Does Good Science Lead to Valuable Knowledge? Biotechnology Firms and the Evolutionary Logic of Citation Patterns

Michelle Gittelman • Bruce Kogut
Department of Management and Organizational Behavior, New York University, Stern School of Business, 44 West 4th Street, Tisch 7-11, New York, New York 10012
The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104, and Centre de Recherche en Gestion, Ecole Polytechnique, 1, rue Descartes, Paris 75005, France
mgittelm@stern.nyu.edu • kogut@poly.polytechnique.fr

This study looks at the United States biotechnology industry as a community of practice caught between two evolutionary logics by which valuable scientific knowledge and valuable innovations are selected. We analyze the publications and patents of 116 biotechnology firms during the period 1988–1995. In models that link scientific capabilities to patent citations, we show that scientific ideas are not simply inputs into inventions; important scientific ideas and influential patents follow different and conflicting selection logics. Publication, collaboration, and science intensity are associated with patented innovations; however, important scientific papers are negatively associated with high-impact innovations. These results point to conflicting logics between science and innovation, and scientists must contribute to both while inhabiting a single epistemic community. We identify individuals listed on patents and scientific papers and find they effectively integrate science with innovation, leading to more successful innovations. Our findings suggest that the role of the small, research-intensive firm is to create a repository of knowledge; to act as an organizational mechanism to combine the capabilities of versatile scientists within and outside the boundaries of the firm; and to manage the selection of scientific ideas to produce valuable technical innovations.
(*Science; Citations; Patents; Scientists; Epistemic Community; Biotechnology*)

Introduction

One of the important lessons of the sociology of science is that the creation of scientific knowledge is an activity that is institutionally constructed and organized. Until the sixteenth century, scientific endeavors were cloaked in secrecy to withhold knowledge and the powers it conferred from the "vulgar multitude" (David 1998). The institutionalization of science encouraged the validation and diffusion of ideas as open to public scrutiny (Merton 1973). To support these institutions, norms that standardized the language and presentation of results developed under the auspices of academic journals. The careers of scientists were tied to their success in publishing these results in prestigious journals and withstanding subsequent public criticism. As science evolved, it also fragmented into distinct communities, with separate identities, journals, and models of experimentation and validation.

The sociology of science, though rich and variegated, broadly agrees with the view of science as embedded in distinctive communities. As Merton (1973) and economists such as Dasgupta and David (1994) and Stephan (1996) have noted, these norms

0025-1909/03/4904/0366$05.00
1526-5501 electronic ISSN

create incentives that are efficient insofar as professional ranking is related to effort, excess duplication of effort is contained, and scientists desire to broadly disseminate results to earn respected reputations. A priority-based publication system is an important principle that promotes rapid dissemination of knowledge within scientific communities and organizes individual contributions in the form of a series of races. De Solla Price (1970, p. 6) observes, "scholarship is a conspiracy to pool the capabilities of many men, and science is an even more radical conspiracy that structures this pooling so that the totality of this sort of knowledge can grow more rapidly than any individual can move by himself." At the bench level, science is "manufactured" in laboratories in which scientists seek power and alliances to persuade each other that they occupy important positions (Knorr Cetina 1999, Latour and Woolgar 1979). In this context, a published paper is a legitimate tool of persuasion and a symbol of achievement.

Even though science is "manufactured" in the context of academic communities, firms can usefully apply scientific knowledge to develop new technologies; indeed, the argument that science could drive commercial innovation was a major justification for public support of a nation's scientific infrastructure. The dilemma for firms seeking to profit from scientific knowledge, however, is that science is not available as ready-made inputs, but is produced by scientists situated in these scientific communities. The useful equation of science as an input to technology is problematic when scientific inputs are also seen as *not only* producing technology but also *manufacturing* scientific outputs valued by other scientists.

The importance of a community operating within a well-organized social structure, sharing a strong epistemic culture, has not been addressed in studies on the economics of science and technology, which have been concerned with showing the functional relation between scientific inputs and technological outputs. The focus on the "production function" is seen in recent studies that have taken a more fine-grained approach to understanding this relationship; these have found that better science leads to more technology (Cockburn and Henderson 1998, Henderson and Cockburn 1994). What happens when the inputs, called professional scientists, care about their perception of what they do, why they do it, when, and for what kinds of rewards?

We propose that the logic of scientific discovery does not adhere to the same logic that governs the development of new technologies, and that these conflicting logics pose potential problems for science-based innovation. The communities in which scientific ideas circulate and the logics by which they are selected mean that value calculations in science and industry are different. Innovation builds on knowledge made in science, but science that is "good" for innovation is propelled by a logic that is different than that employed by a scientific community to determine "valuable" or "important" science. While industry may need scientific insights to resolve technological problems or find new projects, firms do not directly benefit from contributing to important or controversial scientific questions. We use the difference in the processes by which highly cited scientific discoveries are acknowledged by other scientists and by which valuable innovations are selected by market forces to show how different evolutionary logics weaken the science-technology linkage. Consequently, we generate evidence regarding how firms differentially manage the disconnect between the two evolutionary logics by which science and innovations are rewarded. In all, we find evidence for a single epistemic community, for different and conflicting evolutionary logics between science and technology, and for the important role of bridging scientists in the production of valuable innovations.

Science, Epistemic Communities, and Citation Analysis

The community perspective implies a stickiness in the flow of scientific knowledge to firms. The literature on the motives for firms to publish their research in scientific journals has implicitly or explicitly acknowledged the central importance of forming ties to this community, via boundary-spanning "gatekeepers," to access socially embedded knowledge (Allen 1977, Hicks 1995, Lieberman 1978, Tushman 1977). Research in pharmaceutical and biotechnology companies reveals, in particular, the importance of personal links between commercial firms and universities. The fortunes of biotechnology companies are linked to their engagement of "star scientists" and their location in research regions (Zucker et al. 1998). Indeed, because star scientists often maintain university employment, colocation with research centers is a natural and necessary requirement of successful biotechnology firms. In numerous studies in pharmaceuticals and biotechnology, collaboration with researchers external to the firm and an internal science orientation leads to higher research productivity (Cockburn and Henderson 1998, Cockburn et al. 2000, Gambardella 1995, Powell et al. 1996, Zucker et al. 2002). These studies suggest the operation of absorptive capacity to include direct participation in external communities as a means of acquiring knowledge outside the firm's boundaries (Cohen and Levinthal 1990, Rosenberg 1990). In other words, *how* firms access and practice science—in particular, whether they establish credible linkages with the scientific community—matters to the production of valuable innovations.

Of course, investing in science represents a cost to the firm. Stern (1999) argues that research-oriented scientists have an inherent "taste" for research and publication. The skills of these scientists are needed for translating research into product development. The firm allows them to engage in research and publish as a form of payment in exchange for these other activities; these scientists receive lower wages than scientists who are not allowed to publish. Stern's thesis raises the intriguing implication that the relationship of commercial innovations and scientific knowledge is problematic for the scientist whose activities are aimed at commercial outcomes, but whose identity remains embedded in the values and reward systems of a scientific community.

