Strategic Management Journal
Strat. Mgmt. J., **25**: 909–928 (2004)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/smj.384

SCIENCE AS A MAP IN TECHNOLOGICAL SEARCH

LEE FLEMING[1]* and OLAV SORENSON[2]
[1] *Harvard Business School, Boston, Massachusetts, U.S.A.*
[2] *Anderson Graduate School of Management, University of California at Los Angeles, Los Angeles, California, U.S.A.*

A large body of work argues that scientific research increases the rate of technological advance, and with it economic growth. The precise mechanism through which science accelerates the rate of invention, however, remains an open question. Conceptualizing invention as a combinatorial search process, this paper argues that science alters inventors' search processes, by leading them more directly to useful combinations, eliminating fruitless paths of research, and motivating them to continue even in the face of negative feedback. These mechanisms prove most useful when inventors attempt to combine highly coupled components; therefore, the value of scientific research to invention varies systematically across applications. Empirical analyses of patent data support this thesis.

A long line of work supports the idea that scientific research stimulates technological innovation, thereby accelerating economic growth. Though one can trace this notion as far back as Adam Smith (Stephan, 1996), researchers did not begin empirically testing it until the second half of the twentieth century. One set of studies has focused on establishing the relationship between the investments in, or outcomes of, scientific research and their effects on economic growth (e.g., Mansfield, 1972; Rosenberg, 1974; Sveikauskas, 1981). For instance, a paper by James Adams (1990) demonstrates that cumulative research output, in the form of published papers, appears to accelerate growth. Another set—typically drawing on either rich case histories or patent data—has investigated the more proximate linkages between scientific research and technological innovation. Adam Jaffe (1989), for example, finds a positive relationship between university research expenditures and local patenting rates (cf. Jaffe and Trajtenberg, 1996). Together, these complementary lines of research provide strong empirical support for a link between scientific research, technological innovation, and economic growth.

Despite this evidence for the value of scientific research to innovation, inventors across fields vary tremendously in the degree to which they make use of science. Universities, for example, disproportionately patent inventions in two broad sectors: drugs and medicine, and chemicals (Henderson, Jaffe, and Trajtenberg, 1998). Similarly, most studies detailing the beneficial effects of spillovers from academic research to the private sector focus on biotechnology or pharmaceutical firms (e.g., Henderson and Cockburn, 1994; Gambardella, 1995; Zucker and Darby, 1996; Zucker, Darby, and Brewer, 1998).

Why do inventors draw more heavily on scientific research in these areas? Can institutional factors account for the links between academics and

Keywords: science; invention; recombinant search; coupling
* Correspondence to: Lee Fleming, Harvard Business School, Morgan T95, Boston, MA 02163, USA.
E-mail: lfleming@hbs.edu

firms in particular sectors, or do industries differ in the potential returns available to the application of scientific research, thereby affecting the incentives faced by firms? Though institutional factors almost certainly play a role, Mansfield (1995) contends that the benefits of applying science to invention differ across sectors; for instance, pharmaceutical firms in his sample asserted that 27 percent of their inventions required the application of science to avoid costly delays, while electronics firms made the same claim only 6 percent of the time. How do these fields differ such that the application of science appears highly beneficial in one, while of limited use in the next? Answering this question requires us to examine technology at a finer grain of detail; progress in our understanding of the value of science, in the words of Richard Nelson, 'would appear to require a better understanding of what knowledge actually does for inventing' (Nelson, 1982: 455).

To address this issue, this paper investigates the importance of science at the level of the individual invention, explicating one factor—the difficulty of the inventive problem—that might influence the returns to the application of science at that level. We suggest that science proves most useful when inventors seek to combine tightly coupled components. Following a long tradition in the study of innovation (e.g., Gilfillan, 1935; Schumpeter, 1939), we conceptualize invention as a process of searching for better combinations of existing components. When inventors seek to combine relatively independent components (i.e., those with a low degree of coupling), finding useful new configurations proves relatively easy—any search algorithm can locate the most useful combinations. As the space searched becomes increasingly complex, however, local search routines break down, failing to identify the best combinations (Fleming and Sorenson, 2001). In these cases, we posit that science may transform invention from a relatively haphazard search process to a more directed identification of useful new combinations, thus mitigating the complications typically encountered when combining coupled components.

Our empirical analyses estimate the returns to the application of science using patent data. Following a common practice in studying patents, the future citation count provides our measure of the value of the resultant invention. We measure coupling by observing the ease with which subclasses have previously been recombined, and identify the use of science by noting when patents reference published articles in refereed scientific journals. Our results broadly support our expectations: science has no apparent effect when inventors work with relatively independent pieces; it only appears beneficial when inventors seek to combine highly coupled components—a particularly difficult task.

INVENTION AS RECOMBINANT SEARCH

Before considering how science alters the process of invention, one must first ask what inventors actually do. One popular view in the history of technology conceptualizes invention as a process of recombination.[1] Research out of this tradition holds that invention comes either from combining technological components in a novel manner (Gilfillan, 1935; Schumpeter, 1939; Usher, 1954; Nelson and Winter, 1982; Basalla, 1988; Weitzman, 1996), or through reconfiguring existing combinations (Henderson and Clark, 1990). By 'technological components,' we mean any fundamental bits of knowledge or matter that inventors might use to build inventions. For example, Gilfillan (1935) describes the steamship as a combination of the boat with a steam engine, or one might think of the computer workstation as a novel aggregation of several components and subsystems: a CPU, a motherboard, virtual storage and memory, a display, graphics processors, as well as systems and applications software.

Though complex physical systems quite clearly combine multiple elements, many inventions—such as nylon, a polymer—intuitively seem like one component, unbreakable without resorting to atomic-level decomposition. A closer look, however, typically reveals that these inventions also arise from the recombination of discrete components and processes. For example, polymers involve the linkage of several molecules into a substance with new properties (Smith and Hounshell, 1985). Moreover, innovation in these substances frequently occurs in the processes for producing them—often themselves recombinations of existing manufacturing steps. For instance, after the

[1] Other theorists view invention as exogenous shocks (e.g., Klepper, 1996). Though this point of view undoubtedly has validity in many cases, it also relegates the invention process to a black box without first considering whether recombination might usefully characterize the process.

initial synthesis of polymers, Wallace Carothers greatly increased polymer length by applying gentle heat in a molecular still, thereby eliminating interference from water molecules (Smith and Hounshell, 1985). Alternatively, an invention might involve the application of an existing technology to a new purpose. Nylon once again provides a salient example. Though immediately recognized as a replacement for silk, the advent of World War II saw nylon's application to parachutes, flight suits, and glider towropes. Thus, conceiving of inventions as recombinations of existing components does not greatly limit our scope. The question remains, though, how do inventors search for these new inventions?

Local search

Researchers most commonly point to local search —also referred to as exploitation—as the predominant algorithm used in innovation (March and Simon, 1958; Cyert and March, 1963; Hansen and Lovas, 2004; Nelson and Winter, 1982; Stuart and Podolny, 1996). Local search implies that inventors typically alter one component at a time, either reconfiguring it relative to the other components or replacing it with a different component; in other words, inventors search incrementally. Traditional drug discovery, for example, follows this pattern; after identifying a promising candidate through high-throughput screening or luck, researchers attempt to find commercially viable drugs by slightly altering this initial candidate (Drews, 2000). The label 'local' refers to the fact that research activities relate quite closely to prior research activities, by definition implying some experience with the technologies being developed.

Explanations for the prevalence of local search frequently focus on the value of experience. These accounts assume that inventors have an extremely limited understanding of the elements being recombined and the ways in which those elements interact—due to either cognitive limits or fundamental uncertainty (March and Simon, 1958; Nelson, 1982; Vincenti, 1990). As inventors gain experience, they develop some (possibly mistaken) understanding of the components that allows them to invent with greater reliability by avoiding elements that did not work in the past (Vincenti, 1990). Nevertheless, local search also has a downside: focusing on familiar combinations can preclude the inventor from investigating more distant—and potentially more useful—possibilities; as inventors continue to work with a particular set of components, they may exhaust the set of useful combinations (March, 1991; Fleming, 2001).

Empirical research supports the prevalence and effects of local search across a variety of levels. At the firm level, Stuart and Podolny (1996), in an analysis of semiconductor patents, show that firms' new patents concentrate in areas technologically proximate to their existing patent portfolios—a tendency that increases as firms mature, potentially leading to obsolescence (Sørensen and Stuart, 2000; see also Ahuja and Lampert, 2001). Similar results appear in research on new products: Audia and Sorenson (2001) find that computer workstation manufacturers tend to introduce new products with features similar to their existing offerings (cf. Martin and Mitchell, 1998), a practice that retards future sales growth. In research at the community of practice level, Fleming (2001) demonstrates that experience at this level does indeed appear to enhance search, both by increasing the average value of inventions and by decreasing the variability of outcomes, but this beneficial effect eventually succumbs to exhaustion.

