Made in Academia:
The Effect of Institutional Origin on Inventors' Attention to Science

Michaël Bikard
London Business School
mbikard@london.edu

January 2018

ABSTRACT

Inventors cannot exploit new scientific discoveries if they do not pay attention to them. However, allocating attention to science is difficult because the scientific literature is vast, fast-changing, and often unreliable. Inventors are therefore likely to rely on informational cues when screening new publications. I posit that inventors pay significantly less attention to discoveries "made in academia" than to those "made in industry" because they believe that the work of academic scientists will be less useful to them. I test this proposition by examining inventors' patent references to the scientific literature in the case of simultaneous discoveries made by at least one team based in academia and another based in industry. I find that inventors are 23% less likely to cite the academic paper than its twin from industry. My results highlight the importance of inventors' attention as a hitherto underexplored bottleneck shaping the translation of science into new technologies.

Keywords: innovation; attention; academic science; simultaneous discoveries; patents; use of science in inventions

Acknowledgments: This paper was originally distributed with the title, "Is Knowledge Trapped inside the Ivory Tower?" I would like to thank Scott Stern, Keyvan Vakili, and Isabel Fernandez-Mateo for extremely insightful comments and suggestions. I also thank Pierre Azoulay, Julian Birkinshaw, Kevin Boudreau, Sen Chai, Gary Dushnitsky, Sendil Ethiraj, Roman Galperin, Patrick Gaulé, Carolin Häussler, Benjamin Mako Hill, Matt Marx, Fiona Murray, Abhishek Nagaraj, Michael Roach, Henry Sauermann, Yanbo Wang, Ezra Zuckerman, and seminar participants at the NBER Productivity Lunch and at numerous conferences and seminars. Financial support from the Ewing Marion Kauffman Foundation, an NSF SBE Doctoral Dissertation Research Improvement Grant, the Center for Biomedical Innovation at MIT, and the Deloitte Institute of Innovation and Entrepreneurship at LBS is gratefully acknowledged. All errors are my own.

1. INTRODUCTION

Science can accelerate technological innovation (Cohen, Nelson, and Walsh 2002; Mokyr 2002; Fleming and Sorenson 2004), but only if inventors[1] pay attention to it. Yet, identifying useful scientific discoveries is a difficult task. Not only is scientific knowledge generally complex and tacit (Polanyi 1958; Zucker and Darby 1996), but the literature is vast and fast-changing (Jones 2009; Alberts 2010) and the reliability of published findings is uncertain (Freedman, Cockburn, and Simcoe 2015). When information technologies allow easy access to millions of scientific publications, how can boundedly rational inventors possibly keep track of all this knowledge?

I argue that inventors, in order to avoid drowning in the scientific literature, select—whether deliberately or not—how much attention each scientific publication will receive. Selective attention plays a crucial role in both individual and organizational behavior because it bounds individual rationality and determines the menu of available actions (Simon 1947). Clearly, scientific publications that receive little or no attention from inventors are unlikely to be translated into new technologies. Inventors' attention is therefore likely to constitute a critical bottleneck at the interface between science and technology. To navigate the scientific literature, inventors are likely to rely on informational cues drawn from those attributes of a publication that are easily observable (Merton 1968; Simcoe and Waguespack 2011; Azoulay, Stuart, and Wang 2013; Polidoro 2013). Inventors might then pay particular attention to scientific publications that carry specific cues while ignoring others that carry different cues, even if their contents are very similar.

I posit that inventors use the publication's "institution of origin"—that is, whether it stems from an academic or an industrial organization—as such a cue. They might pay less attention to discoveries made in academia than to those made in industry because they know that academic scientists tend to be shielded from commercial pressures. They also know that scientists working in industry have a strong

[1] For simplicity, throughout the paper I call individuals involved in scientific publications "scientists" and those involved in patenting "inventors." In studying the interface between science and technology, I specifically focus on inventors' attention to science produced by others.

incentive to produce knowledge that will be technologically valuable. Thus, the well-documented tension between science as an output evaluated by other scientists and science as an input to technology development (Gittelman and Kogut 2003; Stern 2004; Evans 2010; Murray 2010) is likely to affect inventors' attention to the scientific literature.

The challenge in investigating this proposition is considerable. Because scientists working in industry and in academia typically focus on different projects (Aghion, Dewatripont, and Stein 2008), a simple comparison of inventors' propensity to pay attention to papers from academia and from industry would be uninformative. Inventors might simply pay less attention to publications from academia because of their fundamental nature and not because of their institutional origin. The empirical challenge is therefore an identification problem. Researchers can often observe a treated group without being able to observe a suitable control group. To address this challenge, I use a novel empirical strategy based on the common occurrence of simultaneous discoveries in science (Merton 1957). When two or more teams make the same or very similar discoveries at around the same time and send their manuscripts for publication, both papers may be published, forming what I refer to as "paper twins." (I use the term "twin" even when there are more than two such simultaneous publications; sets of two are much more common than sets of three or more.) By focusing on those twins, it is possible to keep the discovery (almost) constant across different institutional origins. Thus, the relationship between the institutional environment of origin and inventors' attention becomes more salient. In practice, I consider 479 patents and their references (or lack thereof) to 90 scientific publications representing 39 sets of paper twins—28 pairs, 10 sets triplets, and 1 set of quadruplets—in which at least one paper came from industry and one from academia.

This empirical strategy has two key advantages. First, because the paper twins reveal instances in which very similar knowledge emerges at around the same time in different environments, they offer a unique opportunity to isolate the impact of the discovery's content from the cues embedded in its context of emergence. Although there can be subtle differences between the manuscripts of teams that make the same discovery, it is possible to measure those differences—even if imperfectly—and to account for them empirically. Second, patent references to the paper twins constitute a useful measure of inventors' attention to the scientific literature. Roach and Cohen (2013) find that patent citations to papers are a better measure of knowledge flow between academia and industry than patent citations to other patents. In the case of paper twins, inventors have in principle a choice of citing one publication, its twin, or both. This study's complementary qualitative work, involving 50 interviews with scientists and inventors, indicates that inventors' citation (or noncitation) of each paper in a set of twins reflects the differential attention that those papers received.