In effect, those scientists working on research for firms engage in commercial innovative endeavors, while operating within a single "epistemic culture," to appropriate Knorr Cetina's (1999) term to differentiate scientific communities. These cultures code for rules by which scientists define their careers, identities, methods of empiricism, and collaboration with others. Experimental physics utilizes large-scale equipment and engages the efforts of hundreds of scientists who publish jointly. Research in the life sciences relies on smaller teams and is less capital intensive; collaborative publishing, though common, is characterized by fewer authors. Clearly, rules regarding empirical validation and publication are powerful cultural expressions of the collective organization of individual contributions in the development of scientific knowledge.

Citation traces are the bibliometric fossils by which to measure the replication success of an idea. These fossil records permit an investigation of the relative success of an idea to influence subsequent work. Citations to papers are an important way that scientists evaluate their relative standing, by which they exchange gifts, acknowledge prestige, and seek to prevail in their arguments (Crane 1972, Latour and Woolgar 1979). Forward patent citations, that is, citations made by later patents to a patent previously issued, are similarly indicative traces of the importance of commercial innovations, although the process by which they are generated is less deferential to status and reputation effects (Hall et al. 2000). Campbell (1974) contended that academic fields advance by an "evolutionary epistemology" in which favored ideas are promulgated and disfavored ones are lost. The sociology of science shows that articles that are not cited within five years are unlikely to be remembered (Crane 1972). Citation patterns to patents show a similar time frame (Jaffe et al. 1993).

Our analysis relies principally upon citations. We analyze citations in scientific papers to other scientific papers, and also look at citations in patent documents to other patents and to scientific papers. We propose that the difference in evolutionary logics that generate paper and patent citations reflects the difficulty faced by private firms to translate knowledge produced in a scientific setting into valuable technologies. Where the scientific and commercial endeavors diverge is seen in the different citation traces generated by the distinctive rules that govern the logic by which a good paper or a valuable patent is selected and replicated.

How do scientists jointly operate in the distinct communities of science and technology, without damaging their credibility in the former, or their efficiency in producing the latter? To demonstrate first the importance of this question, we turn to an analysis of citations to scientific papers included in other scientific papers and in patents. We put forth the hypothesis that both of these artifacts, papers and patents, will share common antecedents, i.e., they will show similar citation patterns to scientific papers. In other words, there is no ascriptive distinction in the acknowledgments paid by scientists when they publish a paper or when a patent is filed, naming them as the inventor (even if the property right is held by the firm). The citation patterns should reflect membership in the relevant scientific community.

If scientists in universities and those in firms inhabit a single epistemic community, should we not then expect that influential patents should be the product of influential research? At a first pass, it would seem that if citations to scientific research in patents follow the same norms as citations in papers, then influential patents should also cite influential papers. However, the evolutionary logic that selects out "better" patents is different than that which selects the more influential papers. The evolutionary dynamics for patents reflect the joint factors of market demand, technological opportunity, and legal claims on property rights. Market demand increases the efforts to commercialize, and hence patent, more in particular sectors, thus favoring only those related and relevant patents (Mowery and Rosenberg 1982). Second, technologies differ in their opportunities, with some offering a richer set of opportunities than "dead-end technologies" (Kim and Kogut 1996, Stuart and Podolny 1996). Because of the difference in these selection dynamics, we do not expect influential papers to lead to influential patents. Scientists do not reward papers for their market and technological promise; they reward them for reasons proper to their own epistemic community.

Collaborative activity, we would also expect by this argument, reveals the long hand of the sociology of scientific communities. Certainly, firms desire their scientists to engage in external collaboration to improve their productivity and to acquire property rights to research generated in universities and public laboratories. Scientists collaborate on publications, but they do so in accordance within the norms of doing science. Thus, we expect collaboration to help research, and to result in more or better papers, but we propose that this collaboration will be embedded in the social relations and rituals of the scientific community.

We expect to find firm-level heterogeneity in the relationship between scientific inputs and innovation outputs. We model this heterogeneity as the degree to which firms succeed in integrating the two worlds of science and invention. Because we believe that scientific knowledge is embedded in a community inhabited by scientists, we identify these individual scientists and look at whether they publish or patent. We focus on the intersection of these groups as the set of individuals who inhabit both the world of open science and the world of technology creation, and measure their impact on patenting. We pose the question: Are individual scientists who perform both more productive in producing important patents? In this sense, these scientists are technological gatekeepers, as studied by Allen (1977) and Tushman (1977), but more specifically, as suggested by Lieberman (1978), they help firms patent by bridging the worlds of discovery and innovation.

To summarize, we make the following propositions:

Proposition 1. *Papers and patents exhibit similar citation patterns to scientific papers.*

Proposition 2. *The evolutionary logics that select valuable scientific papers and valuable patents are different, and because of this, influential papers are no more likely to lead to influential patents than other papers.*

Proposition 3. *Collaboration choices are influenced by status in a scientific community.*

Proposition 4. *Scientists who bridge discovery and innovation are able to reconcile these two conflicting logics more effectively than those specializing in either science or technology.*

Science, Publications, and the United States Biotechnology Industry

We address these issues in the context of the United States (U.S.) biotechnology industry, which is characterized by rapid knowledge diffusion and intense technological competition. Biotechnology firms are actively engaged in keeping at the forefront of publishing in the scientific literature (Powell et al. 1996). Because this industry so heavily relies on commercializing on the basis of scientific discovery, the relationship between scientific knowledge and innovation outputs are especially strong.

Biotechnology firms act as organizational vehicles for the private appropriation of knowledge produced in university laboratories and moving it to a commercial marketplace. Indeed, the professional and cognitive divide between individuals engaged in scientific research and those engaged in commercialization is much less sharp in biotechnology than in other technology-intensive industries, e.g., microelectronics. Zucker et al. (1998) show that specialized biotechnology firms are formed with the intent of capturing knowledge held by academic scientists close to the "frontier" of knowledge discovery. A startup company enables them to extract economic value from their valuable knowledge (Audretsch and Stephan 1996).

Biotechnology firms are defined in our study as new firms specializing in the use of molecular technologies to develop new drugs, diagnostic tools, or other novel products. However, their collaboration networks extend beyond their own field. If we were to describe, to use Callon's (1986) terminology, the action networks of a biotechnology firm, they would consist of partnerships with other pharmaceutical firms, research institutes, universities, and, to a lesser extent, other startup companies. We know that these networks have two important properties. First, Powell et al. (1996) showed that firm performance increases with the intensity of interfirm collaboration. Second, Shan et al. (1996) found that networks have a self-replicating property; cooperation reenforces the network by building upon the previous pattern of relationships. Thus, collaboration appears to result in the creation of useful knowledge in a durable web of relationships for the production of patentable knowledge.

In summary, biotechnology startups are characterized by a heavy reliance on scientific knowledge sourced from university and academic laboratories. Their own scientists, while competing to produce valuable technologies, also actively engage them in the production of that knowledge. Thus, the biotechnology industry is a rich field for an analysis of the relationship between scientific capabilities of firms and valuable innovations.