Science and search

Scientific knowledge, by contrast, may lead to a very different type of search; in effect, it provides inventors with the equivalent of a map—a stylized representation of the area being searched. Though local search portrays inventors as rummaging around, blindly bumping along in the search for new technologies, they can also proceed with greater foresight (Vincenti, 1990). For example, studies of engineers have found that they sometimes consult the scientific literature to resolve technological difficulties (Gibbons and Johnston, 1974; Allen, 1977). Scientific knowledge differs from that derived through local search, in particular, because the scientific endeavor attempts to generate and test *theories*. Science attempts to explain *why* phenomena occur, providing a means of predicting the results of untried experiments and the usefulness of previously uncombined configurations of technological components. Having an understanding of the fundamental problem—a map—likely modifies the search process in multiple, complementary ways. Notably, it might lead inventors quite directly to the proper combinations of components to solve a particular technical problem. Even absent such a direct linkage, science could usefully increase the effectiveness of search by identifying useless directions of search beforehand, and by providing a glimpse of the possible.

Theoretical understanding of the underlying properties of technological components and their interactions may facilitate effective search. Lippman and McCall (1976) argue that this advantage stems from the ability to assess alternatives 'offline'—in essence, running the experiment in one's mind or on paper without incurring the costs of actually performing the test. For example, laying the groundwork for a future generation of chip technology, researchers predicted that single-wall carbon nano-tubes would either conduct or semiconduct, depending on the diameter of the tube and the angles of the carbon bonds (Dresselhaus, 2001). This prediction pointed to two specific configurations (out of an essentially infinite number of possibilities) that resulted in the successful fabrication of conducting and semiconducting nano-tubes (Ouyang *et al.*, 2001). The theory did not predict everything and researchers must still engage in trial-and-error search to produce working transistors. Theoretical guidance nonetheless accelerated the researchers' arrival at the proper diameter and angles. Maps typically do not detail every rock and crevice dotting the landscape; similarly, scientific theories do not predict every property and interaction of a new combination of technological components. Theory may still get researchers 'in the ballpark,' thereby leading them to a promising new invention more efficiently.

By eliminating fruitless avenues, science may also play the more modest role of reducing the size of the combinatorial search pace (Nelson, 1982). In other words, science can tell inventors how to avoid wasted effort. Consider the classic goal of alchemy: transforming lead into gold. Once people understood the structure and properties of atoms, it became quite clear that anything short of a nuclear reaction would not yield the desired results. More recently, researchers at Lord Manufacturing Company switched from working with electro-rheological materials to magneto-rheological materials in their effort to develop controllable fluids. Even after 10 years of fruitless effort with electrical approaches, the switch occurred only after the lead researcher, Dave Carlson, investigated the physics of the materials and calculated that the potential power available from the use of magnetic fields exceeded the potential of electrical fields by two orders of magnitude (Carlson, 2001; Stix, 2001). Without realizing the futility of the search, researchers at Lord may have pursued an unobtainable goal indefinitely; thus, science improves the efficacy of search by preventing inventors from wasting valuable effort on the impossible.

Even when science has an inaccurate or incomplete understanding of the problem, it might still alter the search process in a useful manner. When inventors experiment with combinations and configurations of components, most of the iterations that they try will fail to yield useful outcomes. In the absence of a predictive model of the problem, inventors constantly face the question of how long they should persist with a failing approach. Should they quit after 10, 100, or 1000 failures? By suggesting that an avenue of solution might succeed theoretically, science can encourage inventors to continue working with a seemingly unfruitful set of components. Consider the case of Prozac. Bryan Molloy began Eli Lilly's research with local search, deriving hundreds of benadryl compounds (Kramer, 1997: 62). They all failed. His colleague David Wong continued the search, however, because he believed a theory that blocking serotonin uptake, a property of benadryl derivatives, held the key to treating depression. Using a new experimental technique that enabled more accurate evaluation, Wong tested the same compounds as Molloy and demonstrated the efficacy of one of them, fluoxetine (the chemical name for Prozac). In addition to motivating inventors in the face of failure, the belief that something better might exist could also lead them to continue searching even after having found acceptable solutions (Gavetti and Levinthal, 2000, provide a similar argument for the usefulness of managerial frameworks in strategic search).

Understanding the strengths and weaknesses of these search processes also requires an understanding of the spaces being searched; thus, we turn to a discussion of technological landscapes.

TECHNOLOGY LANDSCAPES

Landscapes offer a useful heuristic for thinking about the space that inventors must search when attempting to discover useful new inventions. Wright (1932) first introduced the idea of a fitness landscape to describe the evolution of traits in species. Since then, the idea has been applied usefully to topics ranging from spin-glasses in physics (Weinberger, 1991) to organizational routines (Levinthal, 1997; Rivkin, 2000). This same metaphor can help us understand technological search. Think of each possible set of components as corresponding to a particular technological landscape that inventors search; positions on these landscapes represent different configurations of components. Inventors search new landscapes when they combine new sets of components and move across these landscapes when they reconfigure a particular set of components. The peaks and higher elevations represent more useful configurations. Technological landscapes can take a variety of shapes. At one extreme, the landscape might smoothly slope up to a single peak; at the other, it might jump around with treacherous valleys and soaring zeniths. Thus, one might ask, what determines the topography of these technological landscapes?

One important factor influencing the shape of these landscapes is coupling. By 'coupling,' we mean the degree to which components interact in determining the functionality of the overall invention.[2] A rugged landscape implies that adjacent points of the terrain differ dramatically in their usefulness. Coupled technologies exhibit precisely such a characteristic: the functionality of coupled inventions overall becomes highly sensitive to minor changes in the individual components (Ulrich, 1995). For example, a change of one particle in 10 million of semiconductor dopant can change its conductivity by a factor of 10 thousand (Millman, 1979).[3] In contrast, independent components—those with no coupling—gradually change the functionality of the overall system, with each component contributing individually to overall usefulness. Kauffman (1993) presents this relationship in stark simplicity. His simulations include two parameters: the number of components and the degree of interaction between those components. These simulations characterize the landscapes produced by different types of components; interacting components create more jumbled, uncorrelated landscapes, while independent elements generate a landscape with a single peak. In short, the ruggedness of the technological landscape increases as the degree of coupling among the components intensifies.

Landscapes and search

Combining these landscapes with a search algorithm allows us to make predictions regarding the likely outcomes of inventive search. Consider local search. When inventors search locally, they move across adjacent positions on the fitness landscape. Although gradually climbing uphill will eventually bring the inventor to an optimum, this peak might not represent the 'best,' or even a particularly 'good,' configuration. Rather, the topography of the landscape importantly affects the efficacy of the search algorithm. Local search processes work best on smooth, correlated landscapes because incremental improvements more frequently find the global maximum. In contrast, local search processes operate less effectively on jumbled, uneven, and uncorrelated landscapes. On these landscapes, local search processes yield unpredictable outcomes and often trap the searcher on a local peak because any incremental change yields less desirable outcomes (March, 1991; Kauffman, 1993). Inventors who search incrementally with strongly coupled components make slow progress and have no assurance of ultimate success.

Although coupling between components makes the search process more difficult, it also creates the potential for very useful inventions. On flat landscapes, inventors searching locally can find the optima, but few exist and those that do rise only modestly above alternative configurations (i.e., they offer minimal gains in usefulness). Competitors may also find these innovations easy to imitate (Rivkin, 2000). Rugged landscapes contain a large number of potentially useful configurations, if inventors can only find them. Together these processes generate a non-monotonic relationship between the usefulness of

[2] This idea goes by a variety of monikers in different literatures. Milgrom and Roberts (1990) refer to it as complementarity. Kauffman (1993) and Sorenson (2002) call it interdependence, and many use the term modularity to refer to the inverse of interdependence (e.g., Baldwin and Clark, 2000). Given our technological context, we prefer Ulrich's (1995) definition of coupling: 'Two components are coupled if a change made to one component requires a change to the other component for the overall product to work correctly.'

[3] Other examples abound. The fluoxetine molecule resembles other benadryl derivatives quite closely, yet it provides much more effective relief against depression (Kramer, 1997). Superconductor research continues to find unexpectedly high temperature breakthroughs with novel combinations of common materials, such as boron (Campbell, 2001).

inventions found through local search and the coupling of the components combined (Fleming and Sorenson, 2001). On flat landscapes—those with little coupling—local search can locate the highest peak, but that apex offers limited functionality. At low levels of coupling, somewhat rugged landscapes provide a larger number of taller peaks that inventors can still find through local search. As the coupling increases, however, the difficulty of search rises, and the average usefulness of inventions declines. Coupling also increases the uncertainty of invention. As the number of unpredictable interactions rises, the uncertainty, or variability, of the expected outcomes also increases. Fleming and Sorenson (2001) find strong support for this model, using patent citations as a measure of invention usefulness.

By contrast, inventors can proceed quite differently when science provides them with some understanding of the underlying landscape. The exact effect of this process on inventive outcomes depends on how science influences the search process. Consider the first mechanism: cheap offline experimentation. If science allows a rough prediction of the expected interactions between coupled components, it should allow inventors to move quite directly toward the highest peaks—the most useful configurations—on the landscape, thereby reducing the variability of inventive outcomes. As the potential for useful inventions increases with coupling, it should also allow inventors to exploit synergies more effectively; thus, the average usefulness of inventions should rise with coupling when inventors use science. It might also reduce the number of times that inventors use the same combinations of components since they can efficiently focus on the most useful configurations while avoiding the rest.

