at the individual, team or organizational levels such that re-use in different settings could be value-enhancing.<sup>5</sup>

The phenomenon of using knowledge and ideas from one context to solve problems in another has been examined ethnographically in the context of a product development firm, IDEO, as it sought to solve different problems for various clients, often drawing on an inventory of accumulated knowledge to customize a solution for a given situation (Hargadon and Sutton, 1997). More generally, one could imagine re-application of prior problem-solving strategies into new settings, such as legal research, management consulting, or even composing computer software code. While verbatim importing likely represents a minority of cases in these settings, one could imagine “cutting and pasting” applicable modules to new settings for productive use.<sup>6</sup>

Similarly, in biotechnology, there are anecdotal examples of knowledge bridging. A former colleague of Alejandro Zaffaroni (a founder of 7 biotechnology firms) states: “...he is reading and thinking very widely. He is totally unafraid of any new technology in any area of human creativity. He has wonderful contacts with people in many different areas, so he sees the bridges between otherwise disparate fields” (as quoted in Burt, 2004).

---

system that exchanges intellectual property protection for invention disclosure and the academic promotion system based on publishing papers in publicly-available professional journals. Such knowledge spillovers help explain the significant inter-industry knowledge use patterns documented by Scherer (1982). These knowledge flows from one industry to another have been recognized as a productive source of economic growth.

<sup>5</sup> It is interesting to speculate whether boundary spanning knowledge must be in one person or within a team. Brokerage opportunities must be the result of connecting spaces in which there has been limited amount of connection (competition in connecting those spaces would dissipate rents). Most interconnections do not hold value (or are costly to connect relative to expected benefits). Knowledge specialization of disparate areas can lead to interconnection, but making the investment will not be worthwhile for the average person. Of course having fragmented knowledge across the organization can also be costly (Henderson and Clark, 1990). Most of the prior literature has focused on diversity of knowledge at the team rather than the individual level.

<sup>6</sup> Management scholars have also identified a process of technology melding, taking technologies from two different domains and creating a novel application (Kodama, 1992; Levinthal, 1998).

Consider the case of Kary Mullis and the invention of polymerase chain reaction (PCR) technology. Cetus Corporation hired Mullis in 1979 to synthesize oligonucleotide probes.<sup>7</sup> By 1983, however, oligonucleotide synthesis was becoming reliably automated, and Mullis' job as a chemist was facing obsolescence. With more time on his hands, Mullis began "puttering around" with oligonucleotides and became interested in ways to easily detect single base pair changes (against a known sequence) in DNA. Since a genetic mutation may indicate the presence or the potential for a disease, Mullis' search was motivated primarily by finding a potential diagnostic application (Mullis, 1990). Planning this experiment led Mullis to the invention of PCR in the spring of 1983 while he was driving to his cabin in California. During this drive, Mullis came up with the breakthrough idea that using two oligonucleotide primers working in opposite directions on each strand of denatured DNA could create instructions to continually "amplify," or replicate, specific DNA targets (Yoffe, 1994). Mullis had been spending a lot of time writing computer programs and recognized the power of reiterated loops; he envisioned PCR to be such a loop. When he got back to Cetus, Mullis spent three months experimenting with different buffer stocks, reaction temperatures, and reagent concentrations before running his first successful PCR experiment in December 1983. Mullis won the 1993 Nobel Prize in chemistry for his PCR invention.

While Mullis relied more on his in-depth knowledge of chemistry in relation to his knowledge of computer science for the PCR invention, importing ideas and concepts from across academic fields appeared important in his discovery. In other instances, deep knowledge of two or more disparate academic fields was important for the innovative process. Consider George Church, a professor specializing in bioinformatics at Harvard Medical School: "Church's ability to bring together information technology and experimental genetics has made him a *'force majeure* in science,' according to Philip Leder, Andrus professor of genetics and head of the genetics department at Harvard Medical School. Far from being 'just a computer geek,' Leder says, Church is a polymath who 'has terrific ideas that nobody else would think of putting together, because of the many disciplines he has mastered'" (Thomas, 2004).

---

<sup>7</sup> An oligonucleotide is a short chain of specifically-sequenced nucleotide bases. The oligonucleotide can bind specifically with a string of complementary nucleotide bases in single-stranded DNA, and when radioactively labeled, engineered oligonucleotides can serve as probes for detecting whether a sample of DNA contains a particular gene or nucleotide sequence.

Apart from highlighting the importance of knowledge bridging, these anecdotes from the biotechnology industry also suggest an interesting possibility: knowledge bridging may not just lead to good ideas, but also those that are seminal, and with many wide-ranging applications. In this paper, we show an empirical relationship between knowledge bridging and the creation of inventions that are general purpose in nature. Economists have investigated the idea that knowledge can flow across industrial boundaries through general purpose technologies (GPTs), defined as “key functional components embodied in hardware that can be applied as elements or modular units of the engineering designs developed for a wide variety of specific operations or purposes” (David, 1990: 355). While GPTs such as machine tools, electricity, and semiconductors have been investigated through extended historical accounts (e.g., Rosenberg, 1963) and in theoretical treatments (e.g., Bresnahan and Trajtenberg, 1995), empirical studies at lower levels of aggregation are still rare.<sup>8</sup>

To discussion in this section points to *three predicted mechanisms of overcoming local technological search: through strategic alliances, labor market mobility of scientists, and knowledge bridging. We expect these mechanisms to result in higher-quality inventions, both in terms of their direct economic value as well as in their potential to be general-purpose technologies with a broad range of applications.*

### *C. What Affects Knowledge Bridging Use?*

If our hypothesis that knowledge bridging based technological search is correct, one naturally would like to know what affects firms’ use of this mechanism. The empirical literature that addresses triggers of change in technological search is surprisingly limited. Cyert and March (1963) discuss the effects of organizational failure as a proximate cause of organizational search, and Singh’s (1986) empirical results corroborate the notion that risk taking will depend on organizational performance. Consequently, organizational search should depend on feedback from product development efforts and other signals from with the market: positive feedback should reinforce exploitative search, while negative feedback should trigger exploratory technological search.

It is well recognized that for start-up firms, resource constraints, such as access to financial and human capital, often limit business development. During periods of abundant

---

<sup>8</sup> See the recent paper by Hall and Trajtenberg (2005) for an exception.

financial resources, such as “hot” public markets, firms within the favored sector(s) may enjoy more organizational slack and surplus resources, and can therefore afford to engage in more exploratory behavior. In contrast, when such resources are relatively scarce, organizational slack diminishes and these organizations will be more reluctant to engage in exploratory behavior.

Finally, there may be entry order effects associated with competitive dynamics within an industry. As time passes, we predict that the knowledge base of entrepreneurial teams will need to be more specialized and/or span more boundaries in order to reach the competitive frontier. This results because the rent-generating potential of intellectual capital dissipates as knowledge diffuses and is commoditized, thus leading to potential opportunity exhaustion. Knowing this, potential new entrants will evaluate the expected value of their innovations (as compared to their opportunity costs) before making their entry decision. There is an interdependence of own and competitors’ knowledge base in shaping search direction (Stuart and Podolny, 1996). This effort at avoiding head-to-head competition with industry incumbents represents a second mechanism, reinforcing the opportunity exhaustion effects.