Modeling Science as an Input to Discovery

Sample of Firms and Data Collection

The first step in the sampling procedure was to create a representative sample of U.S. biotechnology firms. To accomplish this, we made use of an existing database established by one of the authors (Gittelman 2000). This database includes some 14,000 biotechnology patent records, each corresponding to a single invention filed by U.S. organizations during the period 1982–1997. The source of the data is *Derwent Biotechnology Abstracts*, a comprehensive database of biotechnology patents. Patents are restricted to those as classified by Derwent as relating to genetic engineering and/or biopharmaceuticals. Patents relating to plant and agricultural uses, and other industrial applications outside of human health care, are excluded. From the patent data, a number of sources were used to identify which of the assignees were U.S. biotechnology firms. All identifiable biotechnology firms were included, subject to the criterion that the company was based in the U.S., and was granted at least one U.S. patent during 1988–1995. Biotechnology firms that are subsidiaries of other firms, but have maintained an independent identity are included, e.g., Genentech, once partially owned by Roche. Biotechnology divisions of pharmaceutical firms are not included. The primary sources used to identify biotechnology firms include BioScan, a proprietary directory of the biotechnology industry, Ernst & Young annual biotechnology reports, member directories of the Biotechnology Industry Organization, and company sources.

These criteria yielded a sample of 116 U.S. biotechnology firms. We collected four kinds of data for the firms in our sample: (1) data on publications in the scientific literature during the period 1988–1994 (these data also reveal research collaborations with external institutions), (2) patents issued to the firms in the U.S. during 1992–1995, (3) individual scientist data, and (4) data on firm-level characteristics.

Publications and Collaborative Research Among Sample Firms

The sample firms produce many more publications than they do patents. Our sample of firms published nearly 7,000 articles in the scientific literature (about 1,000 per year) during the sample period (1988–1994), with 30% of articles published in just 10 journals (including the prestigious publications *Nature, Science*, and *Cell*). The total number of articles published by the firms has been rising at about 10% per year, from 711 articles in 1988 to 1,258 articles in 1994.[1]

The firms were granted some 1,200 U.S. patents, a rate of about 300 per year, during 1992–1995. This is a rough indication that a significant portion of publications did not lead to a patent. The average firm in our sample published 60 articles, was granted 10 patents, was founded in 1984, and employed 262 persons. There is a great deal of heterogeneity among these firms. One firm (Genentech) accounted for 1,400 publications, while some firms had no publications. Size differences are also great, from 3,000 employees (Amgen) to 5 employees (Symbollon and Immunologic Pharmaceuticals).

Collaboration with external organizations is a defining feature of the research activities of the sample firms. We measure research collaboration as represented by articles in which both the firm and an outside organization are listed as institutional affiliations of one or more of the authors. The share of collaborative publications is about 70% of total articles published, and this has remained steady during the sample period. This high proportion is not skewed by firm size.

The great majority of shared research is between a biotechnology firm on the one hand and a university on the other; firm-to-firm collaborations are a small portion of the total. In all, some 1,800 organizations are listed as collaborating institutions with the sample firms. An analysis of the top 200 of these research partners shows that only 15 were other firms; the rest were universities, research institutes, and government labs, U.S. and foreign. Extrapolating from this, it is estimated that 90% of the research partners were universities or other research institutions (government labs, hospitals, or research institutes). The data indicate that copublications allow the firms in our sample to tap into high-quality networks of academic scientists, with prestigious universities and research institutes in the life sciences dominating the population of collaborators.[2] Given the sensitivity to prestige and grounding in scientific practice, it is not surprising that past studies found that many collaborations are not formalized in legal contracts (Liebeskind et al. 1996).

Dependent Variable: Forward Patent Citations

We are interested in exploring whether scientific research impacts the value of a firm's innovations, as captured by its patents. Our dependent variable is the cumulative forward citation frequencies to an individual patent.[3] Forward citations count the number of times a patent (the "cited patent") is included in the prior art of subsequent patents. The evidence strongly supports the conclusion that patent citations contain information about a patent's technological importance, and that they can also be used as a proxy for economic value to the innovator (see Hall et al. 2000 for a review). In biotechnology, where patents are a key means of appropriating returns to innovation, citation rates are more likely than in other fields to contain information about the technological and economic value of a given invention. We count all forward citations received by each patent at of the end of 1999. We call this measure CITES TO PATENT.

[1] We use the *ISI Science Citation Index* to collect information on all publications in which the firm is listed as an institutional author during the period 1988–1994.

[2] The top institutional collaborators are (with percent of all collaborative articles): Harvard University (6%), University of California–San Francisco (5%), University of Washington (5%), National Cancer Institute (5%), Stanford University (4%), University of Texas (4%), University of California–Los Angeles (3%), Scripps Clinic and Research Institute (2%), Johns Hopkins University (2%), and University of California–San Diego (1%). In total, the University of California (UC) system accounts for 13% of the collaborations, reflecting the pronounced role of the UC system in fostering a California biotechnology industry and the linkages between UC researchers and scientists at those firms.

[3] Our data give information on the full patent family, comprising the full portfolio of patents issued around the world on a given invention. We utilize the first U.S. patent in the family issued during 1992–1995, inclusive.

Citations may accrue to a patent for reasons that do not reflect its importance but rather its vintage: older patents are likely to be cited more than, but are not necessarily more important than, younger patents. Technological field effects may also influence citation rates: patents in crowded fields may be cited more than patents in sparse fields because there are more citing patents. On the other hand, patents in sparse fields may have higher odds of being cited by subsequent patents because there are fewer cited patents. In these cases, the reasons for citations are likely unrelated to the importance of the patent that we are aiming to measure. We are, therefore, careful to include variables in our regressors that control for patent age and technological field, as well as other characteristics of the firm and cited patent that could affect the frequency with which it is cited.

Our models seek to capture knowledge capabilities at the level of the firm as a whole; to do this, we aggregate various data from all of the firm's patents and publications. For firms with only one patent, the data likely give a poor measure of firmwide capabilities. We, therefore, leave out firms with only one patent in our sample when we estimate our models. This eliminates 15 patents (corresponding to 15 firms). As we are not modeling the *level* of innovation effort, but rather citations to patents, our models should be interpreted as estimating the relative success of the innovative effort, conditional on the firm having a capability to innovate. Additionally, to help ensure that the research effort covered by our patents does not precede the research effort represented by publications, we eliminate 96 patents that were filed prior to 1987, corresponding to an expected 1988 journal publication date, the first year of our publication data.

Independent Variables: The Bibliometrics of Publishing, Copublishing, and Patenting

First, we consider the effect of investing in science on the firm's patents. To measure this effect, we develop several variables from the bibliometric and patent data.

Firm Publish Dummy. This takes a value of 1 if the firm published at least one article up to the year in which the observed patent was filed.

Publication Volume. The total number of firm publications, cumulated up to the year in which the observed patent was filed. This gives an indication of the volume of publishing. The variable is specified in log form to take account of the highly skewed distribution of publications.

Percent of Copublications. Percentage of all publications by the firm that were collaborative publications with an external organization, cumulated up to the year the observed patent was filed.