Even if science plays the more modest role of allowing inventors to rule out unfruitful directions, our second mechanism, this narrowing of options should have a similar—though less pronounced—beneficial effect. One would still expect an improvement in the average usefulness of inventions, as researchers avoid the worst regions of the solution space. The variability of outcomes would likely remain high, however, as even the best regions in fitness landscapes of highly coupled components include peaks and valleys (Kauffman, 1993).

Under the third mechanism, motivation, the expected effects of science on search may differ. Having some sense of the height and location of the tallest peak may lead inventors to better outcomes by preventing them from becoming trapped in local optima (Gavetti and Levinthal, 2000). However, in this case, inventors might actually search *more* when directed by science because they will continue to look for better alternatives even after finding 'good' ones. In the words of Gavetti and Levinthal (2000), they may begin 'chasing cognitive rainbows.'

EMPIRICAL ANALYSIS

Studying how science affects the process of invention can prove quite difficult and, as a result, much of the work on this topic relies on careful case studies. Though useful, these studies lack the statistical power necessary to distinguish potential patterns linking science and invention from random noise. Thus, we analyzed patent data to examine the role of science.

Patent data admittedly offer imperfect measures of invention. Companies sometimes avoid patenting, and industries vary in their propensities to patent (Levin *et al.*, 1987). Patents also do not allow us to observe all points on the fitness landscape. Because inventors may limit their patent applications to their most successful inventions, patents likely represent only the higher positions on these landscapes. This implies that we must infer the topography of the underlying landscape from truncated data.[4] Despite these imperfections, patents do offer a window into the process of invention, allowing us to develop a quantitative measure of fitness across a broad range of technologies, to identify those inventions in which science likely played a role, and to develop a measure of component coupling—the variables needed to investigate our proposition empirically. We collected information on all U.S. patents granted in May and June of 1990 ($n = 17{,}264$)[5] listed in the Micro-Patent product; from these, we excluded the

[4] Simulations demonstrate, however, that increases in the mean of a normal distribution generate corresponding increased means in right truncated observations; thus, these truncated observations should still allow us to observe the rough structure of the technology landscapes (Fleming and Sorenson, 2001).

[5] We chose 1990 to allow sufficient time to observe future citation patterns, our dependent variable; we used a random number generator to select the starting month, May, and included only 2 months of data due to resource limitations (calculation of the coupling measure required months of CPU time and the

442 patents that do not cite any prior art (previous work in the area) because these patents may lie beyond our theory, involving the discovery of fundamental new components rather than the recombination of previous technologies (their inclusion does not change the results). Thus, our analyses include 16,822 patents granted in May and June of 1990.

Key variables

Science

We began by identifying when science likely influenced the process of invention. Much of the work linking science to invention focuses only on research that occurs at a university or public research institution (e.g., Henderson *et al.*, 1998). Since science also gets applied outside the university context (Narin, Hamilton, and Olivastro, 1997; Cockburn, Henderson, and Stern 2000), we took a different approach, identifying the usage of science through non-patent references (Narin *et al.*, 1997; Perko and Narin, 1997).[6]

We took a reference to the scientific literature as an indication that the inventor made use of—or at least had some awareness of—scientific knowledge in the process of invention. In addition to citing the prior art (previous patents), U.S. patents also cite a variety of non-patent literature. Our sample made 16,698 references to the non-patent literature (Table 1). The majority of these non-patent references (67%) cite journals listed in the *Scientific Index*: a publication that covers peer-reviewed scientific journals.[7] Unlike references to patent prior art, often added by the patent examiner, non-patent references more frequently come from the inventors themselves; Tijssen (2001) reports that 81 percent of patents with non-patent references include an inventor's citation to their own scientific work. Thus, non-patent references provide a reasonable indicator of the influence of science; as Tijssen notes, 'these citations at the very least indicate an awareness of scientific results with some indirect bearing on elements of the invention. In the best case, they reflect strong evidence of substantial direct contributions of scientific inputs to breakthrough technological innovations' (Tijssen, 2001: 52).[8] A dummy variable indicates if a patent referenced one of these publications.[9]

coding of the non-patent references involved hundreds of hours of research assistance).

[6] Though Schmoch (1993) and Meyer (2000) question the use of the non-patent references as indicators of the direct application of science, their contention that these references simply denote a relationship between the technology and science fits well with our assumption of their meaning.

[7] References to science most commonly appear in the following classes: 505 Superconductor technology (85%), 530 Natural resins and derivatives (81%), 435 Molecular biology (78%), 512 Perfumes (71%), 548 Organic compounds (69%).

[8] In our own survey of inventors (detailed below), 62 percent of inventors indicated an awareness of the specific scientific papers cited on their patents and 71 percent reported a awareness of the broader scientific literature on the subject. Though our results suggest a weaker link than Tijssen (possibly because we draw a representative sample of all patents while Tijssen selects on patents citing Dutch science), non-patent references to the scientific literature still appear a good proxy for the availability of scientific knowledge to inventors.

[9] Using a count of the number of scientific articles referenced in place of this dummy variable produced qualitatively equivalent results; including both a dummy and the count reveals that the count adds no significant information to the models. We

Table 1. Descriptive statistics of references to non-patent literature

	Number of patents making references	Conditional statistics			
		Mean	S.D.	Min.	Max.
Scientific Index journal (reference to science)	2,919	3.56	4.90	1	75
Non-Index journal (non-science reference)	290	1.56	1.19	1	8
Corporate non-technical (non-science reference)	700	2.11	2.42	1	34
Corporate technical	331	1.91	2.26	1	27
Book	870	1.59	2.24	1	55
Technical report	320	1.54	1.30	1	12
Conference proceedings	465	1.66	1.30	1	11

Coupling

Next we characterized the difficulty of the inventive problem according to the degree of coupling among the components. Our measure essentially observes the degree to which an invention's components have been previously recombined.[10] To calculate this measure, we used the subclass references on each patent. In effect, we treated the subclass assignments as proxies for the underlying components that combine to create that invention. The U.S. Patent Office uses subclass references to indicate which technologies relate to the patent, developing and updating these subclasses such that they consistently track technology back to 1790.[11] The approximately 100,000 subclasses allow for fine-grained classification of inventions (Trajtenberg, Henderson, and Jaffe, 1997). Though in many cases subclasses correspond quite closely to physical components (such as the example in Figure 1), they do not always match so well. Our measure, however, only requires that these subclasses define pieces of knowledge rather than identifiable physical components. Combining some pieces which interact sensitively to each other proves more difficult than connecting relatively independent chunks of knowledge.

We calculated our measure of coupling in two stages. Equation 1 details our measurement of the observed ease of recombination, or inverse of coupling, of an individual subclass i used in patent j. The score increases as a particular subclass combines with a wider variety of subclasses, controlling for the total number of applications; thus, it captures the observed ease with which a component has been recombined. Since most patents (92%) belong to more that one subclass of technology, we averaged the inverse of these subclass scores to create a patent-level measure

also investigated classifying technical references together with *Scientific Index* journal cites as indicators of science; using this more inclusive definition yielded weaker, but substantively identical, results.

[10] Economists seeking to identify complementarity—a closely related concept—have also proposed using the realized combinations of activities to identify the pattern and strength of these interdependencies (Athey and Stern, 1998).

[11] The Patent Office often redraws the boundaries of patent subclasses and retroactively reassigns patents to subclasses based on these changes. This system has the advantage of making subclasses usable over time; however, this reclassification likely introduces noise into our measure of coupling, making our tests of its effects more conservative. We used the concordance of 1996 for our calculations.

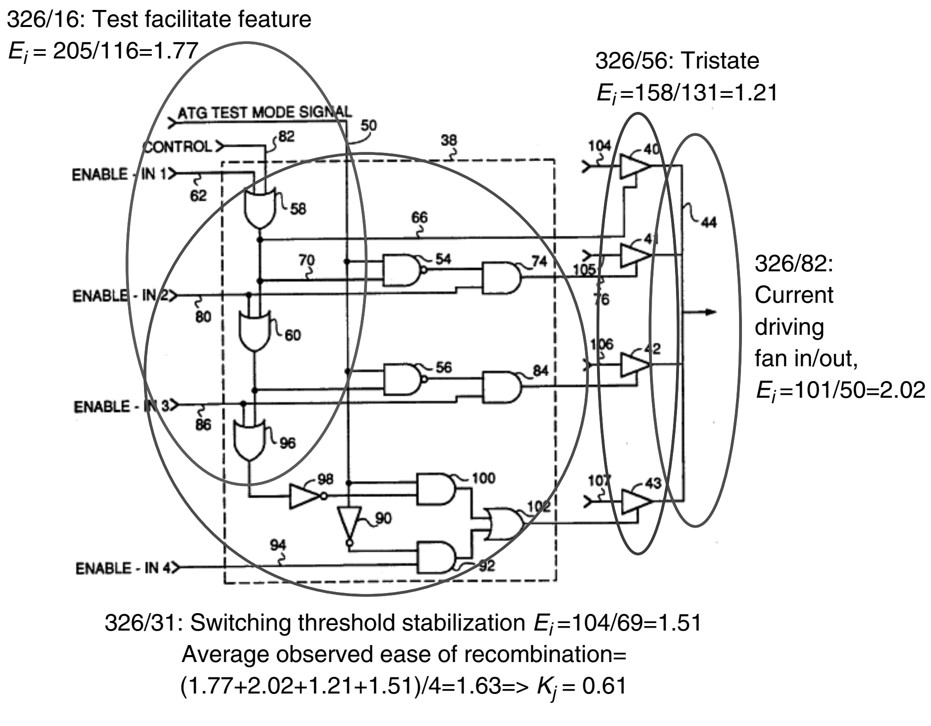

Figure 1. Calculation of K for patent 5,136,185

(Equation 2). We computed this measure using the entire 200 years of pre-sample history and tested its robustness to temporal bias with a 10-year window.