My results confirm that an institution-of-origin effect shapes inventors' attention to science. In my data, an inventor's patent is 23% less likely to cite a paper from academia than its twin from industry. This large effect is robust to a number of specifications and does not appear to be driven by status, social network, or subtle differences between twins. These findings point to the strategic role that attention plays in the process of science-based invention. Failure to pay attention to promising scientific findings might lead to missed opportunities, whereas productivity might suffer if inventors pay too much attention to unpromising or unreliable research results. Absent an understanding of inventors' attention, apparent low rates of knowledge transfer might be attributed to the often-studied tacit nature of knowledge when they in fact stem from a lack of attention from inventors. The challenges raised by tacitness and attention are very different. While issues of tacitness call for investments in codification, issues of inventors' attention call for investments in visibility- and credibility-enhancing measures, such as replication by third parties and broader communication.

In investigating how informational cues shape inventors' attention to science, this study attempts to make three contributions. First, it highlights that science-based invention is complicated not only by the tacit and complex nature of scientific knowledge, but also by the fact that the scientific commons is vast, fast-changing, and often unreliable. While the benefits of using science as a map have been well-documented (Cohen, Nelson, and Walsh 2002; Fleming and Sorenson 2004; Arora, Belenzon, and Patacconi 2017), this study points to the challenges and cost that this map entails. Inventors' attention is scarce (Simon 1947), which constitutes a hitherto understudied bottleneck at the interface between science and technology. Second, while prior studies of the informational cues used in science and innovation highlight the impact of status and certification (Merton 1968; Simcoe and Waguespack 2011; Azoulay, Stuart, and Wang 2013; Polidoro 2013), this paper highlights another cue—the academic or industrial origin of a scientific publication. Finally, this paper is the first to use simultaneous discoveries to conduct a "twin study of scientific knowledge," introducing a new method for social scientists interested in science and its dissemination.

2. INVENTORS' ATTENTION TO SCIENCE

2.1. Challenges in navigating the scientific literature

Inventors can use science, which describes natural phenomena and regularities, as a "map" suggesting the way to new ideas, to solutions for complex problems, and generally to more efficient work (Mokyr 2002; Cohen, Nelson, and Walsh 2002; Fleming and Sorenson 2004). The exploitation of scientific knowledge can therefore drive the success of inventors, firms, and entire industries (Zucker, Darby, and Brewer 1998; Cockburn and Henderson 1998; Murmann 2003). However, the fact that many economists have traditionally described science as a "public good" (e.g., Arrow 1962) does not mean that its exploitation is straightforward or costless (Rosenberg 1990). In fact, inventors face fundamental challenges in identifying which scientific publications might help them develop new technologies.

First, scientific knowledge tends to be complex and tacit. Polanyi (1958: 57) noted that "the large amount of time spent by students of chemistry, biology and medicine in their practical courses shows how greatly these sciences rely on the transmission of skills and connoisseurship from master to apprentice. It offers an impressive demonstration of the extent to which the art of knowing has remained unspecifiable at the very heart of science." This insight has considerable implications. Knowledge developed by scientists can remain tacit many years after its discovery. Transferring it might be costly, rendering the knowledge sticky (von Hippel 1994) or "naturally excludable" (Zucker and Darby 1996). In principle, academic scientists codify new knowledge to facilitate its dissemination. However, this is not always in their interest because codification can be costly (Nelson and Winter 1982; Zucker, Darby, and Armstrong 2002).

Second, the scientific literature is vast and fast-changing. Large amounts of new knowledge are published every day and staying up to date is costly. Over 15 million scientists have authored at least one article indexed by Scopus between 1996 and 2011 (Boyack et al. 2013), and this sheer volume makes it practically impossible to keep track of all the work published in even one area of interest. While the "Renaissance Man" ideal of a scientist well-versed in disparate disciplines once seemed achievable, the growing burden of knowledge constrains today's scientists to lengthen their doctoral programs, specialize, and collaborate (Jones 2009; Agrawal, Goldfarb, and Teodoridis 2016; Teodoridis 2017). The challenge in keeping track of the scientific literature is likely to be even more salient for those involved in technology development. Technologies tend to draw on multiple scientific specialties and inventors are generally—though not always—less embedded in the scientific literature than scientists themselves (Rosenberg 1994).

Third, evaluating scientific manuscripts is not a straightforward process, as their reliability tends to be uncertain. Differing evaluation criteria are possible, which can lead to controversies (Engelhardt and Caplan 1987; Boudreau et al. 2016). In the past few years, practitioners, the media, and many scientists have loudly complained that there is no mechanism by which inventors can guarantee the usability of academic findings (*The Economist* 2013; Freedman, Cockburn, and Simcoe 2015). The integrity of research tools is sometimes questionable (Furman and Stern 2011), but the key issue is reproducibility, with estimated rates of reproducibility of academic studies ranging from a worrisome 49% (Hartshorne and Schachner 2012) to an alarming 11% (Begley and Ellis 2012). This is a crucial area of concern for those involved in science-based invention. In a recent editorial piece, Merck's Chief Medical Officer noted that "the consequences of building on a shaky platform are so great that most biopharmaceutical companies and venture capitalists expend considerable resources in attempts to replicate initial data before committing to further research" (Rosenblatt 2016: 1).

Overall, large quantities of highly technical scientific publications constantly compete for inventors' attention and their value is often uncertain. Inventors' understanding of the scientific literature will therefore always be incomplete. This, in turn, raises the question of how they allocate their scarce attention. If they could observe the value of scientific knowledge directly and costlessly, they would be able to allocate their attention efficiently to the most valuable publications. However, the fact that the scientific literature is so vast, fast-changing, complex, and unreliable means that the allocation of attention is likely to have significant implications for inventors' performance and for technology development in general.