Science Intensity of Firm's Patents. We measure the closeness of the firm's technologies to knowledge produced in open science. A patent is required to list the prior art that it builds upon: this includes both other patented inventions and publications in the scientific literature that are not patented. Firms that seek to integrate scientific findings into their inventions are more likely to cite scientific findings in their patents, hence have more nonpatent references in their prior art. Science intensity is a "backwards" citation count, measured as the number of times a patent references nonpatented literature in its prior art. Deng et al. (1999) find that science intensity is positively associated with subsequent financial performance for a group of technology- and science-based companies. There is great variation in the degree to which patents in our sample build upon science: the mean is around 40 citations to published works, with up to as many as 1,675 such citations. For the firm, we calculate the mean number of citations to nonpatented literature across all of its patents.

Firm Average Cites to Publications. This variable captures citations to a firm's publications and is our primary measure of firm-level scientific research capabilities. Raw citation counts to each article are normalized by the mean and standard deviation of citations received by all sampled articles in its publication year. Normalizing the raw citations by year allows citations to be summed across years for each firm; the aggregate citation counts are then divided by the number of the firm's publications, to yield an average citation measure for the firm as a whole. Averaging the citations this way removes bias toward large-volume publishers; we have separately estimated the effect of publication volume and here we want to isolate the effect of publication quality. In our models, normalized citations are aggregated up to the year the observed patent was filed. The measure, therefore, represents the relative quality of the firm's stock of scientific knowledge, cumulated from the start of the publication period up to the time of the observed innovation.

Control Variables

We also include a number of control variables, to account for heterogeneity among the firms, as well as to control for age and field effects.

Patent-Level Controls.

Age of Patent. Years elapsed since the patent was filed. This control is particularly important, because we expect citations to patents and to papers to increase with age.

Patent Family Size. This variable is an indicator of the value of the invention to the firm, as evidenced by the number of the patents the firm issued or renewed in different countries (Cockburn and Henderson 1998). We count the total number of patents in the patent family, including patents granted overseas, whose forward citations are captured in our dependent variable. It is costly to maintain multiple patents; this variable, therefore, acts as a fixed-effect control for each invention, allowing for random luck in the innovation process and firm efforts to promote their innovations in multiple markets.

Patent Number of Inventors. We hypothesize that the research effort is associated with the number of people assigned to that effort, and that this is reflected in the number of people listed on the patent. We therefore include this measure as a proxy for the resources invested in the research project that resulted in the observed patent.

Technology Class of Patent (*Patent Drug*, *Patent Test*). We wish to control for technology segments that may be inherently more cited than others. We expect that patents that are in drug-related categories may be less cited than technique-based patents, as the former may represent a stopping point in further innovation once a patent has been issued, whereas techniques may spawn a host of incremental innovations. As these patterns would not necessarily reflect the underlying importance of the innovations (and, indeed, may mask importance in the case of a drug patent), we need to create controls for them, as the classification system does not automatically distinguish patents in this way. Using international patent classification codes, we create two main categories of technologies: Patent drug indicates whether the patent is classified in A61K, medicines and pharmaceuticals; patent test indicates whether the classification is C12Q or G01N, which cover measuring, testing, and immunoassays using genetic materials. These two categories account for 345 patents.[4]

Firm-Level Controls.

Firm Age. Number of years since the firm was founded. Older firms have had more time to accumulate a knowledge base that can be applied across a range of innovations, however, they may represent knowledge of an older vintage than younger firms.

Firm Pharmaceutical Strategy. This is a broad measure of the technological orientation of the firm, to identify those firms that are seeking to develop biopharmaceuticals against firms that are primarily specialized in research tools, tests, instruments, and information-based products and services. From the BioScan (1994) data, we coded for a dummy variable that takes a value of 1 if the firm is involved in research oriented toward discovering new human biotherapeutics. Sixty-seven firms are coded as belonging to this category. We expect that on average investments in science will have a greater payoff for firms engaged in drug discovery.

Table 1 gives summary data for the dependent and independent variables, and Table 2 gives the bivariate correlations. None are high enough to suspect multicollinearity, further confirmed by regression results and additional tests (discussed further below).

Model Specification

Because the data are counts of citation frequencies, we employ a count model that makes use of the information contained in the numerous observations that

[4] In models not shown here, we control for all technological subfields by adding dummies for the main (first) patent class listed on each patent, but these controls were not significant and did not add to the power of the model.

Table 1 Summary Statistics

	Mean	Median	Maximum	Minimum	Std. dev.
Cites to patent	12.3	7	463	0	21.5
Patent age	8.88	9	13	5	2
Patent family size	6.14	6	39	1	5.32
Patent number inventors	3.05	3	20	1	2.12
Firm age	16.9	17	25	7	3.98
Firm pharma strategy	0.69	1	1	0	0.46
Patent drug	0.28	0	1	0	0.45
Patent test	0.08	0	1	0	0.28
Firm publication dummy	0.93	1	1	0	0.26
Firm publication volume	139	31	1,395	0	260
Firm % copublication	0.66	0.7	1	0	0.24
Firm % joint patent-publishers	0.6	0.63	1	0	0.21
Firm average cites to publications	−0.29	−0.19	5.63	−3.36	1.12
Firm science intensity	39.9	28.8	373	0	32.7
Patent science intensity	37.9	17	1,675	0	82.8
Patent % joint patent-publishers	0.69	0.86	1	0	0.37

are never cited. Count data are frequently estimated with one parameter Poisson models. Poisson models are nested within the negative binomial model, a two-parameter model that estimates an overdispersion parameter and produces correct standard errors for count data that is overdispersed (Cameron and Trivedi 1998). Because patent citations exhibit a great deal of overdispersion, we estimate negative binomial models. We test for robustness by estimating cluster regression models.

Models of Publishing Effects on Patent Citations

Models 1–4 (see Table 3) include the control variables and add different science investment variables. Model 1 includes only the control variables. As expected, older patents and patents from large families receive more citations. The proxy for cost of the project, number of inventors on the patent, is also positive ($p < 0.01$). There is a small negative effect of firm age but it is not significant. Firms coded as working on drug development do receive higher cita-

Table 2 Bivariate Correlations

Variable		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	Cites to patent	1														
2	Patent age	0.36	1													
3	Patent family size	0.33	0.47	1												
4	Patent number inventors	0.24	0.12	0.26	1											
5	Firm age	0.08	0.17	0.05	0.05	1										
6	Firm pharma strategy	0.12	0.04	0.11	0.11	0.24	1									
7	Patent drug	0.06	0.12	0.22	0.08	0.11	0.25	1								
8	Patent test	0.00	−0.05	−0.07	−0.03	−0.11	−0.22	−0.20	1							
9	Firm publication volume	−0.04	−0.22	−0.06	0.01	0.55	0.28	0.08	−0.04	1						
10	Firm % copublications	0.09	0.05	0.06	0.09	0.15	0.24	0.06	0.09	0.07	1					
11	Firm % joint patent-publishers	0.08	0.05	−0.01	−0.05	0.22	0.26	0.09	−0.16	0.32	−0.09	1				
12	Firm average cites to publications	0.07	0.20	0.11	0.05	0.29	0.33	0.09	−0.07	0.33	0.25	0.14	1			
13	Firm science intensity	0.09	−0.01	0.09	0.10	0.40	0.32	0.12	−0.08	0.47	0.03	0.24	0.19	1		
14	Patent science intensity	0.07	−0.08	0.16	0.14	0.14	0.11	0.11	−0.04	0.18	0.02	0.06	0.07	0.27	1	
15	Patent % joint patent-publishers	0.09	0.07	−0.03	−0.07	0.11	0.06	0.05	−0.09	0.14	−0.10	0.47	0.03	0.07	0.09	1