$$\text{Observed ease of recombination of subclass } i \equiv E_i = \frac{\text{Count of subclasses previously combined with subclass } i}{\text{Count of previous patents in subclass } i} \quad (1)$$

$$\text{Coupling of patent } j \equiv K_j = \frac{\text{Count of subclasses on patent } j}{\sum_{j \in i} E_i} \quad (2)$$

As an example of the coupling calculation, consider one of the first author's patents, #5,136,185. Figure 1 illustrates calculation of the measure and the correspondence between the USPTO classification scheme and the components used, all standard elements described in digital design textbooks at the time of invention (see, for example, McCluskey, 1986). 326/16 identifies the 'Test facilitate feature' subclass, which implements a testing mode within a semiconductor chip. Prior to the author's use of this component (i.e., subclass), it had been recombined 116 times with 205 other components, implying an observed ease of recombination score of 205/116 = 1.77. 326/56 indicates the 'Tristate subclass', and 326/82 points to 'Current driving fan in/out.' 326/31 identifies the 'Switching threshold stabilization' subclass (essentially a priority encoder). Figure 1 illustrates the location of these components on the circuit, and the calculation of their ease of recombination scores, as well as the calculation of the patent's coupling score of 0.61. This score, slightly below the mean value of 0.63, seems reasonable as digital hardware recombines with relative ease.

To validate the measure across a wider range of technologies, we surveyed inventors about the coupling of the components of their inventions. Randomly choosing 0.2 percent of U.S. patents granted to U.S. inventors in the year 2000 gave us a sample of 199 patents. Of these, we located current contact information for the patent holders for 189, from which we received responses in 64 cases. To assess the coupling between an invention's components, we asked (from Ulrich, 1995), 'Modules are said to be *coupled* when a change made to one module requires a change to the other module(s) in order for the overall invention to work correctly. How coupled were the modules of your invention?' The Pearson correlation coefficient between the responses and our measure of coupling is 0.30 ($p \sim 0.03$); though not perfect, the relationship is highly significant.[12] Most importantly, our automated procedure allows us to analyze a large set of patents, and appears to provide at least a rough estimate of coupling.

Invention usefulness

We defined usefulness as the number of citations that a patent receives in the 5 years following its grant date.[13] Each patent by law must cite previous patents that relate closely to its own technology. Research demonstrates that the number of citations a patent receives correlates highly with its technological importance, as measured by expert opinions, social value, and industry awards (Trajtenberg, 1990; Albert *et al*., 1991). It also corresponds closely to the economic value of an invention (Harhoff *et al*., 1999, Hall, Jaffe, and Trajtenberg, 2000). Thus, citation counts offer a means of measuring inventive usefulness across a broad range of technologies.[14] Descriptive statistics for these variables, as well as the others used in the models, appear in Table 2.

Science, coupling, and citations

One can see the relationship between science, coupling, and the usefulness of the inventions even without running a complex statistical analysis. Figure 2 depicts this relationship graphically: the points plot the mean number of citations received against the average level of coupling for each quintile of the coupling distribution. It plots this relationship separately for patents that cite a scientific

[12] While we provided inventors with specific definitions of how they should think about their subsystems and components, they may still have placed different boundaries in the organization of their inventions. This would increase the error of our measure. Increased error, however, should only dampen our estimates of the effects of coupling, making our empirical tests more conservative.

[13] We chose this period to capture the bulk of the citations to a patent, as citations typically peak about 3 years after the grant date (Jaffe, Trajtenberg, and Henderson, 1993). Although citation rates to patents appear quite consistent over time, we tested the sensitivity of this assumption by also using a 10-year window of citation counts; the results remain robust across both.

[14] The propensity to cite does vary across technologies as a function of the level of activity in that technology. We introduce multiple methods for controlling for this heterogeneity below.

Table 2. Descriptive statistics of the variables used in the models

	Full sample $N = 16{,}822$		Cite science $N = 2{,}919$		Do not cite science $N = 13{,}903$	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Cites	2.75	3.54	3.43	4.25	2.60	3.34
Mean technology control	1.19	0.41	1.27	0.45	1.17	0.39
Number of prior art cites	7.63	6.99	7.90	8.66	7.27	6.04
Single subclass dummy	0.08	0.27	0.08	0.28	0.08	0.27
Number of subclasses	4.31	3.31	4.97	4.86	4.02	2.79
Number of classes	1.78	0.95	1.90	1.06	1.74	0.91
Number of trials	2.94	15.23	2.55	18.01	2.98	14.07
Recent technology	3.96	0.65	4.17	0.49	3.92	0.67
Science stock	0.21	0.22	0.45	0.24	0.16	0.17
Reference to science	0.18	0.38	1.00	0.00	0.00	0.00
Coupling	0.63	0.35	0.50	0.28	0.66	0.36

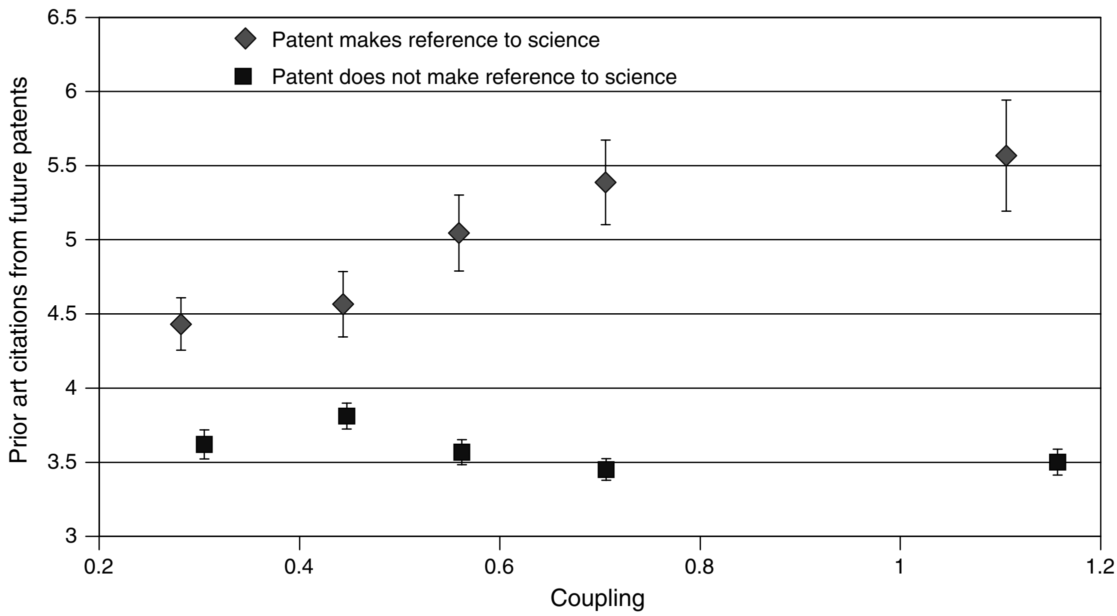

Figure 2. Mean citations (±1 S.E.M.) across quintiles of the coupling variable

article and for those that do not. Patents that reference science receive more cites on average. More importantly, the figure illustrates the predicted positive interaction effect. One can clearly see that while the relationship between coupling and citations first rises and then falls for patents that do not cite science, those that do reference an article in the *Scientific Index* show a positive relationship between coupling and citations. Even at this cursory level, the data support the claim that science helps most when inventors attempt a difficult recombination task.

Multivariate analysis

To investigate the process in more detail and to control for a variety of factors, we estimated fixed effects negative binomial models of future citations. The dependent variable of citation counts takes on only whole number values (i.e. 0, 1, 2, 3...). Using linear regression on such data can yield inefficient, inconsistent, and biased coefficient estimates. Count models can avoid these problems (Cameron and Trivedi, 1998). We employed one particular variant suggested by Hausman, Hall, and Griliches (1984): the fixed effects negative binomial. This procedure usefully allows for overdispersion—where the variance exceeds the mean—which these data exhibit. It also allows us to control broadly for differences across technological domains by defining fixed effects at the level of the USPTO class.[15] In this case, using the fixed effects negative binomial amounts to taking the number of citations received by all patents in a particular class as a given and estimating which factors predict the distribution of those citations among the patents within the class.