The discussion in this section yields *three predictions of events that will increase a firm’s use of knowledge bridging-oriented technological search: (1) negative feedback in a firm’s product development progress, (2) munificence in the external funding environment, and (3) later entry into the industry.*

### **III. Data and Method**

To test these hypotheses, we require a sample of firms that were founded to exploit a given technological opportunity. While we understand that firms will start from different initial conditions, constructing a sample that is relatively uniform in the basic technology upon which they are capitalizing is important. This allows us to observe differences in initial conditions along with the subsequent evolutionary development patterns for firms that are homogenous with respect to initial technology (while recognizing that firms still differ at inception due to, among other things, founder capabilities).

The circumstances surrounding the commercialization of recombinant DNA following its discovery in 1973 by University of California-San Francisco scientist Herb Boyer and Stanford scientist Stan Cohen provides a fortuitous empirical setting. The history of the landmark Cohen Boyer patent is recounted elsewhere, and so we will not duplicate those efforts here (see e.g.,

Reimers, 1987; Hughes, 2001; Galambos and Sturchio, 1998). Instead, we merely note that Stanford University conducted an open non-exclusive licensing program of the patent, and so we are able to observe with great precision de novo firms founded to commercialize recombinant DNA technology (users of the technology that did not participate in the licensing program would be infringing the patent and subject to litigation).<sup>9</sup> Because we are interested in studying the evolutionary development of knowledge bridging search, we track new ventures established to commercialize the Cohen-Boyer patent. Aside from the clear scientific importance of the Cohen-Boyer innovation, the patent was also clearly important commercially: over its lifetime, the patent yielded approximately \$200M in licensing revenues, which implies product sales based on the innovation of some \$40B.<sup>10</sup>

#### *A. Method*

Researchers in the literature have used two types of measures of technological search, both of which are based on patent citation data. Neither of the approaches, however, suits our purpose of capturing knowledge bridging-based search. The first type of exploratory search measure is used by Rosenkopf and Nerkar (2001). In the context of optical disk drive firms these authors use citations to non-disk patents as a measure of technological exploration and non-self citations as a measure of organizational exploration. While these are useful measures of exploration, we wish to develop measures that are more specific to the knowledge bridging concept. A second class of measures examine the degree to which a firm overlaps with another firm's technical knowledge (Mowery, Oxley and Silverman, 1996), or the extent it overlaps with a common stock of scientific knowledge in a network-centric sense (Stuart and Podolny, 1996). Both of these approaches use measures that fundamentally rely on knowledge overlap across firms. In contrast, our conceptualization of knowledge bridging search emphasizes the overlap

---

<sup>9</sup> The Cohen Boyer invention was covered by three patents, with the most important being a process patent, U.S. patent number 4,237,224, entitled "Process for Producing Biologically Functional Molecular Chimeras." This patent, which became the backbone of the Stanford Technology Licensing Office's licensing efforts of recombinant DNA, was issued on December 2, 1980, and expired 17 years later, in 1997. Stanford offered licenses to the patent for a modest fee (\$10,000 annual payments, with 0.5% royalty rates on end products).

<sup>10</sup> As well, between 1980 and 2000, the patent was cited 235 times by other patents, while the average patent of this vintage in this technology class was cited 9.64 times (Jaffe and Trajtenberg, 2002). Despite the economic value of this patent, which yielded such products as recombinant growth hormone and recombinant insulin, its legal validity was not subsequently challenged.

between the technical domain upon which a firm relies upon and the technical area in which it produces new knowledge..

We therefore elect to develop a new measure, which is based on the patent classes of citing and cited patents, to better capture knowledge bridging search processes. Following Jaffe (1986), we believe that using patent class data is a good way of capturing the concept of technological position. Comparing patent classes of cited patents (a measure of technological knowledge inputs) with classes of focal patents (a measure of knowledge outputs) allows us to construct a measure of technological knowledge bridging.

The first step in our method is to identify start-up firms that entered as a result of opportunities afforded by a biotechnology innovation. To identify new venture entrants commercializing recombinant DNA, we rely on the Stanford University Technology Transfer Office's records of licensees to the technology. We include firms into this sample if: (1) they are de novo firms (as opposed to established pharmaceutical firms), and (2) licensed the Cohen Boyer patents at the time of founding, or within a time window of two years after their founding. This process yielded a total of 19 firms. Using U.S. patent database records, we then identified the history of patent applications submitted by each of the firms. We conduct our analysis at both the patent and the firm-year level of analysis. The following section describes the variables and empirical tests used in the analyses. The summary statistics and descriptions of all variables are found in Table 1.

### *B. Patent level analyses*

To assess the potential impact of knowledge bridging, we investigate two patent-level dependent variables: forward patent citations and patent "generality."<sup>11</sup> The forward citation measure is the count of future patents that reference the focal one, evaluated at five years post focal patent issue (mean = 3.3 in the dataset), and through the year 2004 (mean = 6.6). The

---

<sup>11</sup> One reason we investigate impact at the patent level is because the biotechnology industry (and biopharmaceuticals more generally), is notorious for the length of product (drug) development. Evaluators (e.g., venture capitalists) use patent information as a signal of value, and patent-based measures have been validated in a number of academic studies (e.g., Jaffe and Trajtenberg, 2002).

*PATENT GENERALITY* (Henderson, Jaffe, and Trajtenberg, 1998) measures the diversity of patent classes of the forward patent citations (mean = 0.54).<sup>12</sup>

The right hand side variables include measures of *PATENT KNOWLEDGE BRIDGING*, patent characteristic controls, patent application year fixed effects, and start-up fixed effects. The key variable of interest is *PATENT KNOWLEDGE BRIDGING*, which we define as [1 – (share of cited patents that are in the same primary class as the focal patent)] (mean = 0.52). A patent usually cites prior knowledge within the same area, a behavior that can be considered exploitation of a firm’s current knowledge. The extent to which inventors cite patents in different areas indicates the degree of knowledge bridging. High measures of *PATENT KNOWLEDGE BRIDGING* imply substantial use of scientific knowledge originating from outside the focal patent area.

We also use a number of control variables in the analysis. *PATENT ORIGINALITY* (mean = 0.53) is defined similarly to *PATENT GENERALITY*, using backward citations instead (Henderson, Jaffe, and Trajtenberg, 1998). *PATENT SCOPE* (mean = 2.71) is the number of international patent classes assigned to the focal patent, which has also been correlated with economic value (Lerner, 1994). Patent cohort effects, as embodied in a set of indicators of *PATENT APPLICATION YEAR*, are important controls when explaining forward citation rates. Finally, the variable *INVENTOR PATENT EXPERIENCE AT OTHER FIRMS* is defined as the number of patents issued to a focal patent’s inventors when employed by *other* organizations as of the application date of the focal patent (mean = 6.6). The measure is meant to capture the mobility of scientific knowledge through human capital.