Table 3 Negative Binomial Models of the Effect of the Quality of a Firm's Science Capability on the Production of Highly Cited Patents

Model	1. Control variables only		2. Firm publication dummy		3. Collaboration and science measures for firms that publish		4. Citation to publications without other firm-level science indicators	
Sample	All firms with > 1 patent		All firms with > 1 patent		All firms with > 1 patent and > 1 publication		All firms with > 1 patent and > 1 publication	
Variable	β	S.E.	β	S.E.	β	S.E.	β	S.E.
Constant	−2.64***	0.34	−2.80***	0.37	−3.67***	0.42	−3.35***	0.40
Log (patent age)	2.02***	0.16	2.05***	0.16	2.38***	0.19	2.19***	0.19
Patent family size	0.05***	0.01	0.05***	0.01	0.03***	0.01	0.04***	0.01
Patent number of inventors	0.06***	0.01	0.06***	0.01	0.08***	0.02	0.08***	0.02
Firm age	−0.00	0.01	−0.00	0.01	−0.02*	0.01	0.00	0.01
Firm pharma strategy	0.21***	0.07	0.21***	0.07	0.23***	0.09	0.33***	0.08
Patent drug	−0.05	0.07	−0.05	0.07	−0.06	0.07	−0.06	0.07
Patent test	0.23**	0.11	0.23**	0.11	0.19	0.12	0.24**	0.12
Firm publication dummy			0.13	0.12				
Log (publication volume)[a,b]					0.05**	0.03		
Firm % copublications[b]					0.23	0.15		
Firm science intensity					0.00***	0.00		
Firm average cites to publications[b]					−0.14***	0.04	−0.11***	0.03
Overdispersion parameter	−0.14***	0.05	−0.14***	0.05	−0.15***	0.05	−0.13***	0.05
N	1,120		1,120		934		934	
Log likelihood	−3,748		−3,748		−3,044		−3,052	

*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.
[a]A value of 1 is added to each observation.
[b]Cumulated to the year of patent filing.

tions to their patents. On the other hand, drug patents do not receive significantly more citations, although patents in test-related categories patents do. These are not necessarily contradictory results: firms with a pharmaceutical strategy also patent in test-related categories.

Model 2 adds a dummy variable to show whether the firm had any publications up to the file date of the observed patent. It is positive but not significant. This surprising result may indicate that science has no significant impact on innovation outcomes; or it may indicate that there are groups of firms within our sample, and that for some of the firms, science does not impact innovation. Finally, poor specification of the variable is a plausible explanation. Because most of the firms in our sample had at least one publication (only 7% of patents were issued by firms with no publications), this rather crude measure of science investments may not pick up real differences in research levels and capabilities.

The models that follow use measures taken from each firm's publications; in all subsequent models, we, therefore, only include patents of firms that have published at least one article up to the filing date of the cited patent. Model 3 includes publication volume, percent of collaborative publications, science intensity of the firm's patents, and the average citations to the firm's publications. We consider each in turn.

Publication volume raises the patent citation rate and is significant ($p < 0.05$). This result is counter to the findings of Cockburn and Henderson (1998) and Gambardella (1995) among pharmaceutical firms, who find (respectively) that the volume of publication does not appear important in predicting patent performance, and that only recent publications appear to matter in estimates of patent outputs. However, if we remove Genentech from the analysis shown in Model 3, the variable is no longer significant ($p = 0.14$), while the coefficients and standard errors of all the other variables remain unchanged. Because Genentech is such a large part of our sample—accounting for 10% of patents and 20% of publications—and places such strong emphasis on science-driven discovery, we ran all our models both with and without it. We do not report those separately; in all cases, coefficients and standard errors of the reported variables only slightly change significance levels are unchanged.

The degree to which the firm collaborates is positive but not significant in Model 3 (in subsequent models, it is just significant at the 10% level). Empirical evidence has shown the positive role of collaboration in innovation (Cockburn and Henderson 1998, Powell et al. 1996); the measure here may not be sufficiently fine grained to pick up these effects. Science intensity is positive and highly significant ($p < 0.01$). Patents that build on science are more likely to be cited, hence, more likely to generate further innovations and value. This result supports the notion that integrating scientific research with innovation has a positive impact on a firm's innovation performance, and adds to earlier findings that science intensity is associated with firm value (Deng et al. 1999).

The variable of central importance to our study is firm average cites to publications. It has a negative effect on the citation rate ($p < 0.01$). Highly cited papers are associated with less cited patents. The negative sign implies that the production of high-quality publications actually detracts from the innovation effort. This is a strong result, for it indicates that successful patents and successful papers follow different selection logics, and that these logics are opposing. To confirm this, we construct the scientific quality variable in a variety of different ways. We measure research quality as the firm's percentile ranking among all firms in the sample, ranked according to citations to their publications; by including the citations to the firms' single most highly cited publication across all years in the sample period; and by estimating separate models for citations to firm-only versus collaborative publications. The results are robust to all specifications of the variable: negative and significant effect ($p < 0.01$) on patent citations.

It is important to note that the bivariate correlation between highly cited patents and highly cited papers is positive at 0.36, and a regression of important papers on important patents shows a significant and positive relationship. However, once age of the patent is included, the relationship turns negative and significant; a likelihood test shows the added variable significantly improves the fit. Theoretically, this is to be expected: older patents cite older papers and both have more forward citations. Thus, the bivariate correlation is spurious, reflecting age, and, hence, disappears once a control is added. As a further check, in Model 4, we regress average citations to publications without any of the other firm science measures. The coefficient is still negative and significant and the standard error is stable, indicating that the relationship is not due to collinearity with the other firm science measures.

Robustness: An Independent Test to Investigate Measurement Error

To test the validity of this important finding, we collect additional data and perform additional tests. We want to see whether patents that build upon highly cited scientific articles are more influential than patents that build on undistinguished scientific articles. If influential patents are associated with highly cited scientific articles, that would contradict our model result of publication citations having a negative impact on patent citations.

We construct two separate samples, ranked sample and firm sample, to measure important science on an absolute scale and on a firm-specific scale, respectively. Each sample includes two groups: highly cited articles and a control group. We first consider the ranked sample. The highly cited group in the ranked sample includes articles falling into the top 0.05 percentile of all articles published by the firms in 1990 (49 articles) or 1991 (57 articles), for a total of 106 highly cited articles. The average publication in this group was cited 470 times; the group includes major scientific findings published in prestigious journals such as Nature, Science, and Cell. It is an indicator of the excellent science carried out by the firms publishing these articles that one-quarter of the articles (26 out of 106) were firm-only publications; the remainder were coauthored with scientists at universities or research institutes. To create the control group, each of these top-ranked articles is matched to 106 randomly sampled articles from our database published in 1990 (49 articles) or 1991 (57 articles) that are not in the highly cited group. The average publication in the control group was cited 42 times, with one-quarter (27/106) as firm-only publications.