Several measures captured other potential sources of heterogeneity. First, we developed a *technology control*, which essentially refines the fixed effects. In the first stage, we calculated the average number of citations that each patent in a particular USPTO class received from patents granted between January of 1985 and June of 1990 (Equation 3).[16] We weighted these parameters according to the patent's class assignments (Equation 4), where p indicates the proportion of patent k's subclass memberships that fall in class i.

$$\text{Average citations in patent class } i \equiv \mu_i = \frac{\sum_{j \in i} \text{Citations}_j \text{ (before 7/90)}}{\text{Count of patents } j \text{ in subclass } i} \quad (3)$$

$$\text{Technology mean control patent } k \equiv M_k = p_{ik}\mu_i \quad (4)$$

To augment our technology control, we included several additional variables. We incorporated a count of *the number of major classes*. Patents that cover a broad range of technologies may carry a higher risk of being cited simply because future inventions from each field might potentially cite them—similar to what happens when an academic article spans several literatures. This measure also controls for the interdisciplinary breadth of the invention.[17] We also included the *number of prior art citations*—references to prior patents assigned by the U.S. Patent Office—as a control for the degree of local search (Podolny and Stuart, 1995); it may also capture idiosyncratic differences in citation propensity that our technological class controls miss. We additionally measured previous local search as the *number of repeated trials* on a particular landscape (Fleming and Sorenson, 2001). Specifically, we counted the number of previous patents that combined exactly the same set of subclasses.

We control for the number of components or problem domains in an invention by including the *number of subclasses*. Eight percent of the patents in our data belong to only one subclass. Although we suspect that these inventions also arise from a process of recombination—as the prior art references in these patents reveals, they do build on previous inventions—this combination occurs at a finer grain than our measures can capture. Therefore, we added a *single subclass dummy* variable to capture any systematic differences between these inventions and those assigned to multiple subclasses. A *recent technology* control accounts for any differences in citation patterns for technologies closer to the 'cutting edge' (Katila, 2002); we calculated this variable by averaging the patent numbers of the prior art cited.[18] Finally, we controlled for the cumulative *science stock* in a field of technology by calculating the percentage of patents up to 1990 that cite a non-patent reference in the same primary subclass as the focal patent. Table 3 presents the correlations between these variables, as well as those of theoretical interest.

The results of the fixed effects negative binomial models appear in Table 4. Model 1 provides a baseline. After controlling for a variety of factors, strong coupling generates an even stronger negative effect on citations than suggested by Figure 2. Science clearly has a positive effect on citation counts. The fact that this effect persists while controlling for the science stock reveals that these benefits accrue specifically to the patents drawing

[15] In unreported analyses, we also defined fixed effects at the level of the inventing firm, with qualitatively equivalent results. Thus, firm-level differences cannot account for our results.

[16] We allow all patents issued between January 1985 and June 30, 1990 to enter the estimation of the technology control, meaning that the patents used to calculate it vary in the time during which they can receive citations. Alternatively, we could select a small set of patents from 1985 and base the measures on the subsequent 5 years of citations; however, this approach would ignore the patent activity just prior to our sample.

[17] Unreported models indicated that coupling actually decreases the likelihood that a patent receives citations from other classes. Hence, our measure of coupling does not appear to pick up the breadth of application.

[18] Because patenting occurs at a relatively constant rate, this measure correlates at roughly 0.99 with the average age of the patent's prior art measured in terms of days.

Table 3. Correlation matrix

Correlation matrix	Cites	Mean	Priors	Single	Subclass	Classes	Trials	Recent	Science
Mean technology control	0.298								
Number of prior art cites control	0.117	0.020							
Single subclass dummy control	−0.056	0.002	−0.065						
Number of subclasses control	0.110	0.013	0.079	−0.288					
Number of classes control	0.074	−0.021	0.063	−0.244	0.513				
Number of trials control	−0.028	−0.028	−0.007	0.448	−0.168	−0.143			
Recent technology	0.208	0.377	−0.160	−0.005	0.066	0.044	−0.012		
Reference to science	0.095	0.089	0.018	0.006	0.107	0.061	−0.012	0.146	
Coupling	−0.029	−0.057	0.017	0.194	−0.309	−0.275	0.364	−0.135	−0.178

Table 4. Negative binomial estimates of citation counts (5-year window, standard errors in parentheses)[a]

	Model 1 baseline	Model 2	Model 3 10-year variable	Model 4
Mean technology control	**0.246**	**0.245**	**0.212**	**0.246**
	(0.036)	**(0.036)**	**(0.036)**	**(0.036)**
Number of prior art cites	**0.015**	**0.015**	**0.015**	**0.015**
	(0.001)	**(0.001)**	**(0.001)**	**(0.001)**
Single subclass dummy	**−0.186**	**−0.187**	**−0.201**	**−0.185**
	(0.036)	**(0.036)**	**(0.036)**	**(0.038)**
Number of subclasses	**0.027**	**0.027**	**0.029**	**0.027**
	(0.003)	**(0.003)**	**(0.002)**	**(0.003)**
Number of classes	**0.059**	**0.059**	**0.062**	**0.059**
	(0.009)	**(0.009)**	**(0.009)**	**(0.009)**
Number of trials	−0.000	−0.000	−0.001	−0.000
	(0.001)	(0.001)	(0.001)	(0.003)
Recent technology	**0.316**	**0.316**	**0.290**	**0.316**
	(0.016)	**(0.016)**	**(0.016)**	**(0.016)**
Science stock in subclass	0.083	0.085	0.073	0.083
	(0.052)	(0.051)	(0.051)	(0.052)
Cite to scientific publication	**0.098**	0.105	0.079	**0.098**
	(0.022)	(0.058)	(0.052)	**(0.022)**
Coupling	**0.462**	**0.485**	**0.632**	**0.471**
	(0.071)	**(0.076)**	**(0.079)**	**(0.078)**
Coupling^2	**−0.116**	**−0.136**	**−0.155**	**−0.121**
	(0.028)	**(0.030)**	**(0.036)**	**(0.032)**
Coupling × cite to scientific publication		−0.097	−0.023	
		(0.127)	(0.125)	
Coupling^2 × cite to scientific publication		**0.118**	**0.099**	
		(0.049)	**(0.051)**	
Number of trials × coupling				−0.000
				(0.003)
Number of trials × coupling^2				0.000
				(0.001)
Constant	**−1.891**	**−1.893**	**−1.783**	**−1.894**
	(0.084)	**(0.085)**	**(0.078)**	**(0.085)**
Fixed effects	363 Classes	363 Classes	363 Classes	363 Classes
Log-likelihood	−33,619.4	−33,614.5	−33,571.3	−33,619.4
N	16,822	16,822	16,822	16,822

[a] Bold type indicates that the coefficient estimate differs signification from zero with 95% confidence.

on science, rather than to all patents in a particular technological area. The primary test of the role of science appears in Model 2; this model interacts coupling with whether or not the patent cites a journal in the *Scientific Index*. The interaction significantly improves the model ($\chi^2 = 9.8$, d.f. $=$ 2)[19] and shows that while coupling has a non-monotonic relationship with citations for patents that do not cite science, the relationship remains positive for patents citing science. Because of this fact, the benefits to science increase with coupling: at the mean level of coupling, patents that reference science receive 9.5 percent more citations than those that do not; at one standard deviation above the mean on coupling, the advantage increases to 13.1 percent; and at two standard deviations it reaches 20.3 percent. Model 3 tests the robustness of our measure of coupling by substituting a measure based only on the 10-year window preceding our sample rather than on the entire history of the U.S. Patent Office. The relationship between science and the effects of coupling remain the same.

Still, one might also ask whether scientific knowledge influences the process of search in a different manner from empirical knowledge (experience). Model 4 investigates this proposition by interacting the number of trials—a measure of the technological community's experience with a set of components—with coupling. The results reveal that, unlike science, experience does not mitigate the detrimental effects of high coupling, strongly corroborating the notion that search with the aid of science differs fundamentally from local search. Unreported models also attempted interactions, following Podolny and Stuart (1995), with the number of prior art citations as an indicator of local search; these models similarly failed to find an interaction between local search and interdependence.

If science helps inventors predict the results of their recombinant search, it should also reduce the variance of search outcomes—the uncertainty of invention. Thus, we analyzed the residuals from the negative binomial models to determine whether science reduced the variation in the usefulness of inventive outcomes. Table 5 presents ordinary least squares estimates of the absolute value of the residuals from Model 2.[20] We included the number of future citations in the model to account for the relationship between the mean and the variance of a count variable. As one would expect if science helps direct inventors, science reduces the dispersion of outcomes as the degree of coupling rises. At the mean level of coupling, the application of science reduced variability by 6.5 percent; at one standard deviation above the mean, this reduction in risk increases to 9.0 percent.

Table 5. OLS analysis of absolute residuals (standard errors in parentheses)[a]

	Model 5
Citation count	**0.967**
	(0.001)
Coupling	**0.068**
	(0.007)
Cite to scientific publication	−0.008
	(0.013)
Coupling × cite to scientific publication	**−0.257**
	(0.022)
Constant	**−0.076**
	(0.006)
R^2	0.99

[a] Bold type indicates that the coefficient estimate differs signification from zero with 95% confidence.

Robustness checks

Some additional robustness checks appear in Table 6. Models 6 and 7 examine the importance of the definition of the dependent variable. In particular, Model 6 excludes self-citations (when the future citing patent belongs to the same assignee) from the dependent variable. Self-citations may differ in their meaning from citations by competitors—notably Sørensen and Stuart (2000) interpret self-citations as an indicator that the organization's patents may not fit the current state of the market; thus, self-citations have an ambiguous role as a measure of patent importance. The results hold, however, even when excluding self-citations from the dependent variable. Model 7 increases to 10 years the length of the future

[19] Haberman's (1977) chi-squared, two times the difference in log-likelihoods, provides a test of the statistical significance of the difference between two nested models.