### C. Firm-year analyses

We also investigate the antecedents to knowledge bridging based search. Before describing the variables used in that analysis, let us first discuss a patent classification issue. For the firm-level analysis, we categorized the current US primary classes of these patent applications (and the primary classes of the patents cited) using a modified version of the

---

<sup>12</sup> The measure is computed using the formula:  $G_i = \left[ 1 - \sum_{j=1}^J \left( \frac{N_{ij}}{N_i} \right)^2 \right] \left( \frac{N_i}{N_i - 1} \right)$ , where  $i$  indexes the patent,  $j$  indexes patent classes, and  $N$  represents counts of forward citations. The expression outside of the square brackets adjusts for bias associated with small numbers of forward patent counts (Hall, 2005).

categorization proposed by Jaffe and Trajtenberg (2002, Appendix 1).<sup>13</sup> For our purposes, the Jaffe and Trajtenberg (2002) classes are a good starting point because it collapses an almost unwieldy number of three digit US patent classes into industry family groups. While this classification meets our needs for industry groups outside the biotechnology and drug discovery areas, the classification is too coarse for the two focal areas in which most of our sampled firms are patenting. Specifically, Jaffe and Trajtenberg's class 33, biotechnology, contains a high fraction of the primary patent classes in our sample. We use the U.S. Patent and Trademark Office's Manual of Classification scheme (at the six digit level) to sub-classify biotechnology patent classes. This scheme therefore contains greater classification detail of biotechnology patent classes and less for non-biotechnology classes. It is this patent class categorization scheme that is used to construct the dependent variable at the firm-year level of analysis, *FIRM KNOWLEDGE OVERLAP*.

At the patent level, we measure knowledge overlap as the extent to which a focal patent cites other patents in the same technological class. However, most firms are likely to obtain more than one patent in a given year, and possibly in multiple technology areas. As such, for the firm-year analysis, we need to construct a multi-dimensional measure of knowledge overlap that takes into consideration the entire set of patents and patent classes cited each year. The first step is to obtain a vector of focal patents' current classes,  $\mathbf{a}$ , for a given firm-year. Similarly, obtain a vector of patent classes for the patents cited by the focal patent,  $\mathbf{b}$ . Next, construct the vector dot product,  $\mathbf{a} \cdot \mathbf{b}$ , which gives us the degree to which the vectors  $\mathbf{a}$  and  $\mathbf{b}$  are aligned. The variable *FIRM KNOWLEDGE OVERLAP* is measured by calculating the following

expression:  $\left[ -\arccosine \left( \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}| |\mathbf{b}|} \right) \right] \left[ \frac{180}{\pi} \right]$ . This measures the angle, in degrees, between the two

vectors. A higher value of *FIRM KNOWLEDGE OVERLAP* (mean = -40.5 degrees) corresponds to greater similarity between the patent classes of the focal patent and those it cites. Intuitively, the item we want to measure is the angular separation between the technological classes of a focal patent and the technological classes of the patents it cites. The actual magnitudes do not matter and are normalized in the denominator of the expression. At one extreme, both the citing

---

<sup>13</sup> U.S. patents are classified at both the three digit and six digit levels (with the latter being more detailed than the former). See Lerner (1994) for more institutional details about U.S. patent classifications. Since patent classes are sometimes reclassified over time, it is important to use patent classes in the sample at a given time, which in our case is the 2004 classification.

and cited patent classes could be parallel, meaning that the knowledge produced is directly in line with the knowledge cited (in this case, *FIRM KNOWLEDGE OVERLAP*=0). At the other extreme, the citing and cited patent classes could be orthogonal, so that the knowledge produced is in a completely different direction than the knowledge cited (in this case, *FIRM KNOWLEDGE OVERLAP* would be  $-90$  degrees).

We employ three sets of independent variables in the analysis: a measure of initial conditions, firm-specific measures, and munificence of the external funding environment. Each is discussed in turn.

Our review of the prior literature suggests that taking account of initial search conditions is important. We adopt Cockburn, Henderson, and Stern's (2000) philosophy of documenting variables that correlate with organizational strategy while taking into account initial conditions, while recognizing the potential dissipation of imprinting effects of initial conditions over time. We do this by constructing a variable, *FOUNDER OVERLAP*, that calculates a dot product (as before) comparing the degree to which the patent classes of each firm's founders' patents across time overlaps with those granted to the firm in a given year (mean =  $-56$  degrees). We use founders' patents across time due to the relatively small number of such patents. Using this definition is not very restrictive: we observe that in our data set, founders tend to patent in a stable set of classes throughout their careers.

A second group of variables examine firm effects, which we break down into independent variables associated with product development, and control variables associated with firm size and resources. Since the overwhelming majority of the firms in the dataset innovate in the human therapeutics space, we collect information about each firm's lagged annual count of Federal Drug Administration (FDA) drug approval stages. In the pre-clinical stage, animal studies are conducted in an effort to determine if there are harmful effects of the potential drug. Experimental results from this phase are submitted to the FDA in an Investigational New Drug (IND) application. The FDA reviews this application to determine whether human study is warranted. If so, clinical trials can be conducted on human subjects—which takes place in three phases. Phase I involves a small number of healthy people in an effort to determine appropriate dosage. With each passing phase, more human subjects are used under increasingly controlled trials. Conditional on passing through phase III trials, firms apply for

FDA drug approval by submitting a New Drug Approval (NDA) application.<sup>14</sup> The variable *CUMULATIVE FDA DRUG DEVELOPMENT APPROVALS IN T-2* tallies the cumulative number of FDA drug approvals the firm received t-2 years ago (mean = 2.2). At any time during the FDA drug approval process, firms can decide to terminate projects. We interpret these terminations as negative signals to the firm, which may be due to various factors, including disappointing clinical trials and powerful competition in the product segment. The variable *CUMULATIVE DRUG DEVELOPMENT TERMINATIONS IN T-2* counts the stock of such terminations two years from the observation period (mean = 3.03). Data to code these variables are from the Recombinant Capital database.

To round out the firm-level variables, we employ two control variables that proxy for firm size and resources. The variable, *EMPLOYEES IN T-2* (mean = 720) is a firm's headcount in year t-2. The variable *REVENUES IN T-2*<sup>15</sup>, measures gross revenues of a firm t-2 years in the past (mean = \$171.4M), is not used in the empirical analysis because it is almost perfectly collinear with *EMPLOYEES IN T-2*. Data to code these firm size variables is taken from the CorpTech directory of technology firms and from firms' public filings, such as 10-K and annual report to shareholders documents. The variable *CUMULATIVE STRATEGIC ALLIANCES IN T-2* is a count of the total number of strategic alliances a firm has undertaken up until two years before the observation year (mean = 17.5). The alliance data is obtained from Recombinant Capital. We use this as a control variable because researchers have identified use of alliances as a means to overcome local search (e.g., Mowery, Oxley, and Silverman (1996)), and so may reflect organizational efforts to change their strategic direction.

Finally, because there might be entry order effects, we code the order in which firms entered the industry within the innovation wave, based on founding year and month data, and label this variable *ENTRY ORDER*. The earliest entrant is assigned a value of one, and later entrants are assigned increasing numbers. If two or more firms were founded in the same month and year, they are assigned the same entry order number (there is one instance in which we were not able to establish entry sequence for three firms).