The firm sample is constructed to measure science quality on a scale relative to a firm's total scientific outputs. We select the most highly cited article published by the firm in any year to create the highly cited group (total = 93 articles, 1 per firm).[5] The average publication in this group received 321 citations, and 24 were firm-only articles. For the control group, we match the firm's most highly cited article to the median article published by the firm in the same year (total = 93 articles; average citations = 30, 26 firm-only articles). We run separate tests for ranked sample and firm sample.

Table 4 shows that for both the ranked sample and the firm sample, highly cited articles generate many more patents than the control group. This is true for total patents as well as patents by the firm that authored the article (self-patents). Important science attracts innovation efforts, both by firms that generated the findings and by other organizations.

These data establish that firms cluster their innovate efforts around important scientific findings. Interestingly, this is true whether science importance is measured in an absolute sense (ranked sample) or relative to a firm's own scientific research outputs (firm sample).

We next consider whether these patent clusters around important science are more valuable as innovations. We test whether patents that build on highly cited scientific articles receive more citations, through the year 2000, than patents that build on the control groups. We regress the citations received by each patent against a dummy variable, indicating whether

Table 4 Number of Patents Citing Articles: Highly Cited Articles Versus Control Groups (Self-Patents Shown in Parentheses)

Group	Ranked sample	Firm sample
Highly cited articles	232 (46)	194 (29)
Control group articles	44 (6)	58 (13)

the patent references a highly cited article in its prior art (variable takes a value of 1) or a control group article (variable takes a value of 0). In the ranked sample, only 3 patents cite both a highly cited article and a control group article, indicating that these are different clusters of innovations. In the firm sample, 12 patents cite both a highly cited and a control group article. We remove these patents from our estimations. A positive and significant coefficient for the dummy variable would indicate that patents that cite important articles are more highly cited themselves: valuable science leads to valuable innovations. This finding would contradict our earlier model results.

Table 5 reports the results of these regressions. Both are negative binomial estimations. We include controls for the age of the cited patent, the total number of articles cited in the patent's prior art, and whether or not the patent was a self-patent (assigned to the same firm as authored the article). The results provide strong corroboration for our earlier result. In the ranked sample, highly cited articles are negatively associated with patent citations ($p < 0.01$). In the firm sample, the sign is positive but not significant ($p < 0.68$). The findings are strongly suggestive that highly cited patents do not build upon valuable

Table 5 Negative Binomial Models of Citations to Patents Citing Sampled Articles

	Ranked sample		Firm sample	
Variable	Coeff.	S.E.	Coeff.	S.E.
Intercept	-2.88^{***}	0.49	-3.17^{***}	0.51
Log (patent age)	2.29^{***}	0.23	1.98^{***}	0.19
Science intensity of patent	0.30^{***}	0.09	0.30^{***}	0.09
Self-patent	0.27	0.27	0.49^{*}	0.28
Cites a highly cited article = 1	-0.81^{***}	0.33	0.11	0.26
Overdispersion parameter	0.67^{***}	0.13	0.38^{***}	0.14
Log likelihood	-452		-444	
N	229		212	

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

[5] This sample is smaller than our full sample of firms, because it is limited to firms that published at least two articles during a given year of our sample period.

science, and, indeed, important science leads to innovations that poorly perform.

The tests of patenting and patent citations indicate that firms cluster their innovation activities around important scientific discoveries, both their own and those made by other firms. However, innovation clusters around highly cited scientific discoveries do not result in highly cited patents, indeed, patent clusters around important science produce less cited patents. This provides support for our earlier finding of conflicting logics between the selection of valuable scientific ideas and valuable technologies.

Bridging Scientists: Joint Patent-Publishers

Why, then, do firms invest in carrying out excellent science, and why do they associate with prestigious scientists? One possibility is that some firms are "captured" by star scientists whose reputations bring prestige and tangible resources to firms. The most skilled practitioners of their art are those who are most wedded to its selection logic yet our results suggest that the process of selecting among ideas comes into conflict with the selection logic of patents. The firm's ability to integrate and mediate these conflicting logics becomes important: as shown by Cockburn et al. (1999), the balancing of incentives based on science with rewards that are market oriented becomes a fulcrum for explaining firm heterogeneity in innovation performance.

We model firm heterogeneity in innovation as the degree to which firms succeed in integrating the two worlds of science and invention at the level of the individual scientist. We identify two overlapping sets of scientists. The first group, called publishers, includes those scientists who are listed on at least one publication in our sample. In total, we identified 19,638 different names of publishers; while our data do not allow us to identify where these individuals actually work, we suspect that a significant proportion are employed by outside institutions, mainly universities. The second group, called inventors, includes scientists involved in developing new technologies, as revealed by their being listed on a patent. On the patents, we identified 2,035 names. This yields a ratio between publishers to inventors of about 10 to 1. Our data indicate that these two groups of scientists are largely distinct, though they overlap. Among inventors, the overlap with the publishers is relatively large: 57% of inventors are also publishers (1,170 out of 2,035 inventors). We identify these individuals as forming a group of joint patent-publishers. However, only 6% of publishers are also inventors (1,170 out of 19,638). This indicates that the firms are intensely leveraging the contributions of scientists working in academic institutions, because few of this large group ever appear on a patent for the firm.

We expect that this measure of overlap between scientists who publish and patent at the firm level is an important indicator of the degree to which a firm is able to successfully translate research into invention. To capture this heterogeneity, we construct a variable called percent of joint patent-publishers. We calculate this variable first at the patent level, as the percentage of all individuals listed on a patent who are also listed on at least one publication. We then calculate this variable for the firm as a whole, by aggregating all scientists listed on the firm's patents during the sample period. These give measures of the degree to which scientists who patent are also active (or have been active) in scientific research, at the level of the individual project (patent) and at the aggregate firm level.

Model 5 (See Table 6) adds the effect of joint patent-publishers at the firm level to the earlier models of patent citations. The variable is positive and significant ($p < 0.05$) and including it does not affect our previous findings about science intensity and science quality. Model 6 replaces the measure at the firm level to show the percent of joint patent-publishers at the level of the cited patent itself. We also include the science intensity of the cited patent rather than for the firm as a whole. These measures can pick up effects at the more detailed project level rather than the aggregated data for the firm. Both variables are positive and significant ($p < 0.01$). Finally, to show robustness, in Model 7, we re-estimate Model 6 by using a clustering regression that controls for random firm effects. A fixed-effect model, with firm dummies included but not shown, is reported in Model 8; several firms are

Table 6 Negative Binomial Models of Effects of Bridge Scientists on the Production of Highly Cited Patents