[20] In unreported analyses, we estimated the mean and variance effects simultaneously, using the variance decomposition negative binomial (King, 1989). Simultaneous estimation produced equivalent results. The variance decomposition negative binomial, however, has the disadvantage of not being able to incorporate fixed effects; thus, we report the two-stage estimation.

Table 6. Negative binomial estimates of citation counts (standard errors in parentheses)[a]

	Model 6 no self-citations	Model 7 10-year citations	Model 8	Model 9 Selection
Mean technology control	**0.256**	**0.135**	**0.248**	**0.242**
	(0.038)	**(0.028)**	**(0.036)**	**(0.036)**
Number of prior art cites	**0.014**	**0.015**	**0.015**	**0.014**
	(0.001)	**(0.001)**	**(0.001)**	**(0.001)**
Single subclass dummy	**−0.181**	**−0.176**	**−0.188**	**−0.200**
	(0.038)	**(0.031)**	**(0.036)**	**(0.037)**
Number of subclasses	**0.029**	**0.024**	**0.026**	**0.027**
	(0.003)	**(0.002)**	**(0.003)**	**(0.003)**
Number of classes	**0.061**	**0.050**	**0.059**	**0.059**
	(0.010)	**(0.008)**	**(0.009)**	**(0.009)**
Number of trials	−0.000	−0.000	−0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)
Recent technology	**0.275**	**0.218**	**0.318**	**0.297**
	(0.016)	**(0.013)**	**(0.016)**	**(0.019)**
Science stock in subclass	**0.145**	−0.013	0.068	0.087
	(0.054)	(0.045)	(0.052)	(0.052)
Cite to scientific publication	0.085	0.062	**0.097**	0.118
	(0.062)	(0.053)	**(0.022)**	(0.060)
Cite to non-scientific publication			0.106	
			(0.112)	
Coupling	**0.423**	**0.418**	**0.459**	**0.441**
	(0.080)	**(0.064)**	**(0.072)**	**(0.100)**
Coupling^2	**−0.123**	**−0.121**	**−0.112**	**−0.109**
	(0.032)	**(0.025)**	**(0.028)**	**(0.035)**
Coupling × cite to scientific publication	−0.072	−0.016		−0.133
	(0.136)	(0.119)		(0.127)
Coupling^2 × cite to scientific publication	**0.108**	**0.094**		**0.121**
	(0.053)	**(0.048)**		**(0.047)**
Coupling × cite to non-scientific publication			0.044	
			(0.254)	
Coupling^2 × cite to non-scientific publication			−0.078	
			(0.112)	
Selection (likelihood of citing scientific publication)				**−0.469**
				(0.239)
Selection × coupling				**2.045**
				(0.609)
Selection × coupling^2				**−1.104**
				(0.403)
Constant	**−1.785**	**−1.304**	**−1.898**	**−1.837**
	(0.089)	**(0.069)**	**(0.084)**	**(0.091)**
Fixed effects	363 classes	363 classes	363 classes	363 classes
Log-likelihood	−31,411.3	−46,458.2	−33,613.9	−33,607.8
N	16,822	16,822	16,822	16,822

[a] Bold type indicates that the coefficient estimate differs signification from zero with 95% confidence.

citation window observed. If knowledge diffuses at different rates across technologies, and these rates correlate to the usage of science, this process might influence our results. Regardless, our models continue to show a non-monotonic relationship between coupling and citations for patents that do not cite a *Scientific Index* journal, and a strictly positive relationship between coupling and future citations for those referencing science.

One potential alternate explanation for these results involves the diffusion of knowledge. References to non-patent sources might speed the dissemination of information, allowing communities of practice to build knowledge more rapidly.

This account implies that the mere act of publishing increases the flow of information, thereby increasing citation rates. Model 8 tests this possibility directly by interacting the effects of coupling with whether or not the patent references a *non*-scientific publication.[21] A reference to these publications does nothing to mitigate the effect of coupling, failing to support this account. The diffusion story further suggests that references to science should uniformly increase citations, rather than primarily increasing citation rates for inventions combining coupled technologies. Since the citation to science main effect has no impact on future citations, the models also fail to support this account on that basis. Variation in knowledge diffusion does not appear to account for our findings.

Another potential explanation for our findings concerns the effort invested in solving the inventive problem. Science may denote a greater level of effort that proves particularly useful in the face of difficult technological challenges. Although we cannot measure effort directly, the number of inventors may proxy for this factor.[22] Nevertheless, in both cases using and not using science, coupling correlates *negatively* and significantly with the number of inventors listed on the patent. The evidence, therefore, suggests that we can dismiss this alternative.

A more serious critique might contend that patents that cite science differ in some other systematic way from those that do not. To investigate this possibility, we first estimated a logistic regression of whether or not a patent cited a *Scientific Index* journal (see Table 7). Even after controlling for differences across technologies using class fixed effects (Model 11), these estimates reveal some differences between those patents that reference science and those that do not. First, patents that draw on science tend to come from highly active research domains, as the positive and significant coefficients for both the number of prior art citations and the number of trials reveal. This finding suggests that science does not differ from non-science by identifying fundamental new components. Second, patents drawing on science more likely come out of academia, though university patents account for only a small proportion (4.8%) of the total number of patents citing science in our sample. Notably, the models fail to show that science produces knowledge of any greater generality; neither the number of classes nor subclasses has a significant relationship with the use of science in the invention process. In Models 12 and 13, we compared the factors that predict a citation to science to those that correlate to university ownership (Model 12) and citations to non-scientific references (Model 13).

Though this analysis increases our confidence that non-patent references to the scientific literature capture a connection to science, it also reveals one difference that requires further investigation: patents that cite science exhibit lower coupling. Our theory actually accommodates this fact. Because we measure coupling as a function of the history of recombination, the prior application of scientific knowledge in a technological domain should facilitate prior recombination, thereby lowering the observed coupling score. One might worry, however, that this correlation represents some other type of selection rather than an unfortunate (but necessary) weakness in our measure of coupling. To test the possible influence of this selection on our models, we created a selection term from the logistic regression results (Model 10) representing the probability that each patent would cite science given its characteristics.[23] We then interacted this selection parameter with the coupling score and its quadratic term to determine whether selection effects might account for our previous findings. The results—see Model 9—demonstrate that the effects of science remain robust even after explicitly accounting for potential selection issues.

DISCUSSION

The results provide strong evidence that the returns to science depend on the difficulty of the inventive

[21] We coded any patent that referenced a journal not in the *Scientific Index* (e.g., *Time*) or a non-technical corporate publication—usually a product catalog or advertisement—as having a non-scientific reference because these publication types most clearly do *not* involve the application of scientific theory.

[22] Thanks go out to Riitta Katila and Steven Klepper for suggesting this test.

[23] The logic for this specification follows that of a Heckman selection model. The *application period*—the number of years between the application and granting of the patent—allows us to identify the procedure. Although this factor positively relates to both university assignees and the likelihood of citing science, we have no reason to expect that it would influence the number of future patents that would cite the focal patent following the grant date and the lack of correlation ($r = 0.02$) between application period and future cites supports this notion.

Table 7. Logit models of the correlates of citing scientific and non-scientific publications and of university ownership (standard errors in parentheses)[a]

	Model 10 cite science	Model 11 cite science (fixed effects)	Model 12 university	Model 13 cite non−science
Number of prior art cites	**0.035**	**0.041**	0.009	**0.057**
	(0.003)	**(0.003)**	(0.008)	**(0.003)**
Single subclass dummy	−0.041	−0.097	0.322	0.055
	(0.100)	(0.106)	(0.246)	(0.144)
Number of subclasses	0.013	0.010	−0.031	0.015
	(0.008)	(0.009)	(0.022)	(0.012)
Number of classes	−0.028	0.015	**0.175**	−0.020
	(0.029)	(0.031)	**(0.078)**	(0.042)
Number of trials	**0.004**	**0.004**	0.005	0.001
	(0.002)	**(0.002)**	(0.003)	(0.002)
Recent technology	**0.296**	0.028	0.164	**−0.345**
	(0.047)	(0.053)	(0.141)	**(0.052)**
Science stock in subclass	**5.594**	**4.806**	**3.010**	**1.830**
	(0.122)	**(0.159)**	**(0.300)**	**(0.176)**
Application period	**0.113**	**0.071**	**0.109**	0.008
	(0.022)	**(0.023)**	**(0.053)**	(0.032)
Coupling	**−0.613**	**−0.297**	−0.401	0.171
	(0.100)	**(0.127)**	(0.283)	(0.126)
University dummy	**0.921**	**0.900**		−0.324
	(0.137)	**(0.169)**		(0.287)
Cite to non-scientific publication	**0.274**	**0.538**	−0.115	
	(0.092)	**(0.097)**	(0.289)	
Cite to scientific publication			**1.004**	0.116
			(0.166)	(0.095)
Constant	**−4.253**		**−6.415**	**−2.518**
	(0.122)		**(0.632)**	**(0.233)**
Fixed effects	None	249 Classes	None	None
Log-likelihood	−5743.1	−4976.3	−1021.0	−3392.7
N	16,822	14,778[b]	16,822	16,822

[a] Bold type indicates that the coefficient estimate differs signification from zero with 95% confidence.
[b] 2,044 cases drop out of Model 11 because they belong to classes in which none of the patents cite science.

problem being addressed. When inventors work with relatively independent components, science offers little or no advantage. On the other hand, science can allow inventors recombining highly coupled components to avoid the difficulties inherent in local search on a rugged landscape. Though the size of this effect may appear modest, a 10 percent increase in citations at the mean level of coupling, it likely corresponds to a much larger economic benefit. If the exponential relationship between citations and economic value found by Harhoff *et al*. (1999) for German patents holds in the U.S. market, this 10 percent increase in citations would correspond to a 214 percent increase in economic value.[24] At one standard deviation above the mean of coupling, the expected value of applying science would increase the average value of a patent by 281 percent, at two standard deviations, 473 percent. Hence, the economic value of science may increase considerably with the difficulty of the inventive problem.