---

<sup>14</sup> More details of the drug approval process can be found at the FDA's website, [www.fda.gov](http://www.fda.gov).

<sup>15</sup> The time lags associated with this and other independent variables are meant to capture the idea that the current observation of firms' technological search will depend on events that took place in the past. In the empirical analysis, we check for sensitivity to the two year lag used in the reported specifications, and find that the results are generally robust for a one to three year lag, though because of the limited number of observations in this dataset, we report two year lags.

Our final independent variable examines potential effects associated with the munificence of the external funding environment. Because the drug discovery process is long and involved, most biotechnology firms are not able to generate working capital based on sales revenue in the short run. As a result, many (if not most) biotechnology firms are dependent on pharmaceutical firm alliances, venture capital (or other private) funding, and public offerings for financial capital. To examine whether firms' technological search is correlated with the availability of funding for biotechnology investments changes over time, we use Josh Lerner's index of external funding available to biotechnology firms. This measure, described in Lerner (1994), tracks a representative basket of biotechnology firms with respect to availability of outside funding, and for 1978-1999, the index ranges from a low of 1 (the starting position) to a high of 5.9. *HOT BIOTECH FUNDING ENVIRONMENT IN T-2* (mean = 3.24) reflects the funding environment in time t-2.

Notice that *HOT BIOTECH FUNDING ENVIRONMENT IN T-2* is just one component of organizational slack resources—the overall munificence and availability of financial resources over time. We would ideally like to incorporate firm-level measures of organizational slack—but measurement of this concept longitudinally, particularly during periods in which firms are privately-held, is challenging (see the approaches and discussion contained in Bourgeois, 1981; Sharfman et al., 1988; and Nohria and Gulati, 1996).

#### **IV. Empirical Results**

We present the empirical results in three stages. First, we examine multivariable results at the patent level of analysis. Second, we describe knowledge input - knowledge output matrices for firms in the sample to illustrate knowledge bridging over time. Finally, we step through our multivariate results of the correlates of *FIRM KNOWLEDGE OVERLAP*, our firm-level measure of knowledge bridging-based technological search.

##### *A. Patent level regression results*

Table 2 initiates the patent level analysis. In the first three columns, we examine correlates of *PATENT FORWARD CITATIONS* within five years of patent issue while in the last column we analyze these forward citations through 2004 (negative binomial count regressions are used throughout). In each of the specifications, we include *PATENT APPLICATION YEAR*

and *FIRM* fixed effects. The first two columns show that *PATENT KNOWLEDGE BRIDGING* is positive and significant at the 1% level (with an associated incidence rate ratio (IRR) of 1.19), even controlling for *PATENT SCOPE* (which itself is positive and statistically significant, with an IRR of 1.04). Notice that the variable *INVENTOR PATENT EXPERIENCE AT OTHER FIRMS*, the measure of exploratory search via scientific labor market mobility, is not statistically different from zero. One might argue that an alternative definition of the concept of knowledge bridging is captured by the *PATENT ORIGINALITY* variable, which indicates the diversity of primary patent classes a focal patent cites. While this measure is 58% pair-wise (and significantly) correlated with *PATENT KNOWLEDGE BRIDGING*, we believe that the concepts behind the measures are distinct. *PATENT ORIGINALITY* does not make the comparison between inputs (backward citation patent classes) and outputs (focal patent class). Nevertheless, specification (2-3) takes *PATENT ORIGINALITY* as the measure of knowledge bridging, and finds it to be statistically (at the 5% level) and economically significant (with an IRR of 1.14). The final column of Table 2 replicates the logic of specification (2-2), but uses as the dependent variable the logged count of forward citations through 2004. Due to the censoring of forward citations, it is important to include the *PATENT APPLICATION YEAR* fixed effects.<sup>16</sup> The results here are consistent with the prior ones. In all specifications in this table, excluding self-citations from the forward citation counts substantially strengthens the *PATENT KNOWLEDGE BRIDGING* results.

Table 3 reports OLS regressions with patent generality as the dependent variable. Again, in all the specifications, we include *PATENT APPLICATION YEAR* and *FIRM* fixed effects. In addition, the standard errors are robust and take into account clustering by firm. We replicate the logic of the prior table by examining in turn *PATENT KNOWLEDGE BRIDGING* and *PATENT ORIGINALITY* as measures of knowledge bridging search behavior. We also add *FORWARD CITATIONS THROUGH 2004* to the second and fourth specifications (this variable is estimated with a positive and significant coefficient, but with a small economic effect). The main result of the table, however, is that in each case, the measure of knowledge bridging is positive and

---

<sup>16</sup> An alternate approach is to deflate the forward citations by the average value for its scientific field-year cohort as a fixed effect, as indicated in Jaffe and Trajtenberg (2002). Because we do not use the NBER dataset for our patent data (the NUS patent project allows us to access more recent patent data), we do not use these deflators in our analysis.

significant at the 1% level. Taken together, these results suggest that this type of exploratory search is associated with the production of patents that are of a high quality and generality.

### *B. Input-output knowledge matrices*

We now turn our attention to understanding this phenomenon at the firm level. As a first step, we illustrate different typologies of knowledge bridging from our data set. In considering types of bridging, it is useful to compare them against a base case of no bridging. In this organizational search mode, firms are drawing primarily on search within the domain in which they hope to innovate. We contrast this situation with taking knowledge from one area and applying it to another through problem solving strategies employing either analogy or recombination.

Figures 2a through 2d illustrate knowledge bridging at the firm level. Each row shows the primary classes of a firm's patents, while each column indicates the primary classes of the patents it cites. Figure 2a shows the results for Biogen, one of the earliest firms founded to commercialize recombinant DNA technology. The patent class shown is that of the modified Jaffe-Trajtenberg scheme, which is also used in the regression analysis below. Most of the data points are along the diagonal. This means that Biogen's patents in a given technological class mainly cite patents that are also from the same class. In contrast, Figure 2b shows the results for Visys, which produces patents in a much narrower range of classes than Biogen (primarily 33d). However these patents make citations to a broad range of other patent classes, including classes 14, 15, 19, 33e, 33o. This represents a case of bridging from those other classes into class 33d. Figure 2c shows yet another example, that of Therion. It produces patents in a number of different classes, but most of them cite patents that are in class 33j. This illustrates a case of bridging from class 33j to other technological areas. Finally, Figure 2d shows the patenting behavior of New England Biolabs. The firm mainly produces patents in class 33i, but these cite patents in class 33d more often than those in class 33i itself. Interestingly, the firm produces only a small number of patents in class 33d, even though 33d is so heavily cited by patents in class 33i. A likely interpretation of this pattern is that New England Biolabs is taking knowledge from class 33d and transferring it into patentable inventions within class 33i.

A second set of charts explores the differences across firms in the knowledge they cite and whether this changes over time. Figure 3a shows the number of citations made by New

England Biolabs' patents to various primary patent classes each year. The original US patent class is used instead of the Jaffe-Trajtenberg ones, so as to show finer-grained details. From this figure, it is clear that the company relies mainly on knowledge in US patent class 435, and that the pattern persists throughout the 15 year period shown. By comparison, Biogen cites a broader set of technological classes (Figure 3b), including US patent classes 424, 435, 514 and 530. This pattern is also fairly stable across time. Extending this analysis to other firms in the sample reveals a large variation in the observed plots: there are large inter-firm differences that persist over time. This suggests that a firm's initial condition leaves a strong mark on the future search strategies pursued by the firm.