2.2. Allocating attention to some scientific publications and not to others

Much research has investigated the challenges in translating tacit scientific knowledge into new technologies. To do so, innovators must cultivate the appropriate skills (for individuals) or capabilities (for organizations). They can increase their absorptive capacity by conducting research themselves to develop the relevant cognitive structures to acquire, recall, and use knowledge produced elsewhere (Cohen and Levinthal 1990; Gambardella 1992). Firms might also reduce the cost of knowledge acquisition by creating tight relationships with academic scientists (Cockburn and Henderson 1998; Zucker, Darby, and Armstrong 2002) or by fostering individual mobility across the academia-industry divide (Murmann 2003; Furman and MacGarvie 2007). Collocation can facilitate knowledge flow by allowing frequent interactions (Zucker, Darby, and Brewer 1998; Belenzon and Schankerman 2013). Consider the case of recombinant DNA. Early on, knowledge about the new technique diffused slowly because it was largely tacit. To learn it, scientists had to work side by side with more experienced colleagues in laboratories already using it (Zucker and Darby 1996).

Although research has examined how the tacit nature of science complicates science-based invention, it has not examined the role of inventors' attention. This omission is significant. Attention is a necessary—though not sufficient—driver of knowledge transfer. Empiricists who overlook the issue of attention might overestimate the challenge of the tacitness of knowledge. The issue is not only an empirical one, however. Inventors' scarce attention constitutes a hitherto understudied bottleneck in the translation of science into new technology. This bottleneck is different from that imposed by the complex and tacit nature of knowledge. It raises different theoretical questions about determinants of attention and calls for different solutions for practitioners interested in boosting science-based invention.

The way inventors manage their attention to science is likely to have important strategic implications. Attention is known to affect firm performance (Ocasio 1997) and the allocation of attention amongst scientists explains in part why some make Nobel Prize-winning breakthroughs and others do not (Chai 2017). Over 70 years ago, Herbert Simon noted that, considering the sheer volume of information available at any given time, the most important challenge in decision-making is not the scarcity of information but rather the scarcity of attention. "Until they are noticed, opportunities are not opportunities. In the world in which we actually live, at any given time we notice only a tiny fraction of the opportunities that are objectively present (…). A major initial step— and by no means an assured one— in technological or social invention is (…) to attend to the right cues" (Simon 1947: 123).

Indeed, cues play an important role in science and in technology. The imperfection of the system of communication in science has been known for decades. Most studies have focused on the cue conferred by an author's social status (Merton 1968; Simcoe and Waguespack 2011; Azoulay, Stuart, and Wang 2013). The underlying idea is straightforward: since scientists cannot possibly keep track of the entire scientific literature, they use the status of the authors to assess how much attention a publication deserves. Innovators also use cues to assess the potential of various technological niches. Several such cues have been examined, including the social status of the innovators (Podolny and Stuart 1995) and the recognition that they received (Rao 2004). Interestingly, science itself can bolster the prominence of new technologies because the review process acts as a certification mechanism that mitigates some of the uncertainty surrounding new technologies (Azoulay 2002; Polidoro and Theeke 2011; Polidoro 2013). Thus, informational cues shape the evolution of science and of technology, but their impact on the interface between the two remains unknown. This paper constitutes a first step in exploring that question.

2.3. The impact of institutional origin on inventors' attention

Inventors are likely to pay less attention to a discovery "made in academia" than to one "made in industry" because they know that academics are broadly shielded from commercial pressures. At the end of World War II, Vannevar Bush (1945: 241) argued that academic scientists should be protected "from the adverse pressure of convention, prejudice, or commercial necessity." Kuhn (1970: 163) noted that the insulation of scientific communities is unparalleled: "There are no other professional communities in which individual creative work is so exclusively addressed to and evaluated by other members of the profession." This feature is consequential: "Just because he is working only for an audience of colleagues, an audience that shares his own values and beliefs, the scientist can take a single set of standards for granted. He need not worry about what some other group or school will think and can therefore dispose of one problem and get on to the next more quickly" (163). Taking into account that the "other group" might be inventors sheds some light on the challenge that academia as an institutional environment raises for science-based invention. Scientists and inventors do not approach scientific knowledge in the same way (Stokes 1997). The findings of academic scientists are not primarily evaluated for their practical usefulness. Conversely, inventors do not assess scientific discoveries chiefly for their contribution to fundamental understanding (Fini, Jourdan, and Perkmann 2017).

In particular, academic origin might signal poor replicability. Pressures to publish or perish in academia are well known throughout the preclinical literature (Mobley et al. 2013; Begley and Ioannidis 2015).[2] A recent survey of faculty and trainees at MD Anderson Cancer Center illustrates those pressures: "When trainees at the institution were asked if they had ever felt pressure to prove the mentors' hypothesis, even when the data that the trainee generated did not support it, 31.4% reported that they had felt pressure. Furthermore, 18.6% of trainees said that they had been pressured to publish findings about which they had doubts. Additionally, when asked if they were aware of mentors who required a high impact journal publication before a trainee can complete his or her training in the laboratory of their mentor, 48.9% (68/139) reported that they were aware of this requirement" (Mobley et al. 2013).

[2] I thank an anonymous reviewer for these references.

Industry origin, on the other hand, is likely to signal commercial value. From the perspective of inventors, the fact that scientific publications were "made in industry" indicates that a firm believes enough in the economic value of that area of science to invest in its exploration. Polidoro (2013) studies instances in which firms decide to enter new technological niches by developing drugs using new mechanisms of action in 14 therapeutic classes. He finds that rivals are more likely to enter when pioneering companies have published about their technology, presumably in part because these publications draw attention to the practical benefits of the new mechanism of action for the development of new drugs. Industry publications in top scientific journals might also be more replicable than academic papers. Publication is not crucial from a firm's standpoint (Stern 2004). Since scientists in firms have much less pressure to publish than those who work in academia, one might expect that they are less likely to publish shaky results (Casadevall and Fang 2012).