A more nuanced consideration of the results allows us to investigate which proposed mechanisms likely play the most important roles in generating this effect. The results do not, however, allow us to rule out any of the mechanisms through which science might influence the search process of invention. As one would expect if science either points inventors more directly to the most useful configurations of components or simply allows them to avoid swaths of less productive solution space, science increases the average usefulness of more tightly coupled inventions (see

[24] Harhoff *et al*. (1999) find the following relationship for value: Dollars $= 1\,000\,000\, e^{\left(\frac{\ln \text{cites} - 2.076}{0.119}\right)}$.

Tables 4 and 6). Moreover, the strong effects that science has on reducing the variance in inventive outcomes using coupled components suggests that science operates more by leading inventors directly to useful outcomes (see Table 5); reducing the solution space would have a relatively weak effect on the unpredictability in outcomes because even the 'best' regions of rugged landscapes still have highly varied terrain. Inconsistent with this direct link, however, inventors calling on science appear to work in regions with more search activity. Both Models 10 and 11 indicate that inventions that call on science work with components that have been recombined more frequently in the past, whereas a direct path to the best combinations should reduce the number of iterations. This finding suggests that science also aids invention by providing researchers with motivation to continue searching in a particular direction despite negative feedback.

These two mechanisms, one directing researchers to the most useful regions of space, the other encouraging them in the face of failure, likely complement each other in the process of inventing with coupled components. Even in areas with well-developed science, inventors often do not have sufficient knowledge to predict all of the interactions that might occur among a highly coupled set of components. Hence, even after embarking on the right path, they must typically engage in trial-and-error learning (probably through local search) to fill in these missing pieces. Many qualitative accounts, such as Fleming's (2002) history of ink-jet printing, point to just such an interaction.[25] Nucleation physics greatly enhanced the inventors' understanding of the fundamental processes underlying the formation of ink bubbles, but they still went through hundreds of iterations between science and empirical search to find the right combination of materials to produce consistently high-quality printing. As in this case, science likely operates through multiple mechanisms in improving inventive outcomes.

Science may even influence the shape of technological innovation along other dimensions beyond the process that inventors use in recombining components. For example, science might aid in the discovery of new components (Smith and Hounshell, 1985). Although the recombination of elements implies that the potential number of inventions expands very rapidly with the number of fundamental components (Weitzman, 1996), completely new components must obviously arise from somewhere. Some exogenous process might govern these arrivals, but science may also stimulate technological progress by identifying new fundamental components for recombination. Though outside the scope of our analysis, it seems interesting to note that of the 442 patents excluded from our analysis because they had no prior art, 46 percent reference a scientific journal article (vs. 17% in the sample as a whole).

Regardless of the particular mechanisms, the differential returns to the application of science likely influence when inventors call on science and when they do not. Although the analysis of when non-patent science references appear on a patent suggests that these patents involve the recombination of less coupled elements (see Models 10 and 11), this result likely arises as an artifact of the calculation of our coupling measure. We estimated coupling by observing the historical ease of recombining patent subclasses. If researchers have used science in the past to facilitate the use of this component, this practice would bias our coupling measure downward (indeed, the stock of science in a particular subclass correlates negatively, $r = -0.28$, with its observed difficulty of recombination). The effects of the selection terms in Model 9 corroborate this notion; the types of patents on which science references appear exhibit larger marginal effects to coupling. Hence, inventors appear to use science most frequently in those domains that, because they work with highly coupled components, offer the greatest returns to its application. To some degree, then, variation in the application of science across industries likely stems from the types of inventive challenges participants in these sectors face.

These results extend prior research on the link between science and innovation. Much of the existing work of the role of science in stimulating technological advance investigates the research and patenting activities of universities, identifying such features as changes in federal law,[26] the

[25] Rosenberg (1974) offers several additional examples.

[26] Prior to 1980, the property rights to the results of research funded by government grants belonged to the federal government. Thus, university researchers funded by federal grants had little incentive to apply for patents. The Bayh-Dole Act, passed in 1980, eliminated this provision and allowed researchers to retain the property rights to federally funded research.

establishment of university offices for technology transfer, and increased industry funding of university research as factors influencing the propensity of universities to patent (Henderson *et al.*, 1998; Mowery and Shane, 2002). Much of the application of science, however, occurs outside the university; in our sample, only 4.5 percent of the patents referencing an article in the *Scientific Index* had university assignees. Though private firms also face incentives to use science (Cockburn *et al.*, 2000), most prior research has failed to address the question of how these incentives might vary systematically according to the firm's technology. By focusing on precisely how science might aid inventors, this research identifies one technological factor—the difficulty of recombining coupled components—that appears to influence the potential returns to the application of science.

Our findings also inform research on complex systems—particularly research applying Kauffman's NK model to the social sciences. Kauffman's model predicts a non-monotonic relationship between coupling and the outcomes of local search. When actors do not use science—and presumably therefore engage in local search—the results conform closely to Kauffman's expectations. Nevertheless, our results clearly show the *inapplicability* of this model to situations in which the searchers have some understanding or cognitive model of the landscape they search; inventors following science do not behave like local searchers. Thus, researchers wishing to apply the NK model to cognitively aware agents should revisit the model and modify it to account for the search heuristics that people actually use (cf. Gavetti and Levinthal, 2000; Rivkin, 2000).

For the individual inventor, the policy implications seem straightforward. Science offers limited benefits to inventors working with modular components—given the cost of searching and digesting the scientific literature, the price tag for using science likely exceeds its benefits for those working with uncoupled components. In contrast, science offers large potential rewards to inventors operating with highly coupled components. Without science, inventors that search these landscapes must rely on exhaustive search techniques. Although inventors could focus resources on developing methods for testing large numbers of combinations cheaply, efforts along these lines have met with limited success; for example, combinatorial synthesis strategies remain problematic, as demonstrated by the recent controversy in drug discovery over the effectiveness of combinatorial chemistry and high-throughput screening (Drews, 2000). An understanding of the forces creating these interactions through the development of basic science offers an attractive means for accelerating invention in these highly coupled technologies.

In conclusion, we wish to reiterate that this paper proposed and tested one explanation for why the returns to the application of scientific knowledge to invention might vary across technological fields. In particular, by envisioning invention as a process of searching through combinations of technological components for new and useful configurations, we argue that science acts like a map—providing inventors with a sense of the underlying technological landscape they search—thereby allowing them to avoid the difficulties inherent in trying to combine highly coupled components. Our results demonstrate that scientific knowledge mitigates the negative effect that coupling typically has on the outcomes of invention. In doing so, this research elaborates the more proximate role that science plays in technological advance and how it influences the rate and quality of invention. Thus, it brings a somewhat more nuanced view of the benefits of science in stimulating technological advance, and with it economic growth, to the already extensive literature on the subject.

ACKNOWLEDGEMENTS

We acknowledge the Harvard Business School Department of Research for their financial support, Corey Billington and Ellen King of Hewlett Packard for donation of their patent database, and Evelina Fedorenko for her historical research. We especially wish to thank Sue Borges for her painstaking analysis of over 16,000 non-patent citations. Comments from Pierre Azoulay, Michael Darby, Riitta Katila, Steven Klepper, Dan Levinthal, Sue McEvily, Jan Rivkin, and the participants of seminars at UCLA and NBER improved this paper substantially. The usual disclaimer applies.

REFERENCES

Adams J. 1990. Fundamental stocks of knowledge and productivity growth. *Journal of Political Economy* **98**: 673–702.