A third observation in the descriptive data is illustrated in Figures 4a and 4b. For each year, the plots show the number of patents a firm receives in each primary technological class. The trend over time indicates the extent to which a firm enters new technological fields. For some firms such as New England Biolabs, the pattern is stable across time (Figure 4a). This suggests that throughout its history, the firm has primarily been producing knowledge in roughly the same technological space. However, other firms such as Repligen show marked changes in direction (Figure 4b). Early in its history, Repligen produced patents mainly in technological classes 14, 31e and 33e. At a later stage, it stopped patenting in these areas and began patenting in other areas instead (classes 15, 31i and 31xa). In this case, we know that the change in strategy was driven by turbulent organizational changes, as documented in a Harvard Business School case (Lerner and Kane, 1998). However, the differences we observe across firms suggest the need for a more systematic analysis of what triggers changes in each firm's search direction. Consequently, we now turn to a regression analysis examining correlates of knowledge bridging at the firm-year level.

### *C. Firm-year level regression results*

Table 4 presents tobit regressions of *FIRM KNOWLEDGE OVERLAP* with censoring points at negative 90 and 0 degrees. In a first specification, we examine the role of *FOUNDER OVERLAP*, our measure of imprinting, while controlling for *EMPLOYEES IN T-2* and *CUMULATIVE STRATEGIC ALLIANCES IN T-2*. The estimated coefficient on *FOUNDER OVERLAP* is positive and statistically significant at the 1% level, suggesting that founders' knowledge is importantly imprinted on the future direction of the firm. In the second column, we

add the measure of lagged hot funding environment, *HOT BIOTECH FUNDING ENVIRONMENT IN T-2*. This estimate is negative and significant at the 5% level, which is consistent with the proposition that more munificent funding environments are associated with organizational slack, which in turn is associated with more exploratory technological search behavior.

To investigate the effects of product development, we add in the third column of the table two measures of product development in this empirical setting: *CUMULATIVE FDA DRUG APPROVALS IN T-2* and *CUMULATIVE DRUG TERMINATIONS IN T-2*. While these two estimates are not statistically significant, they are estimated with signs in the expected direction: approvals tend to reinforce a firm's search behavior, while terminations are associated with wider technological search behavior.

In the final column of Table 1, we add *ENTRY ORDER*. Interestingly, the estimate is statistically significant (at the 5% level), but carries the opposite sign of what we expected. The result suggests that later entrants are more uniform (rather than divergent, as predicted) in their search patterns. In an effort to better understand the possibility of niche player entry driving this result, we coded firms according to the scope of their product development efforts. Unfortunately, this analysis was not illuminating, both because of the somewhat subjective manner in which we coded niche players, and perhaps more importantly, because of the limited number of firms in the dataset. A more subtle possibility also exists (though it is hard to examine empirically): knowledge bridging may involve importing or exporting small amounts of information as the basis for recombination, and so later entrants may be recombining small amounts of external knowledge (while still relying intensively on the dominant body of knowledge relevant to a field). Nevertheless this phenomenon merits further investigation, most appropriately in a larger dataset.

The main result from Table 4 is the enormous impact of founder "imprinting" in reinforcing future technological search patterns, as well as the importance of funding environment munificence in spurring exploratory search behavior. To check whether fixed firm effects overturn these results, we present Table 5, which replicates the logic of the prior table (with the exception that *ENTRY ORDER* cannot be examined using fixed effects regressions). We might believe that these specifications are important to the extent that firms differ in "style"

of technological search in ways we as researchers are not able to observe and/or measure.<sup>17</sup> Reassuringly, the results are indeed robust to these fixed effects specifications. Furthermore, across the specifications in both empirical tables, the estimated coefficients of *FOUNDER OVERLAP* and *HOT BIOTECH FUNDING ENVIRONMENT IN T-2* are stable and economically significant: a one standard deviation increase in *FOUNDER OVERLAP* is associated with a 3.5 to 5 degree increase in *FIRM KNOWLEDGE OVERLAP*. A one standard deviation increase in *HOT BIOTECH FUNDING ENVIRONMENT IN T-2* is associated with a 2.9 to 4.1 degree decrease in *FIRM KNOWLEDGE OVERLAP*.

## V. Discussion

We investigate a type of technological boundary spanning-oriented search in which firms apply knowledge from one technical domain to innovate in another, a phenomenon we term knowledge bridging. We develop and explore measures capturing the concept of knowledge bridging-oriented exploratory search. In a controlled empirical setting of new firms founded to commercialize recombinant DNA technology, we find that their patents exhibiting a high degree of knowledge bridging are associated with a higher forward-citation rate and a higher generality score. In addition, we find variation in the degree in which firms are bridging knowledge. The strongest predictors of knowledge bridging at the firm level are initial search trajectories and munificence in the external funding environment. As well, we found weaker statistical support for product development feedback. Finally, there is some evidence that later firm entry order into commercialization of recombinant DNA technology was associated with narrower knowledge bridging-oriented search (though due to the small sample size, we could not rule out some possible alternate explanations).

Two interpretational issues merit discussion. First, to what extent does *FIRM KNOWLEDGE OVERLAP*, capture (or not capture) the concept of knowledge bridging at the firm level? We can make an analogy to academic journal articles: those that cite work from a variety of fields are likely to have borrowed, recombined (and possibly extended) external

---

<sup>17</sup> Despite censoring points associated with the dependent variable, we chose to report fixed effects OLS regressions. A fixed effects tobit model is not available because the fixed effect cannot be conditioned out of the likelihood function. While a random effects tobit model is available, we do not feel that the random effects tobit is appropriate here (the results of that model accentuate the funding environment munificence results, essentially doubling the estimated coefficients at conventional levels of statistical significance ( $p$  values of .01), while weakening or eliminating the founder overlap result.

knowledge. As well, because patent citations are used as a legal device to circumscribe patent scope, such citations-based data may be useful data in analyzing the notion of knowledge bridging. A second interpretational issue involves the use of patent citation data—namely that patent examiner-imposed citations may be an important phenomenon (in light of Alcacer and Gittleman, 2003). If true, then our calculation of *FIRM KNOWLEDGE OVERLAP* may not accurately reflect firm search behavior. Ultimately, however, we are mainly concerned with knowledge use, so as long as each patent does depend on the other patents that it cites for knowledge, it does not matter who was responsible for citation being added to the patent.

Finally, we believe that the work described here can and should be extended. For example, in addition to examining more data, perhaps in other industrial contexts, it would be interesting to use the knowledge bridging measure to evaluate how some established firms are using acquisition of innovative start-ups to potentially alter (or not) their strategic direction. As well, we believe that developing the concept and measurement of knowledge uniqueness (which would affect, among other things, relative bargaining position) would be highly complementary to the efforts developed here.