In sum, an institution-of-origin effect might shape inventors' attention to science. Inventors might not pay close attention to discoveries stemming from academia, at least as compared with discoveries published by firms. Since inventors' relative attention to the scientific literature is visible in their references to it in their patents, I posit that:

HYPOTHESIS. *Inventors are less likely to refer in their patents to a publication from academia than to a publication originating from industry that discloses essentially the same discovery.*

3. METHOD AND DATA

3.1. Structure and composition of the dataset

My empirical approach is probably best explained by considering an example. In 2001, a San Diego-based company called X-Ceptor invented a new method for treating a number of diseases, including obesity, using liver X receptor (LXR) compounds. This invention was based on recent scientific developments. Patent US6924311, which describes the invention, includes 28 references to the scientific literature. Those references provide a window onto the attention that the inventors paid to the scientific literature that was relevant to this invention. However, among all the publications that were not referenced, how can one gauge which were relevant but were omitted? One way to address this challenge would be to observe essentially the same discovery “made in academia” and “made in industry” and then to examine whether inventors are more likely to cite one or the other in their patents. I take this approach here by focusing on paper twins. Table 1 presents five such sets of paper twins drawn from my dataset.

[Insert Table 1 and Figure 1 about here]

In particular, consider the example presented in Figure 1. Both papers report the discovery that a point mutation in the OB-R gene (a receptor for leptin) explains the obese phenotype of diabetic mice. X-Ceptor’s inventors could in principle have referenced both papers in patent US6924311. In fact, they cited the former paper, which was published by Millennium Pharmaceuticals in Cambridge, MA, but not the latter one, which was published by Rockefeller University in New York. My empirical strategy consists of focusing on inventors such as those from X-Ceptor and on the citations they make in their patents when they have a choice of “paper twins”— one from academia and one from industry. The level of observation is therefore the patent–paper dyad. Here, patent US6924311 forms one dyad with the Millennium paper and another dyad with the Rockefeller paper. In structuring the dataset that way, I can observe not only realized citations but also citations that were possible but not realized. This, in turn, lets me estimate for each dyad whether the paper’s academic or industrial origin predicts citation.

[Insert Figure 2 about here]

For clarity, the dataset and its structure are presented in Figure 2. It consists of 479 patents (including US6924311) and their citations or noncitations of 90 scientific papers (including the Millennium and Rockefeller papers). Those 90 papers constitute 39 sets of paper twins: 28 pairs, 10 sets of triplets and 1 set of quadruplets. Paper twins tend to be published around the same time in very similar outlets. The combinations between the 477 patents and those 90 papers amount to 924 patent–paper dyads. The data are drawn from several sources. Data for each publication come from ISI Web of Science and Scopus. Details about the corresponding author as well as each author’s address come from an analysis of the text of the publications. Patent data and patent-level variables were collected using

Harvard's IQSS patent database (Li et al. 2014) and impact-factor data come from the ISI website. Table 2 lists the main variables and their definitions.

[Insert Table 2 about here]

I complement my quantitative analysis with 48 interviews with scientists and inventors, one interview with a patent lawyer, and one interview with an editor at a prestigious scientific journal. I do not use these qualitative data to test my hypothesis. Rather, those interviews provide a deeper understanding of how inventors navigate the scientific literature and how this is reflected in their patent citations to scientific papers—especially in the case of paper twins. Those interviews were important for two reasons. First, patent citations to the academic literature remain poorly understood and the process of patent citation in the case of paper twins has not been studied before. My interviews therefore provided key insights about how I should interpret the results from my quantitative analysis. Second, since I study the allocation of attention, my interviews helped me understand how inventors think about the institutional origin of scientific discoveries. This provided more fine-grained insights about my results, which complement my quantitative analysis. The interviews lasted from 30 minutes to two hours and were all recorded and transcribed—except for three interviewees who refused to be recorded.

3.2. Paper twins

The algorithm used to produce the sample of paper twins is based on findings from the sociology of science about the allocation of credit in the case of simultaneous discoveries. Of particular interest is the insight that scientists use citations as "votes" highlighting which team they think deserves the credit for the discovery. When they believe that credit ought to be split, they typically cite two or more papers adjacently (Cozzens 1989). Based on this insight, simultaneous discoveries can be identified automatically by observing instances in which two or more teams share the credit for a specific discovery; that is, two or more papers are consistently cited within the same parentheses or adjacently. In practice, I did not restrict my search to specific disciplines. Rather, I sought to identify all the pairs of papers that were consistently cited adjacently in every research paper published by the top 15 non-review scientific journals between 2000 and 2010. The general method used to identify paper twins, its theoretical foundations, the algorithm, and various robustness analyses are detailed in a separate paper (Bikard 2012). To identify the sample of 39 sets of paper twins that I use here, I used a six-step process.

In step 1, I used ISI Web of Knowledge to collect information about 42,106 research articles that appeared in the top 15 non-review scientific journals (2009 journal impact factor ranking) between 2000 and 2010. Those journals are *Nature*, *Science*, *Cell*, *New England Journal of Medicine*, *JAMA*, *Lancet*, *CA: A Cancer Journal for Clinicians*, *Nature Genetics*, *Nature Materials*, *Nature Medicine*, *Nature Immunology*, *Nature Nanotechnology*, *Nature Biotechnology*, *Cancer Cell*, and *Cell Stem Cell*. In step 2, I gave each reference listed in those papers a unique identifier using PubMed and CrossRef. I identified 744,583 unique references made by those 42,106 scientific articles. In step 3, I generated a database of pairs of all references that (a) were co-cited in at least one paper, (b) were published no more than a calendar year apart, (c) had no overlapping author, and (d) were cited at least five times in the 42,106 papers that I began with. In all, 17,050,914 pairs were considered and 449,417 met those criteria. In step 4, I computed the Jaccard co-citation index for each of those pairs. I selected 2,320 pairs that had a co-citation index greater than 50%. In step 5, I used a parsing algorithm to identify those pairs for which co-citation was always within the same parentheses or adjacent. I excluded 495 pairs for which I was not able to parse at least three co-citing articles. I found that 720 of the remaining 1,825 pairs had been cited adjacently in 100% of the co-citing articles. Finally, in step 6, I collected the addresses of corresponding authors for each of the 1,246 papers that constitute these 720 pairs and coded them as academic or industrial. I found 49 pairs that involved one paper from academia and one paper from industry. In several cases, the same paper was paired with more than one other paper. Those instances formed 10 triplets and 1 quadruplet—which I also refer to as “paper twins” for simplicity. In total, these efforts yielded a sample of 90 papers constituting 39 sets of paper twins. To my knowledge, this is the first sample of discoveries made simultaneously in academia and in industry ever built.