	5. Effect of firm's joint patent-publishers		6. Effect of patent's joint patent-publishers		7. Model (6) with standard errors adjusted for clustering of firm		8. Model (6) with firm fixed effects (firm dummies not shown)		9. Model 7: elasticities	
	All firms with > 1 patent and > 1 publication		Same as (6)		Same as (6)		Same as (6)		Same as (6)	
Variable	β	S.E.	β	S.E.	β	Robust S.E.	β	S.E.	ey/ex	S.E.
Constant	−3.82***	0.43	−4.0***	0.44	−3.98***	0.61				
Log (patent age)	2.33***	0.20	2.38***	0.20	2.34***	0.12	3.11***	0.34	5.05***	0.63
Patent family size	0.03***	0.01	0.03***	0.01	0.03***	0.01	0.04***	0.01	0.18***	0.07
Patent number of inventors	0.08***	0.02	0.08***	0.02	0.08***	0.02	0.10***	0.12	0.23***	0.07
Firm age	−0.02	0.01	−0.01	0.01	−0.01	0.02	0.33	0.44	−0.21	0.27
Firm pharma strategy	0.20**	0.09	0.25***	0.09	0.25	0.14	−0.52	0.85	0.18*	0.10
Patent Drug	−0.08	0.08	−0.08	0.08	−0.07	0.09	−0.09	0.08	−0.02	0.02
Patent Test	0.22*	0.12	0.21*	0.12	0.21	0.16	0.31***	0.12	0.02	0.01
Log (publication volume)[a, b]	0.04	0.03	0.04	0.03	0.04	0.22	0.18*	0.10	0.15	0.13
Firm % copublications[b]	0.29*	0.15	0.26*	0.15	0.26	0.22	0.40	0.36	0.17	0.15
Firm science intensity	0.00***	0.00								
Firm average cites to publications[b]	−0.14***	0.04	−0.13***	0.04	−0.013***	0.05	−0.15*	0.08	−0.05***	0.02
Firm % joint patent-publishers	0.43**	0.21								
Patent science intensity			0.07***	0.03	0.07*	0.04	0.05**	0.03	0.19*	0.12
Patent % joint patent-publishers			0.25***	0.10	0.25**	0.11	0.21**	0.10	0.18**	0.08
Overdispersion parameter	−0.15***	0.05	−0.16***	0.05	−0.16	0.07	−0.46***	0.06		
N	934		934		934		934		934	
Log likelihood	−3, 042		−3, 040		−3, 040		−2, 916			

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.
[a]A value of 1 is added to each observation to take the log.
[b]Cumulated to the year of patent filing.

dropped due to collinearity. Both indicate robustness of the main results to within-firm variance.

In the final column, we report the elasticities of Model 7. The coefficient to the highly cited publication variable implies an elasticity of −0.05, with a 95% confidence interval ranging from 0.01 to −0.35. These effects are not high, until it is recalled that the citations to the more successful papers (see Discussion below) average around 500. The important effect of patent-publishers, shown here at the project level (with an elasticity of 0.18 for Model 7) supports the interpretation that integrating research and innovation at the level of the individual scientist is more important to the innovation effort than firm-level scientific capabilities as measured by the volume and quality of scientific publications. As a note on the side, by far the most influential effect on patent citations is age, with an elasticity near five.

Discussion

In a knowledge-based industry, it is reasonable to expect that firms with access to superior knowledge resources or skills should outperform those with weaker resources or skills. Our models do not provide strong support for this hypothesis. Highly cited patents are associated with science intensity and firm effects, but not with cutting-edge science; they are associated with scientists who publish, but only weakly associated with publication volume. Indeed, we find that the ability to produce excellent science has a strong *negative* impact on the patent citation rate. Taken together, the models indicate that investing in scientific research produces mixed results, and the relationship between research and innovation is more complex than a simple human capital story would predict.

Instead of a smooth internal transfer between firm scientific capabilities and innovation, the results indicate the very different processes involved in acquiring scientific knowledge and generating high-impact innovations. The negative relationship between scientific capabilities and the innovation effort points to a problematic disconnect between the scientific knowledge of the firm and its ability to generate high-impact innovations. Scientific ideas are not simple inputs into inventions; important scientific ideas and influential patents follow different and apparently conflicting evolutionary logics. This raises the question: Why do firms invest in scientific research when those investments do not seem to pay off in terms of more highly cited patents?

Two factors emerge as important in predicting patent citations. High-impact innovations heavily build upon the scientific literature and are made by people who both invent and do research. These factors are not independent of one another. Joint patent-publishers may perform the important function of identifying and applying the scientific research that the firm would most profit from in its projects. This function includes identifying as well as accessing external researchers in the field who are likely to bring new or complementary knowledge to the firm. Put another way, bridging the disconnect between scientific knowledge and innovation appears to depend on access to individuals who perform both activities, rather than on the ability to generate valuable scientific knowledge alone. Papers and patents do not follow the same selection logics and, yet, scientists produce both. Firms recruit scientists who can successfully bridge these logics and provide incentives that support their dual activities. In this regard, our findings imply the firm-level properties that Cockburn et al. (2000) found important.

Conclusions

Scientific knowledge and patents are related, but good publications and good patents are not. This can be easily explained by recalling that the two artifacts are not chosen by the same evolutionary logic of selection. In other words, patent citations are filtered by the conjoint influence of technical richness and market impact. These are very different evolutionary criteria than those faced in the world of publications. As long as these heavily cited patents defer to the papers that influenced them, the process will generate a very different selection citation pattern for influential patents than for patents overall.

This filtering of the technologically valuable patents by the selection dynamics among patent citations means that there is a technological and market component to patenting. Namely, because certain patents open richer technological veins, the subsequent advances in related technical knowledge encourage more innovative efforts in that area and, hence, more patents. These, in turn, cite the initial patents that opened this avenue of technological innovation. It is this feedback that carves a trace in the patent patterns. Patent citation patterns do not acknowledge what Merton (1973) called Matthew effects in science of prestige, attracting citations and resources; they reflect perceived technical and market opportunities.

This conclusion has a simple implication for understanding what firms do in biotechnology. On the most basic level, a firm that has excellent capabilities to do scientific research may not succeed well in producing marketable innovations, as indicated by Stern's (1999) analysis of scientists' wages. However, having a reputation for performing "good" science may be necessary to attract the kinds of people the firm needs to innovate. Firm heterogeneity in innovation performance centers on the ability to translate knowledge produced within the epistemic community of science into knowledge that a market will value. Scientists who simultaneously publish and invent are instrumental in bridging the disconnect between scientific knowledge and important technologies. Heterogeneity in innovation performance comes from firms' abilities to access and create the capability to do science, while bypassing the evolutionary logic that selects among its outputs. This role points to potential differences in the capabilities of firms to recruit and manage intellectual capital, as found in the studies by

Cockburn and Henderson (1998) and Cockburn et al. (2000).

Our study concerns an industry with particularly strong linkages between technological innovation and scientific knowledge (Cohen et al. 2002). Although the reliance on public science sets this industry apart, in a wider perspective, these findings are not unique to science- or technology-based industries. They point to the broad claim that knowledge of firms is created within and shaped by occupational and epistemic communities. Individuals embody knowledge that is useful when moving within the firm (Argote et al. 1995) or between firms (Almeida and Kogut 1999, Gittelman 2000). They are also anchored in identities and in what van Maanen and Barley (1984) call "occupational communities" that span across firm boundaries. These communities influence as well the organizing principles that guide the internal structure and the coordination among people and divisions inside the firm. While a resource to the firm, occupational communities pose potential conflicts in directing the exploration and efforts of their members. These results point to the important influence of membership in communities broader than a firm's boundaries that both abet and hinder the search for commercially valuable technological innovations.