## References

- Alcacer, J. and M. Gittelman (2003). "How Do I Know What You Know? Patent Examiners and the Generation of Patent Citations," working paper, New York University.
- Boeker, W. (1989). "Strategic Change: The Effects of Founding and History," *Academy of Management Journal*, 32: 489-515.
- Bourgeois, L.J. (1981). "On the Measurement of Organizational Slack," *Academy of Management Review*, 6: 29-39.
- Bresnahan, T.F. and M. Trajtenberg (1995). "General Purpose Technologies: Engines of Growth?" *Journal of Econometrics*, 65: 83-108.
- Burt, R. (1992). Structural Holes: The Social Structure of Competition. Harvard University Press, Cambridge, MA.
- Burt, R. (2004). "Structural Holes and Good Ideas," *American Journal of Sociology*, 110: 349-399.
- Cockburn, I., R. Henderson, and S. Stern (2000). "Untangling the Origins of Competitive Advantage," *Strategic Management Journal*, 21: 1123-1145.
- David, P.A. (1990). "The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox," *American Economic Review Papers and Proceedings*, 80: 355-361.
- DiMasi, J.A. (2000). "New Drug Innovation and Pharmaceutical Industry Structure: Trends in the Output of Pharmaceutical Firms," *Drug Information Journal* 34: 1169-1194.
- Fleming (2001). "Recombinant Uncertainty in Technological Search," *Management Science*, 47: 117-132.
- Fleming, L. and O. Sorenson (2004). "Science as a Map in Technological Search," *Strategic Management Journal*, 25: 909-928.
- Gans, J.S., D.H. Hsu and S. Stern (2002). "When Does Start-up Innovation Spur the Gale of Creative Destruction?" *RAND Journal of Economics*, 33: 571-586.
- Galambos, L. and J.L. Sturchio (1998). "Pharmaceutical Firms and the Transition to Biotechnology: A Study in Strategic Innovation," *Business History Review*, 72: 250-278.
- Griliches, Z. (1992). "The Search for R&D Spillovers," *The Scandinavian Journal of Economics*, 94: Supplement 29-47.
- Hall, B. (2005). "A Note on the Bias in Herfindahl-type Measures Based on Count Data," unpublished working paper, UC Berkeley.

- Hall, B. and M. Trajtenberg (2005). "Uncovering GPTs Using Patent Data," in Antonelli, Foray, Hall, and Steinmueller (eds.), *Festschrift in Honor of Paul A. David*, Edward Elgar.
- Hargadon, A. and R. Sutton (1997). "Technology Brokering and Innovation in a Product Development Firm," *Administrative Science Quarterly*, 42: 716-749.
- Helfat, C.E. (1994). "Evolutionary Trajectories in Petroleum Firm R&D," *Management Science*, 40: 1720-1747.
- Henderson, R. and I. Cockburn (1994). "Measuring Competence? Exploring Firm Effects in Pharmaceutical Research," *Strategic Management Journal*, 15: 63-84.
- Henderson, R., A. Jaffe, M. Trajtenberg, (1998). "Universities as a Source of Commercial Technology: A Detailed Analysis of University Patenting, 1965-1988," *The Review of Economics and Statistics*, 80: 119-127.
- Hughes, S.S. (2001). "Making Dollars out of DNA: The First Major Patent in Biotechnology and the Commercialization of Biotechnology," *Isis*, 92: 541-575.
- Jaffe, A. (1986). "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value," *American Economic Review*, 76: 984-1001.
- Jaffe, A., M. Trajtenberg, and R. Henderson (1993). "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *Quarterly Journal of Economics*, 108: 577-598.
- Kodama, F. (1992). "Technology Fusion and the New R&D," *Harvard Business Review*, 70: 70-78.
- Lerner, J. (1994). "The Importance of Patent Scope: An Empirical Analysis," *RAND Journal of Economics*, 25: 319-333.
- Lerner, J. and D. Kane (1998). "Repligen Corp.: January 2002," Harvard Business School case 9-294-082.
- Levinthal, D.A. (1998). "The Slow Pace of Rapid Technological Change: Gradualism and Punctuation in Technological Change," *Industrial and Corporate Change*, 7(2): 217-247.
- March, J.G. (1990). "Exploration and Exploitation in Organizational Learning," *Organization Science*, 2(1): 71-87.
- March, J.G. and H.A. Simon (1958). *Organizations*. John Wiley, New York.
- Mowery, D.C., J.E. Oxley and B.S. Silverman, "Strategic Alliances and Interfirm Knowledge Transfer," *Strategic Management Journal*, 17: 77-91.

- Mullis, Kary B. (1990). "The Unusual Origin of the Polymerase Chain Reaction," *Scientific American*, April, pp. 56-65.
- Nelson, R.R. and S.G. Winter (1982). An Evolutionary Theory of Economic Growth. Harvard University Press, Cambridge, MA.
- Nohria, N. and R. Gulati (1996). "Is Slack Good or Bad for Innovation?" *Academy of Management Journal*, 39: 1245-1264.
- Olsson, O. and B.S. Frey (2002). "Entrepreneurship as Recombinant Growth," *Small Business Economics*, 19: 69-80.
- Reimers, N. (1987). "Tiger by the Tail," *Chemtech*, 17: 464-471.
- Rosenberg, N. (1963). "Technological Change in the Machine Tool Industry, 1840-1910," *Journal of Economic History*, 23: 414-443.
- Rosenkopf, L. and A. Nerkar (2001). "Beyond Local Search: Boundary-Spanning, Exploration, and Impact in the Optical Disk Industry," *Strategic Management Journal*, 22:287-306.
- Scherer, F.M. (1982). "Inter-industry Technology Flows and Productivity Growth," *The Review of Economics and Statistics*, 64: 627-634.
- Schumpeter, J. (1934). The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest and the Business Cycle. Harvard University Press: Cambridge, MA.
- Singh, J.V. (1986). "Performance, Slack, and Risk Taking in Organizational Decision Making," *Academy of Management Journal*, 29: 562-585.
- Sharfman, M.P., G. Wolf, R.B. Chase, and D.A. Tansik (1988). "Antecedents of Organizational Slack," *Academy of Management Review*, 13: 601-614.
- Stinchcombe, A.L. (1965). "Social Structure and Organizations," in Handbook of Organizations, J. March (ed.), Rand-McNally: Chicago, IL, 142-193.
- Stuart, T.E. and J.M. Podolny (1996). "Local Search and the Evolution of Technological Capabilities," *Strategic Management Journal*, 17: 21-38.
- Thomas, Patricia, "DNA as Data," *Harvard Magazine*, January-February, 2004.
- Tushman, M. (1977). "Special Boundary Roles in the Innovation Process," *Administrative Science Quarterly*, 22: 587-605.
- Tushman, M.L. and T.J. Scanlan (1981). "Boundary Spanning Individuals: Their Role in Information Transfer and Their Antecedents," *Academy of Management Journal*, 24: 289-305.