Considering my research question, this sampling strategy has three important advantages. First, to identify paper twins, I looked for frequent co-citations in the most prestigious and highly visible journals. This means that I am excluding poorly cited discoveries. This is a study of inventor’s selective attention to science and it is important to note that I am not measuring their attention to some obscure scientific findings. Second, my focus on frequent co-citation in the scientific literature means that, in principle, those publications are equally citable in patents. It is unlikely that one paper is of much higher quality than its twin. Finally, the fact that I identify co-citation in the text of papers published in the top 15 journals means that the resulting sample of discoveries belongs to the disciplines that those journals publish; that is, the natural sciences in general, with a focus on the life sciences. Those fields are likely to be particularly useful for technology development (Rosenberg and Nelson 1994). In addition, the fact that at least one firm took part in each of the 39 set of paper twins further points to the commercial value of those projects.

3.3. How similar are the twins?

The question of the comparability of simultaneous discoveries was a topic of fierce debates in the 1960s and 1970s (e.g., Schmookler 1966; Merton 1968). Merton (1968: 9–10) noted: "It is no easy matter to establish the degree of similarity between independently developed ideas. Even in the more exact disciplines, such as mathematics, claims of independent multiple inventions are vigorously debated. The question is, how much overlap should be taken to constitute 'identity'?" To avoid getting caught in past debates about identity or similarity, I built three measures of within-twin heterogeneity.

Publication month difference. The occurrence of races in science provides key evidence that the contents of scientific manuscripts overlap—i.e., they are substitutes rather than complements. Failure to submit by the time a competing paper has been published amounts to "getting scooped." Teams that have been scooped are generally constrained to submit to a lower-tier journal and/or to conduct extra work and analyses in order to publish their findings. This supplementary work could create heterogeneity within pairs of papers. As noted above, my algorithm matches papers based on co-citation and not on publication month. It is striking, therefore, that the paper twins in my final sample were published on average only 2.2 months apart. For 36 sets of paper twins (92%), the gap between publications was less than six months.

To recognize a tie in the race for priority, editors who receive papers describing the same discovery around the same time might decide to publish them back-to-back. Even though not all back-to-back publications correspond to simultaneous discoveries (Drahl 2014), simultaneous discoveries appear often back-to-back in scientific journals. Several well-known discoveries were disclosed this way, including the discovery of evolution through natural selection by Darwin and Wallace in the *Journal of the Proceedings of the Linnean Society of London,* published on August 20, 1858, and the discovery by both Richter and Ting of the J/ψ meson, published in *Physical Review Letters,* on December 2, 1974. Interestingly, my sample of 39 simultaneous discoveries includes 24 instances of back-to-back publications.

Semantic difference. Different teams are likely to use different words to describe even very similar discoveries. I measured the semantic similarity of each pair with the PubMed-related citation algorithm (PubMed 2016). For each scientific publication in PubMed, this algorithm ranks every other scientific article based on the words they have in common, adjusted for length. If the pairs in my sample were not closely related, they should use different words and should therefore be ranked far from each other. Specifically, I assessed semantic difference by measuring rank difference in the PubMed-related citation algorithm after dropping articles published more than a calendar year apart. PubMed ranks the two papers in a pair right next to each other for 18 (49%) of the 37 sets of paper twins from my sample that are included in PubMed. The rank difference is less than 10 for 33 (89%) pairs. To put these numbers in perspective, the number of papers included in PubMed between 1994 and 2007 is 7,866,467. Hence, a rank difference of less than 10 out of 1,685,671 on average means that two papers are in the top 0.0006% most-related papers for that time period from a semantic standpoint.

Stylistic differences. To assess the stylistic differences among papers disclosing the same discovery, I hired 10 postdoctoral researchers from prestigious universities with the appropriate background in the life sciences and in applied physics. Each was asked to evaluate 4 to 17 sets of paper twins and to write a few sentences highlighting what they perceived to be the main differences between the manuscripts—which had been previously anonymized. Three researchers assessed each set of paper twins. To be conservative, I considered two papers to be different along one dimension even if only one of the researchers noted a difference. The researchers identified five types of difference and I translated those descriptions into five dummy variables capturing variance in (a) level of detail, (b) richness of the theory, (c) general sophistication, (d) practical emphasis, and (e) clarity.

[Insert Table 3 about here]

Table 3 investigates the differences between the scientific style of the publications "made in academia" and those "made in industry." The left part of the table presents information about how often the postdoctoral researchers found that paper twins differed along each dimension. The most common difference is the level of detail. Those experts highlighted that one paper appeared more detailed than its twin in 32% of the cases. The second most common difference is that one paper has more of a practical emphasis—that is, an emphasis on practical implications—in 19% of the cases. In contrast, differences in theory development (8%), general sophistication (7%), and clarity (4%) appeared to be less common. The remaining columns further examine whether this heterogeneity actually means that academic publications and their industrial twins differ significantly. They present the means and standard deviation of each variable for papers of industrial and academic origin respectively followed by a Wilcoxon-Mann-Whitney test of the statistical significance of the difference in means. The academic papers appear significantly more detailed than their industrial twins. However, I do not find statistically significant differences in the papers' level of theory development, sophistication, clarity, or clinical orientation.

3.4. Dependent variable: Patent references to the scientific literature

I measure inventors' attention to science by examining patent references to the scientific literature. Those citations are known to be much less likely than patent-to-patent citations to be added by the examiner or used strategically (Lampe 2012). Roach and Cohen (2013: 521) explore the validity of patent references to the scientific literature as a measure of knowledge flow by comparing them to survey data and find that "compared to citations to other patents, citations to nonpatent references correspond much more closely to managers' reports of the use of public research." Note that I focus on references to paper twins. In those cases, disclosing one paper is technically enough from a legal standpoint, since the two papers disclose the same prior art and, under US Patent and Trademark Office Rule 56, patent applicants are not required to disclose "cumulative" information. The inventors I interviewed indicated that, in practice, when they cited only one paper of a set of paper twins, they cited the one that came to mind first; that is, the one they had paid the most attention to. A senior inventor at an East Coast-based biotechnology company noted:

> It's like checking the box. If you got one that serves the purpose, that's good enough. The fact that we missed one does not necessarily mean that we did not see the paper. It might just be lack of thoroughness.