Ahuja G, Lampert J. 2001. Entrepreneurship in the large corporation: a longitudinal study of how established firms create breakthrough inventions. *Strategic Management Journal*, Special Issue **22**(6–7): 521–543.
Albert M, Avery D, Narin F, McAllister P. 1991. Direct validation of citation counts as indicators of industrially important patents. *Research Policy* **20**: 251–259.
Allen T. 1977. *Managing the Flow of Technology*. MIT Press: Cambridge, MA.
Athey S, Stern S. 1998. An empirical framework for testing theories about complementarity in organizational design. Working paper #6600, National Bureau of Economic Research.
Audia P, Sorenson O. 2001. Aspirations, performance and organizational change: a multi-level analysis. Working paper, London Business School.
Baldwin C, Clark K. 2000. *Design Rules: The Power of Modularity*. MIT Press: Cambridge, MA.
Basalla G. 1988. *The Evolution of Technology*. Cambridge University Press: Cambridge, U.K.
Cameron A, Trivedi P. 1998. *Regression Analysis of Count Data*. Cambridge University Press: Cambridge, U.K.
Campbell A. 2001. Superconductivity: how could we miss it? *Science* **292**: 65–66.
Carlson D. 2001. Personal interview, June 5: Cary, NC.
Cockburn I, Henderson R, Stern S. 2000. Untangling the origins of competitive advantage. *Strategic Management Journal*, Special Issue **21**(10–11): 1123–1145.
Cyert R, March J. 1963. *A Behavioral Theory of the Firm*. Prentice-Hall: Englewood Cliffs, NJ.
Dresselhaus M. 2001. Burn and interrogate. *Science* **292**: 650–651.
Drews J. 2000. Drug discovery: a historical perspective. *Science* **287**: 1960–1964.
Fleming L. 2001. Recombinant uncertainty in technological search. *Management Science* **47**: 117–132.
Fleming L. 2002. Finding the organizational sources of technological breakthroughs: the story of Hewlett Packard's thermal inkjet. *Industrial and Corporate Change* **11**: 1059–1084.
Fleming L, Sorenson O. 2001. Technology as a complex adaptive system: evidence from patent data. *Research Policy* **30**: 1019–1039.
Gambardella A. 1995. *Science and Innovation: The U.S. Pharmaceutical Industry During the 1980s*. Cambridge University Press: Cambridge, U.K.
Gavetti G, Levinthal D. 2000. Looking forward and looking backward: cognitive and experiential search. *Administrative Science Quarterly* **45**: 113–137.
Gibbons M, Johnston R. 1974. The roles of science in technological innovation. *Research Policy* **3**: 220–242.
Gilfillan S. 1935. *Inventing the Ship*. Follett: Chicago, IL.
Haberman S. 1977. Log-linear model and frequency tables with small expected cell counts. *Annals of Statistics* **5**: 1148–1165.
Hall B, Jaffe A, Trajtenberg M. 2000. Market value and patent citations: a first look. NBER working paper 7741.
Hansen MT, Lovas B. 2004. How do multinational companies leverage technological competancies? Moving from single to interdependent explanations. *Strategic Management Journal*, Special Issue **25**(8–9): 801–822.
Harhoff D, Narin F, Scherer FM, Vopel K. 1999. Citation frequency and the value of patented inventions. *Review of Economics and Statistics* **81**: 511–515.
Hausman J, Hall B, Griliches Z. 1984. Econometric models for count data with an application to the patents–R&D relationship. *Econometrica* **52**: 909–938.
Henderson R, Clark K. 1990. Architectural innovations: the reconfiguration of existing product technologies and failure of established firms. *Administrative Science Quarterly* **35**: 9–30.
Henderson R, Cockburn I. 1994. Measuring competence? Exploring firm effects in drug discovery. *Strategic Management Journal*, Winter Special Issue **15**: 63–84.
Henderson R, Jaffe A, Trajtenberg A. 1998. Universities as a source of commercial technology: a detailed analysis of university patenting, 1965–1988. *Review of Economics and Statistics* **80**: 119–127.
Jaffe A. 1989. Real effects of academic research. *American Economic Review* **79**: 957–970.
Jaffe A, Trajtenberg M. 1996. Flows of knowledge from universities and federal labs: modeling the flow of patent citations over time and across institutional and geographic boundaries. *Proceedings of the National Academy of Sciences* **93**: 12 671–12 677.
Jaffe A, Trajtenberg M, Henderson R. 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics* **108**: 577–598.
Katila R. 2002. New product search over time: past ideas in their prime. *Academy of Management Journal* **45**: 995–1010.
Kauffman S. 1993. *The Origins of Order*. Oxford University Press: New York.
King G. 1989. Event count models for international relations: generalizations and applications. *International Studies Quarterly* **33**: 123–147.
Klepper S. 1996. Entry, exit, and innovation over the product life-cycle. *American Economic Review* **86**: 562–583.
Kramer P. 1997. *Listening to Prozac*. Penguin: New York.
Levin R, Klevorick A, Nelson R, Winter S. 1987. Appropriating the returns from industrial research and development: comments and discussion. *Brookings Papers on Economic Activity* **3**: 783–831.
Levinthal D. 1997. Adaptation on rugged landscapes. *Management Science* **43**: 934–950.
Lippman S, McCall J. 1976. The economics of job search: a survey. *Economic Inquiry* **14**: 155–189.
Mansfield E. 1972. Contribution of R&D to economic growth in the United States. *Science* **175**: 477–486.
Mansfield E. 1995. Academic research underlying industrial innovations: sources, characteristics and financing. *Review of Economics and Statistics* **77**: 55–65.

March J. 1991. Exploration and exploitation in organizational learning. *Organization Science* **2**: 71–87.
March J, Simon H. 1958. *Organizations*. Blackwell: Cambridge, MA.
Martin X, Mitchell W. 1998. The influence of local search and performance heuristics on new design introduction in a new product market. *Research Policy* **26**: 753–771.
McCluskey E. 1986. *Logic Design Principles with Emphasis on Testable Semi-custom Circuits*. Prentice-Hall: Englewood Cliffs, NJ.
Meyer M. 2000. Does science push technology? Patents citing scientific literature. *Research Policy* **29**: 409–434.
Milgrom P, Roberts J. 1990. The economics of modern manufacturing: technology, strategy and organization. *American Economic Review* **80**: 511–528.
Millman J. 1979. *Microelectronics*. McGraw-Hill: New York.
Mowery D, Shane S. 2002. Introduction to the special issue on university entrepreneurship and technology transfer. *Management Science* **48**: v–ix.
Narin F, Hamilton K, Olivastro D. 1997. The increasing linkage between U.S. technology and public science. *Research Policy* **26**: 317–330.
Nelson R. 1982. The role of knowledge in R&D efficiency. *Quarterly Journal of Economics* **97**: 453–470.
Nelson R, Winter S. 1982. *An Evolutionary Theory of Economic Change*. Belknap: Cambridge, MA.
Ouyang M, Huang J, Cheung C, Lieber C. 2001. Energy gaps in 'metallic' single-walled carbon nanotubes. *Science* **292**: 702–705.
Perko J, Narin F. 1997. The transfer of public science to patented technology: a case study in agricultural science. *Journal of Technology Transfer* **22**: 65–72.
Podolny JM, Stuart TE. 1995. A role-based ecology of technological change. *American Journal of Sociology* **100**: 1224–1260.
Rivkin J. 2000. Imitation of complex strategies. *Management Science* **46**: 824–844.
Rosenberg N. 1974. Science, invention, and economic growth. *Economic Journal* **84**: 90–108.
Schmoch U. 1993. Tracing the knowledge transfer from science to technology as reflected in patent indicators. *Scientometrics* **26**: 193–211.
Schumpeter J. 1939. *Business Cycles*. McGraw-Hill: New York.
Smith J, Hounshell D. 1985. Wallace H. Carothers and fundamental research at DuPont. *Science* **229**: 436–442.
Sørensen J, Stuart T. 2000. Aging, obsolescence and organizational innovation. *Administrative Science Quarterly* **45**: 81–112.
Sorenson O. 2002. Interorganizational complexity and computation. In *Companion to Organizations*, Baum JAC (ed). Blackwell: Oxford; 664–685.
Stephan P. 1996. The economics of science. *Journal of Economic Literature* **34**: 1199–1235.
Stix G. 2001. Project Skyhook: a 'smart' material that transforms from a liquid to a solid state on cue is beginning to show up in prosethetics, automobiles, and other applications. *Scientific American* May: 28–29.
Stuart T, Podolny J. 1996. Local search and the evolution of technological capabilities. *Strategic Management Journal*, Summer Special Issue **17**: 21–38.
Sveikauskas L. 1981. Technological inputs and multifactor productivity growth. *Review of Economics and Statistics* **63**: 275–282.
Tijssen R. 2001. Global and domestic utilization of industrial relevant science: patent citation analysis of science–technology interactions and knowledge flows. *Research Policy* **30**: 35–54.
Trajtenberg M. 1990. A penny for your quotes: patent citations and the value of innovations. *Rand Journal of Economics* **21**: 172–187.
Tratjenberg M, Henderson R, Jaffe A. 1997. University versus corporate patents: a window on the basicness of invention. *Economic Innovation and New Technology* **5**: 19–50.
Ulrich K. 1995. The role of product architecture in the manufacturing firm. *Research Policy* **24**: 419–440.
Usher A. 1954. *A History of Mechanical Invention*. Dover: Cambridge, MA.
Vincenti W. 1990. *What Engineers Know, and How They Know It*. Johns Hopkins University Press: Baltimore, MD.
Weinberger E. 1991. Local properties of Kauffman's N-K model: a tunably rugged energy landscape. *Physical Review A* **44**: 6399–6413.
Weitzman M. 1996. Hybridizing growth theory. *American Economics Review* **86**: 207–212.
Wright S. 1932. The roles of mutations, inbreeding, crossbreeding and selection in evolution. *Proceedings of the 11th International Congress of Genetics* **1**: 356–366.
Zucker L, Darby M. 1996. Star scientists and institutional transformation: patterns of invention and innovation in the formation of the biotechnology industry. *Proceedings of the National Academy of Sciences* **93**: 709–716.
Zucker L, Darby M, Brewer M. 1998. Intellectual human capital and the birth of U.S. biotechnology enterprises. *American Economic Review* **88**: 290–306.