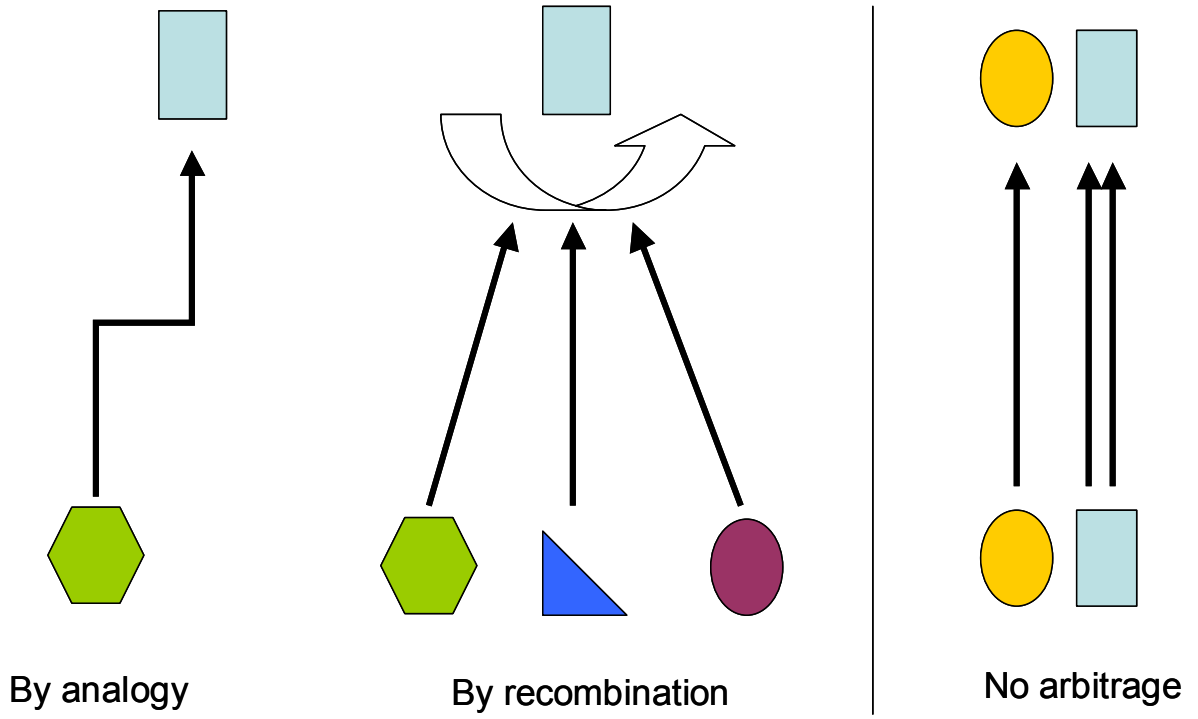
Usher, Abbott Payson (1954). A History of Mechanical Inventions, 2<sup>nd</sup> edition, Harvard University Press.

Weitzman, Martin (1998). "Recombinant Growth," *Quarterly Journal of Economics*, 63: 331-360.

Yoffe, Emily (1994). "Is Kary Mullis God? Nobel Prize Winner's New Life," *Esquire*, July.

Zucker, L., M. Darby and M. Brewer (1998). "Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises," *American Economic Review*, 88: 290-306.

Figure 1: Types of Knowledge Arbitrage



**Figure 2a: Citing versus Cited Patent Classes for Biogen**

Patent classes cited by Biogen

Biogen's  
patents, by  
class

Class	<>	11	14	15	19	21	23	31a	31f	31g	31i	31j	31k	31l	31m	31xa	32	33	33e	33e	33f	33i	33j	33k	33n	33o	33p	39	59
12								1								2													
14			4	2						1						3			3					1					
15	1		20	4					1	6	4					2	1		7	11			2	1					
31a																1													
31f				11					21	2					1	3			1	5					3				
31g			6	7	4		1		1	4						6			2	8				1					
31k			1										1							1	1					1			
31p				3																1									
31xa	4		1	10				1			4	1		1		25			3	11				2		1		2	
33		1		4						1				1				1											
33d	1		1	7	6					5	12					4			9	13	1		6	1	1		1		
33e	7		2	15		2			1	1	6			1		6			8	49	1	2	2	5	2	3		1	
33i	2			1										2					1				6						
33j																						1	1						
33k																				1					1				
33n																				1									
66	1				1																							1	

**Figure 2b: Citing versus Cited Patent Classes for Visys**

Patent classes cited by Visys

Visys'  
patents, by  
class

Class	14	15	19	22	31xa	32	33d	33e	33n	33o	39	43	44	61
14	11						6							
33d	21	4	18		1		41	7	1	3				1
44				1		1					1	10	15	

**Figure 2c: Citing versus Cited Patent Classes for Therion**

Patent classes cited by Therion

Therion's  
patents, by  
class

Class	31k	33d	33e	33f	33j
31i					1
31k	2				
33e					2
33f		1	2	1	2
33j					6
33k					1
33o					1

**Figure 2d: Citing versus Cited Patent Classes for New England Biolabs**

Patent classes cited by NE Biolabs

NE  
Biolabs'  
patents,  
by class

Class	<>	14	15	31xa	33d	33e	33f	33i	33n	33o	69
14		1			3	5		6			
15			2	1	1	2		3	1		
33d	1				3	1		1			
33e		1			5	1		4	1		
33f					1		1	2			
33i		6	2	1	78	49	13	66	13	1	1

**Figure 3a: Primary US Patent Class Cited Each Year by New England Biolabs Patents**

Primary US Class	No of Citations Per Year													
	1986	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
blank		1												
428			1											
435	2		11	35	39	1	26	49	19	6	19	30	19	5
514									1			1		
530									2			2		
536					1		2	4			1			

**Figure 3b: Primary US Patent Class Cited Each Year by Biogen Patents**

Primary US Class	No of Citations Per Year																		
	1980	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1997	1998	1999	2000
blank	1		3	2	1		1	1	2	3	2								
100										1									
340							1												
345											1								
379					1														
424		5		2	5		5		10	3	6	3	1	13	25	2			
435	5	3	1	7	19	6	10	9	10	4	13	10	11	7	51	2		6	8
436					1				1		1	1		1	7	2			
504										1									
514					3	2		1	3	2	7	5	1	1	21		6		
521									1										
530		1	1	1	5		8	4	7	1	6	6	7	7	25				
536							1		4			1			4	4			
560														1					
588										1									
604													1						
623									1	1	1								
800							1												

**Figure 4a: Primary Patent Classes in which New England Biolabs Obtained Patents by Year**

Class	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
14								1							1
15													1		
33d		4	2	2		1	1		1	1		2	1		
33e	1	2		1		1	2		1	1	1	1	1		
33f				1	1		1	1		1			1		
33i		1	2	3	7	9	2	7	4	7	2	5	6	9	1
33l												1			

**Figure 4b: Primary Patent Classes in which Repligen Obtained Patents by Year**

Class	1982	1983	1984	1985	1986	1987	1988	1989	1990	1992	1993	1994	1995	1998	1999
blank										1					
11					1	1		1							
14	2	1		1											
15			1			2	1	2	2	1		1	1		
31e		2		1	1		2	1							
31i													2		
31xa			1	1					2	1	1	1		1	1
33d										1					
33e	2		3		1	1									
33i					2										
33k											1				
33n						1									
33p					3										