Differential patent citations in the case of paper twins therefore appear to measure of the differential attention that those papers had received.

I tracked my main outcome measure, $Reference_{ijk}$, by examining the references in the patent literature to the 90 publications making up the 39 sets of paper twins. To find these references, I searched each paper in the Dataverse patent database (Li et al. 2014), using the first author's last name, the journal name, the publication year, and the paper's first-page number. I excluded all self-citations by screening out all instances in which inventors of patent *j* work in an organization in which one or more of paper *i*'s authors were also based. To correct for the fact that older sets of paper twins have had more time to be referenced, only patents awarded within five years of the discovery's publication were considered. $Reference_{ijk}$ is an indicator variable taking the value 1 if patent *j* cites paper *i* of set of paper twins *k*.

3.5. Empirical analysis

I examine inventors' referencing (or nonreferencing) of academic and industry paper twins in their patents. As described above, the unit of analysis is the patent–paper dyad. Since unobserved characteristics of the inventor or invention (such as the inventor's familiarity with the scientific literature) might be correlated with the origin of the scientific discovery (academic or industrial), I used fixed effects at the level of the referencing-patent/set-of-paper-twins dyad to avoid an omitted-variable bias. The binary nature of the outcome variable could be modeled using a logistic regression. However, considering the small number of observations per patent/set-of-paper-twins dyad, such a model would not be consistent. The well-known incidental parameter problem can be solved by using a conditional likelihood function instead of the usual maximum likelihood. I therefore carry out the estimation using a conditional logit model (Chamberlain 1982) to model inventors' choice to refer to paper *i* of set of paper twins *k* in patent *j*. My baseline empirical test for the impact of academic origin on whether patent *j* cites paper *i* of set of paper twins *k* is

$$Reference_{ijk} = f(\varepsilon_{ijk};\ \alpha_0 + \alpha_1 Academic\ origin_i + \alpha_2 \bar{X}_{ij} + \gamma_{jk})$$

where γ_{jk} is a fixed effect for patent *j* citing set of paper twins *k*, $\alpha_2 \bar{X}_{ij}$ is a vector of control variables, and $Academic\ origin_i$ is my main explanatory variable. Robust standard errors are clustered at the level of the set of paper twins.

4. RESULTS

4.1 Sample Characteristics

Table 4 shows the summary statistics for the 924 dyads. It presents the means and standard errors of the main variables for the full sample as well as for the "industrial origin" and "academic origin" dyads separately. Sixty-eight percent of the potential citations are realized and one can already see that the mean of citation is considerably higher for "industrial origin" dyads (74%) than for "academic origin" ones (62%). The fact that the papers in the sample tend to have practical implications is visible in the high rate of patent–paper pairs (61%). Note also the high impact factor (23.17) of the journals in which paper twins are published. Of the 90 papers included in this study, 40 were published in *Nature*, *Cell*, and *Science*. The other papers were published in *Nature Genetics*, *Immunity*, *PNAS*, *Journal of Biological Chemistry*, *Nature Immunology*, *Journal of Experimental Medicine*, *Molecular Cell*, *Current Biology*, *Molecular and Cellular Biology*, *Nature Materials*, *Physical Review Letters*, *Toxicological Sciences*, *Genes & Development*, *Journal of Leukocyte Biology*, *Nature Medicine*, *NEJM*, *Physical Review B*, *American Journal of Human Genetics*, *EMBO Journal*, and *Nature Biotechnology*. The average number of authors per paper is 17.3 due to large collaborations in several gene-sequencing sets of paper twins. The set with the largest collaborations was for the first sequencing of a full Arabidopsis thaliana chromosome in 1999 by a team led by Celera Genomics (37 authors) and an academic consortium (231 authors)—and my results are robust to its exclusion. The articles in the data were published between 1994 and 2007 and the patents' application dates were between 1995 and 2008. The sum of each author's count of past publications is 347 papers on average per authoring team by the time of patent application.

[Insert Table 4 about here]

4.2. Main effect

Table 5 presents my main results. Note that the number of observation is not 924 but 523. In 401 (43%) of the 924 observations, inventors cite both paper twins in their patent. The patent/set-of-paper-twins fixed effects mean that those observations are dropped from the conditional logit regressions. Model 5-1 shows the main control variables. Inventors tend to cite US-based publications more, which suggests the possible existence of a "country-of-origin effect" (e.g., Verlegh and Steenkamp 1999) whereby inventors systematically pay more attention to science "made in America." However, this apparent US bias might also be driven by my choice of the US Patent and Trade Office as a data source. Note also that my measures of a paper's style—such as clarity, level of detail, and emphasis on practical implications—appear to predict citation, whereas author characteristics (such as past publications and past patents) do not. This absence of author-level effect seems surprising. It is worth noting that inventors acquire scientific knowledge primarily through open channels such as journal articles, informal interactions, and conferences (Cohen, Nelson, and Walsh 2002). They might not use author characteristics as a cue because they do not know those authors or because they do not consider that past publication and patenting activity are good signals of the value of a person's current work.

Model 5-2 tests my main hypothesis—that inventors are less likely to cite a paper in a patent if that paper emerges from academia (as opposed to industry). Indeed, I find that *Academic origin* is a strong and negative predictor of citation ($p<0.006$). The effect is large. When inventors cite only one of the paper twins in a patent, they are twice as likely to cite the corporate paper as they are to cite its academic twin. After accounting for the cases in which all the papers in a set of twins are cited, this amounts to a 23% discount in the citation of papers "made in academia" as opposed to those originating from industry. Model 5-3 tests the same hypothesis but using a linear probability model (LPM). Because LPM, unlike conditional logit, does not drop instances in which inventors cite all the papers of a set of twins, I can use my entire dataset of 924 patent/paper dyads. This advantage comes at a cost, however. Since I do not observe cases in which inventors could have cited a set of twins but did not cite any of the papers in that set, the introduction of the cases in which all paper twins are cited might actually bias my estimates. I therefore use conditional logit in the rest of my analysis, but my results are generally robust to using LPM.