Acknowledgments

The authors thank Andreas Noetzel and Mervyn Tan for exceptional research assistance. This paper benefited from comments received from participants in New York University's Management Seminar, and the Harvard Strategy Seminar, as well as from Pierre Azoulay, Lee Fleming, Diana Hicks, Maryellen Kelley, Steve MacMillan, Richard Nelson, Brian Silverman, Olav Sorensen, Brian Uzzi, Andrew Wang. The authors also thank the editors of this special issue, the referees, and associate editor. This research was financed by the National Institute of Standards and Technology and the Reginald H. Jones Center, The Wharton School, University of Pennsylvania.

References

Allen, T. 1977. *Managing the Flow of Technology*. MIT Press, Boston, MA.

Almeida, P., B. Kogut. 1999. Localization of knowledge and the mobility of engineers in regional networks. *Management Sci.* **45**(7) 905–918.

Argote, L., S. L. Beckman, D. Epple 1995. The persistence and transfer of learning in industrial settings. *Management Sci.* **41**(11)

Audretsch, D., P. Stephan. 1996. Company-scientist locational links: The case of biotechnology. *Amer. Econom. Rev.* **86**(3) 641–652.

BioScan. 1994. *Knowledge Access International*, Vol. 63. Mountain View, CA.

Callon, M. 1986. The sociology of an actor-network: The case of the electric vehicle. M. Callon, J. Law, A. Rip, eds. *Mapping the Dynamics of Science and Technology*. Macmillan, London, U.K.

Cameron, A. C., P. Trivedi. 1998. *Regression Analysis of Count Data*. Cambridge University Press, Cambridge, U.K.

Campbell, D. 1974. Evolutionary epistemology. P. A. Schilpp, ed. *The Philosophy of Karl Popper*, Vol. 14-I. Open Court, LaSalle, IL.

Cockburn, I., R. Henderson. 1998. Absorptive capacity, coauthoring behavior, and the organization of research in drug discovery. *J. Indust. Econom.* **46** 157–182.

——, ——, S. Stern. 1999. Balancing incentives: The tension between basic and applied research. Working paper 6882, National Bureau of Economic Research, Cambridge, MA.

——, ——, ——. 2000. Untangling the origins of competitive advantage. *Strategic Management J.* **21** 1123–1145.

Cohen, W., D. Levinthal. 1990. Absorptive capacity: A new perspective on learning and innovation. *Admin. Sci. Quart.* **35**(1) 128–152.

——, R. Nelson, J. Walsh. 2002. Links and impacts: The influence of public research on industrial R&D. *Management Sci.* **28**(1) 1–23.

Crane, D. 1972. *Invisible Colleges: Diffusion of Knowledge in Scientific Communities*. University of Chicago Press, Chicago, IL.

Dasgupta, P., P. David. 1994. Toward a new economics of science. *Res. Policy* **23** 487–521.

David, P. 1998. Common agency contracting and the emergence of "open science" institutions. *AEA Papers Proc.* **88**(2) 15–21.

Deng, Z., B. Lev, F. Narin. 1999. Science and technology as predictor of stock performance. *Financial Analysts J.* **53**(3) 20–32.

De Solla Price, D. 1970. Citation measures of hard science, soft science, technology, and nonscience. C. Nelson, D. Pollock, eds. *Communication Among Scientists and Engineers*. Heath Lexington Books, Cambridge, MA.

Gambardella, A. 1995. *Science and Innovation: The US Pharmaceutical Industry in the 1980s*. Cambridge University Press, Cambridge, U.K.

Gittelman, M. 2000. Mapping national knowledge networks: Scientists, firms and institutions in biotechnology in the United States and France. Unpublished Ph.D. thesis, The Wharton School, University of Pennsylvania, Philadelphia, PA.

Hall, B., A. Jaffe, M. Trajtenberg. 2000. Market value and patent citations: A first look. Working paper 7741, National Bureau of Economic Research, Cambridge, MA.

Henderson, R., I. Cockburn. 1994. Measuring competence? Exploring firm effects in pharmaceutical research. *Strategic Management J.* **15**(Winter Special Issue) 63–84.

Hicks, D. 1995. Published papers, tacit competencies and corporate management of the public/private character of knowledge. *Indust. Corporate Change* **4**(2) 401–424.

Jaffe, A., M. Trajtenberg, R. Henderson. 1993. Geographic location of knowledge spillovers as evidenced by patent citations. *Quart. J. Econom.* **108**(3) 577–598.

Kim, D., B. Kogut. 1996. Technological platforms and diversification. *Organ. Sci.* **7** 293–301.

Knorr Cetina, K. 1999. *Epistemic Cultures: How the Sciences Make Knowledge*. Harvard University Press, Cambridge, MA.

Latour, B., S. Woolgar. 1979. *Laboratory Life: The Construction of Scientific Facts*. Princeton University Press, Princeton, NJ.

Lieberman, M. 1978. A Literature citatiuon study of science-technology coupling in electronics. *Proc. Inst. Electr. Electronics Engr.* **66**(01) 5–13.

Liebeskind, J., A. Oliver, L. Zucker, M. Brewer. 1996. Social networks, learning, and flexibility: Sourcing scientific knowledge in new biotechnology firms. *Organ. Sci.* **7**(1) 428–443.

Merton, R. 1973. *The Sociology of Science*. Chicago University Press, Chicago, IL.

Mowery, D. C., N. Rosenberg. 1982. Technical change in the commercial aircraft industry, 1925–1975. N. Rosenberg, ed. *Inside the Black Box: Technology and Economics*. Cambridge University Press, Cambridge, U.K.

Powell, W. K. Koput, L. Smith-Doerr. 1996. Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. *Admin. Sci. Quart.* **41** 116–145.

Rosenberg, N. 1990. Why do firms do basic research (with their own money)?" *Res. Policy* **19**(2) 165–174.

Shan, W., G. Walker, B. Kogut. 1994. Interfirm cooperation and startup innovation in the biotechnology industry. Res. notes and comm., *Strategic Management J.* **15** 387–394.

Stephan, P. 1996. The economics of science. *J. Econom. Literature* **34**(3) 1199–1235.

Stern, S. 1999. Do scientists pay to be scientists? Working paper 7410, National Bureau of Economic Research, Cambridge, MA.

Stuart, T. J. Podolny. 1996. Networks, knowledge, and niches: Competition in the worldwide semiconductor industry, 1984–1991. *Amer. J. Soc.* **102**(3) 659–689.

Tushman, M. 1977. Specialty boundary roles in the innovation process. *Admin. Sci. Quart.* **22** 587–605.

Van Maanen, J., S. R. Barley. 1984. Occupational communities: Culture and control in organizations. B. M. Staw, L. L. Cummings, eds. *Research in Organizational Behavior*, Vol. 6. JAI Press, Greenwich, CT, 287–365.

Zucker, L., M. Darby, M. Brewer. 1998. Intellectual human capital and the birth of U.S. biotechnology enterprises. *Amer. Econom. Rev.* **88**(1) 290–306.

——, ——. J. Armstrong. 2002. Commercializing knowledge: University science, knowledge capture, and firm performance in biotechnology. *Management Sci.* **48**(1) 138–153.

Accepted by Linda Argote, William McEvily, and Ray Reagans; received March 1, 2001. This paper was with the authors 8 months for 2 revisions.