**Table 1**  
**Summary Statistics and Variable Definitions**

VARIABLE	DEFINITION	MEAN	SD
<b>Patent-level measures</b>			
<i>Patent forward citations within 5 years</i>	Count of forward citations within 5 years of patent issuance.	3.33	4.68
<i>Patent forward citations through 2004</i>	Count of forward citations as of December 2004.	6.63	14.70
<i>Patent generality</i>	Concentration of forward-citing patent classes (see Jaffe and Trajtenberg, 2001), adjusted as per Hall (2005).	0.54	0.33
<i>Patent knowledge bridging</i>	1 – (Number of backward citing patents in common with the primary patent class of the focal patent as a share of backward citations of the focal patent).	0.52	0.38
<i>Patent originality</i>	Concentration of backward-cited patent classes (see Jaffe and Trajtenberg, 2001), adjusted as per Hall (2005).	0.53	0.33
<i>Patent scope</i>	Number of primary classes assigned to the patent, evaluated as of December 2004.	2.71	1.43
<i>Inventor patent experience at other firms</i>	Number of patents issued to focal patent's inventors when employed by other organizations as of the application date of the focal patent.	6.63	11.92
<i>Patent application year</i>	Set of dummy variables = 1 for patent application years from 1979 through 2003.		
<b>Firm-level measures</b>			
<i>Firm knowledge overlap</i>	Angle degree of overlap between firms' primary patent classes and those it cites in a given year, based on a vector dot product.	-40.51	19.38
<i>Founder overlap</i>	Angle degree of overlap between firm founders' primary patent classes and the firm's patent classes in a given year, based on a vector dot product.	-56.01	17.30
<i>Employees (t-2)</i>	Number of employees in time t-2.	719.64	1287.59
<i>Cumulative FDA drug approvals (t-2)</i>	Cumulative number of FDA drug approvals in time t-2.	2.22	4.55
<i>Cumulative FDA drug terminations (t-2)</i>	Cumulative number of FDA drug terminations in time t-2.	3.03	6.45
<i>Entry order</i>	Sequential order of entry (1 is first).	9.84	5.51
<i>Cumulative strategic alliances (t-2)</i>	Cumulative number of strategic alliances in time t-2.	17.54	29.13
<i>Hot biotech funding environment (t-2)</i>	Lerner (1994) index of biotechnology funding environment, as of time t-2.	3.24	1.34

**Table 2**  
**Patent Forward Citation Regressions**  
**(Patent level of analysis)**

	<i>Dep. Var.:</i> Patent forward citations within 5 years of patent issue			<i>Dep. Var.:</i> Patent forward citations through 2004
	Negative binomial regressions			Negative binomial regressions
<i>Independent variables</i>	<b>(2-1)</b>	<b>(2-2)</b>	<b>(2-3)</b>	<b>(2-4)</b>
Patent knowledge bridging	0.174*** (0.063)	0.176*** (0.063)		0.113** (0.056)
Patent originality			0.132* (0.078)	
Patent scope		0.042*** (0.016)	0.060*** (0.018)	0.029** (0.014)
Inventor patent exp. at other firms		-0.001 (0.003)	0.001 (0.003)	-0.002 (0.003)
Patent application year fixed effects	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes
Constant	0.410 (1.011)	0.909 (0.607)	1.036 (0.575)	1.828*** (0.402)
Log likelihood / R <sup>2</sup>	-4204.849	-4196.738	-3375.854	-6397.936
Observations	1878	1875	1501	2938

\*, \*\* and \*\*\* denote statistical significance at the 10%, 5% and 1% level, respectively.

**Table 3**  
**Patent Generality Regressions**  
**(Patent level of analysis)**

	<i>Dependent variable: Patent generality</i>			
	OLS regressions (standard errors are robust & clustered by firm)			
<i>Independent variables</i>	<b>(3-1)</b>	<b>(3-2)</b>	<b>(3-3)</b>	<b>(3-4)</b>
Patent knowledge bridging	0.190*** (0.037)	0.190*** (0.036)		
Patent originality			0.237*** (0.023)	0.232*** (0.025)
Patent scope		0.004 (0.004)		0.006 (0.004)
Forward citations through 2004		0.001*** (0.000)		0.001* (0.000)
Inventor patent exp. at other firms		-0.000 (0.001)		-0.001 (0.001)
Patent application year fixed effects	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes
Constant	0.619*** (0.014)	0.605*** (0.018)	0.562*** (0.016)	0.272*** (0.042)
R <sup>2</sup>	0.13	0.13	0.13	0.14
Observations	1488	1487	1237	1236

\*\* and \*\*\* denote statistical significance at the 5% and 1% level, respectively.

**Table 4**  
**Knowledge Overlap Tobit Regressions**  
**(Firm-year level of analysis)**

	<i>Dependent Variable</i> = Firm knowledge overlap (data censored at -90 and 0)			
	N = 163			
<i>Independent Variables</i>	<b>(4-1)</b>	<b>(4-2)</b>	<b>(4-3)</b>	<b>(4-4)</b>
Founder overlap	0.208*** (0.080)	0.274*** (0.085)	0.249*** (0.089)	0.289*** (0.089)
Hot biotech funding environment in t-2		-2.341** (1.150)	-2.532** (1.178)	-3.255*** (1.189)
Cumulative FDA drug dev. approvals in t-2			0.620 (0.697)	1.327 (0.733)
Cumulative drug dev. terminations in t-2			-0.486 (0.511)	-0.665 (0.506)
Entry order				0.984** (0.368)
Employees in t-2	0.002 (0.001)	0.002 (0.002)	0.002 (0.002)	0.002 (0.002)
Cumulative strategic alliances in t-2	0.007 (0.053)	0.064 (0.076)	0.093 (0.118)	0.125 (0.116)
Constant	-25.799*** (5.474)	-7.301 (8.101)	-2.977 (8.308)	-6.738 (8.285)
Prob > $\chi^2$	0.02	0.00	0.00	0.00

\*\* and \*\*\* denote statistical significance at the 5% and 1% level, respectively.

**Table 5**  
**Knowledge Overlap OLS Fixed Effects Regressions**  
**(Firm-year level of analysis)**

	<i>Dependent Variable = Firm knowledge overlap</i>		
	N = 163		
<i>Independent Variables</i>	<b>(5-1)</b>	<b>(5-2)</b>	<b>(5-3)</b>
Founder overlap	0.266*** (0.094)	0.277*** (0.100)	0.289*** (0.103)
Hot biotech funding environment in t-2		-2.313** (1.166)	-2.215** (1.186)
Cumulative FDA drug dev. approvals in t-2			0.039 (0.945)
Cumulative drug dev. terminations in t-2			-0.291 (0.678)
Employees in t-2	0.002 (0.001)	0.002 (0.002)	0.002 (0.002)
Cumulative strategic alliances in t-2	0.038 (0.060)	0.142 (0.097)	0.181 (0.155)
Constant	-25.410*** (6.372)	-11.424 (7.877)	-7.533 (15.716)
R <sup>2</sup>	0.26	0.33	0.33

\*\* and \*\*\* denote statistical significance at the 5% and 1% level, respectively.