[Insert Table 5 about here]

4.3. Robustness analysis

In light of the small number of sets of paper twins in my sample, one might be concerned that a few outliers drive my results. I examine this possibility in Table 6. First, Models 6-1 and 6-2 exclude, respectively, higher-order sets of paper twins (triplets and quadruplets) and the sets of paper twins that receive the most patent citations. To eliminate the cases in which the papers are likely to be the least similar, Models 6-3 and 6-4 exclude, respectively, the sets of paper twins in which the papers appeared more than 10 ranks apart in PubMed's related citation ranking and those in which the papers were published more than six months apart. Finally, to eliminate concerns relating to specific authorship patterns, Models 6-5 and 6-6 exclude, respectively, the sets of paper twins in which at least one paper involved more than 20 authors and those in which at least one corresponding author had a dual academia-industry affiliation. My main result remains strong across all these specifications.

[Insert Table 6 about here]

4.4. Mechanism

Table 7 attempts to shed more light on the mechanism underlying the observed institution-of-origin effect. First, inventors might pay more attention to publications from industry simply because they are better connected to the authors of those publications. I investigate the role of social networks in several ways. Model 7-1 excludes every instance in which any of the inventors on the patent had ever been a coauthor of any of the paper twins' authors. The result remains strong ($p<0.036$). Model 7-2 adds controls for whether the team is mixed between industry and academia and for the stock of every author's prior publications with industry. Admittedly, those authors who collaborate with individuals in firms are likely to be, on average, better connected with the community of inventors. Again, the main result on *Academic origin* remains strong (p<0.004). Model 7-3 examines whether the effect holds for both inventors from firms and those from academia. After all, academic inventors are likely to be better connected with other academics whereas industry inventors are likely to be better connected with scientists working in other firms. I find that the main result holds for both academic (p<0.078) and corporate (p<0.024) inventors. In unreported analyses, available upon request, I also use a number of different measures of geographic distance; my results remain unchanged.

[Insert Table 7 about here]

Model 7-4 examines whether the effect holds for patents that emerge right after the discovery, for those that emerge two to three years later, and for those that emerge four to five years later. The effect holds across the board, but appears to weaken over time (respectively, p<0.023, p<0.035, and p<0.085). This apparent weakening of the main effect is consistent with the idea that the uncertainty surrounding the discovery dissipates over time and that the institution-of-origin cue therefore becomes less important. Finally, Model 7-5 considers whether the effect is visible for both high- and low-status academic institutions. Inventors might pay more attention to higher-status academic institutions and the effect might therefore be less salient for sets of twins involving those institutions. To measure academic status at the time of invention, I considered the institutions that, according to Scopus, were the top 10 sources of publications in *Science* and *Nature* in the years in which the patents emerged (1995 to 2008). Those high status institutions were UC Berkeley, MIT, Harvard University, Oxford University, Harvard Medical School, University of Cambridge, Stanford University, University of Washington (Seattle), UC San Francisco, and Caltech. I then split my *Academic origin* variable in two, depending on whether or not any author of the academic publication was at any of those top 10 institutions. Model 7-5 shows that the effect of *Academic origin* is statistically significant in both cases (p<0.030 and p<0.055, respectively). Prestigious academic institutions do not seem to be immune from the negative effect of academic origin, perhaps in part because top universities are disproportionately involved in retractions (Furman, Jensen, and Murray 2012). Besides, the results from the splines in Models 7-3 to 7-5 should be interpreted carefully. My sample of paper twins is small and the difference in the splines' coefficients might in part reflect differences in the underlying sets of paper twins. Overall, the institution-of-origin effect does not seem to be driven by social networks, inventor type, or social status.

At the same time, my interviews highlighted that publications made in academia generally inspire less confidence among inventors than those stemming from industry. A senior scientist at a fast-growing East Coast biotechnology firm noted:

> If we see that a competing company has published promising data on a project, we will put a lot of weight on that data. We will think that we should pursue this, whereas if the same paper came from an academic lab, I think we would be a lot more skeptical. Is it real or perception, I don't know, but I think that there is some reality to it based on perception in the industry of things not reproducing when you try to bring the project internally.

The academic scientists that I interviewed were well aware of this skepticism. A senior immunologist based in Canada observed:

> One of the issues that we're running into is that companies all think that—at least the ones I'm dealing with—that the academics are all just—we're kind of fudging our data to get their money. So they're kind of suspicious of us. For whatever reason, they don't trust the academics. And it kind of bothers me and my colleagues because science is built on your reputation.

When asked why they have so much more faith in the value of research produced in industry, my interviewees consistently discussed the institutional differences between industry and academia. A senior scientist at a large West Coast biotechnology firm explained:

> The principle that I follow is that in academia, the end game is to get the paper published in a as high-profile journal as possible. In industry, the end game is not to get a paper published. The end game is getting a drug approved. It's much, much, much harder, okay? Many, many more hurdles along the way. And so it's a much higher bar—higher standards—because every error, or every piece of fraud along the way, the end game is going to fail. It's not gonna work. Therefore, I have more faith in what industry puts out there as a publication.

The results from my interviews therefore complemented my quantitative analysis by highlighting that the institution of origin effect often takes the form of skepticism. The inventors that I interviewed appear to have generally less confidence in the reproducibility of scientific work when it was made in academia than when it was made in industry.

5. DISCUSSION AND CONCLUSIONS

Inventors use a discovery's institutional environment of origin—whether industry or academia—as a cue to how to allocate their attention to the scientific literature. Science-based invention involves considerable uncertainty. Not only is scientific knowledge complex and tacit, but the scientific literature is vast, changes rapidly, and is often unreliable. Boundedly rational inventors are likely to struggle in gauging which new discoveries deserve their attention. I argue that inventors resort to informational cues to resolve this uncertainty and to identify the discoveries that they will attend to. The institutional environment of origin is one such cue. Scientists in industry face strong incentives to produce knowledge that will have commercial value and that will be replicable, while academics tend to be shielded from commercial pressures. Inventors, aware of those institutional differences, are likely to pay significantly less attention to discoveries "made in academia" than to those "made in industry."

Because scientists in firms work on more applied projects than their academic colleagues, inventors might naturally pay more attention to their publications anyway. However, I hypothesize that the institution shapes the amount of attention that a discovery receives, above and beyond the amount warranted by the nature of the discovery. In other words, I propose that the same discovery receives different levels of attention depending on whether it emerges in industry or in academia. To test this proposition, I use a novel empirical approach based on the occurrence of simultaneous discoveries in science. Specifically, I focus on simultaneous discoveries in which at least one discovery emerged from academia and one from industry. I measure inventors' differential attention to these papers by tracking citations in patents and find that inventors are more likely to refer to a scientific publication if it was "made in industry" than if it was "made in academia." The size of the effect is large. In the 57% of cases in which inventors do not refer to both paper twins, they are twice as likely to refer to the firm paper than to its academic twin. This effect is robust to a number of specifications. It does not appear to be driven by social networks, status, location, or subtle differences between the paper twins. Moreover, the effect holds whether the inventor is based in academia or in industry.

This manuscript makes several contributions. First, it highlights a hitherto underrecognized bottleneck in the translation of science into new technologies. Prior literature has focused on the challenges in translating tacit knowledge (e.g., Cockburn and Henderson 1998; Zucker, Darby, and Armstrong 2002). My findings complement this work by highlighting another challenge in this process. The scientific literature is vast, fast-changing, and often unreliable, while inventors' attention is scarce. Promising scientific discoveries might therefore not garner the attention they deserve if they emerge from researchers working under inventors' radar. For scientists, inventors' attention is likely to influence how much impact their research has in industry. More broadly, the way inventors manage their scarce attention is likely to affect the number and type of scientific discoveries that will be translated into new technologies.

Innovators' scarce attention is therefore likely to shape the evolution of technology. In exploring the cognitive drivers of technological change, prior research on technology evolution has focused on the equivocal nature of technology; that is, the actors' tendency to interpret new knowledge and technology in different and sometimes conflicting ways (Kaplan and Tripsas 2008; Murray and Kaplan 2010). Attention scarcity is another cognitive driver of technology evolution but it remains poorly understood. Yet, my results point to the importance of attention as a prerequisite to cumulative innovation. The way innovators allocate their attention will determine which interpretations they consider. Here I explore the use of cognitive cues in this process, but other elements, such as geographic location (Bikard and Marx 2018), and innovators' degree of specialization (Teodoridis, Bikard, and Vakili 2017) are also likely to influence how they allocate their attention.

Second, my results highlight the existence of an institution-of-origin effect shaping the process of science-based invention. Prior literature, by examining the role of informational cues in the cumulative processes of science and technology separately (Merton 1968; Podolny and Stuart 1995; Simcoe and Waguespack 2011; Azoulay, Stuart, and Wang 2013; Polidoro 2013), has overlooked some of the challenges that the institutional boundary raises for cumulative work across institutions. By focusing on science-based invention across academia and industry, I can identify the role of institutional origin as an informational cue and can document one way in which institutional boundaries affect cumulative work.

Third, this is the first paper to use simultaneous discoveries to conduct what amounts to a "twin study of scientific knowledge." Variance in the nature of scientific discoveries is an important challenge in empirical studies of science and its dissemination (e.g., Azoulay, Ding, and Stuart 2009). An empiricist who wishes to study the impact of location or of team composition, to take two examples, always faces a challenge: it is relatively easy to observe a treated group but very difficult to identify a suitable control. To solve this issue, one can use simultaneous discoveries as a series of natural experiments in which both the treated and the control deal with essentially the same underlying knowledge. This approach opens the door to studying important issues in a new way. Thus, this paper also attempts to contribute a new method to the toolbox of researchers interested in scientific knowledge and its dissemination.

This study also has implications for firms' innovation strategies. The benefits of academic knowledge for firms' R&D productivity are well known and there is a productive division of labor in which firms commercialize insights from academic labs (Rosenberg and Nelson 1994; Aghion, Dewatripont, and Stein 2008; Arora, Belenzon, and Patacconi 2017). However, the finding that inventors pay (relatively) little attention to work made in academia highlights an important friction in this division of labor. Taken together, the benefits of science for innovation and the challenge in identifying useful academic knowledge mean that firms' allocation of attention is likely to be a source of competitive advantage. Innovation performance will be shaped by a firm's ability to pay attention to valuable academic discoveries and to avoid paying attention to less-valuable ones. Thus, this paper raises important questions about how firms allocate their attention to science and how those choices compare with the actual distribution of useful discoveries.

The findings from this research have their limits, opening several doors for follow-up studies. Most importantly, this study relies on citations in the patent literature to a relatively small sample of 39 sets of paper twins emerging in academia and in industry, primarily in the life sciences. Although my focus on paper twins makes for a relatively clean empirical strategy, it also makes it difficult to assess what would have happened had there been only one source for the knowledge. On one hand, the interviews suggest that the presence of paper twins improves inventors' confidence in the reliability of a scientific discovery; my results might therefore understate the negative impact of academic origin. On the other hand, I cannot examine the effect of institutional origin for cases in which no paper twin exists. In the absence of a paper twin, the baseline against which citation rates could be compared remains elusive.

In addition, these results do not allow any normative conclusion about whether or not inventors should pay more attention to discoveries "made in academia." Systematically paying less attention to discoveries made in academia than to those made in industry could be a good strategy in light of the challenges in identifying academic publications that are promising and replicable. Future studies might explore the extent to which publications from academia are indeed less replicable than those originating in industry. They might also examine how replication studies affect inventors' attention.

Finally, the fact that the vast majority of the inventors in my data are life scientists raises questions about the generalizability of my findings to other scientific fields. The crucial role of scientific research in the life-science industry is visible not only in patent citations but also in surveys and in collaborations between industry and academia (Henderson, Jaffe, and Trajtenberg 1998; Zucker, Darby, and Brewer 1998; Cohen, Nelson, and Walsh 2002). The fact that I find an institution-of-origin effect in a field in which industry and academia have such a tight relationship raises the possibility that the effect might be even larger in other fields, but this remains an open question.

In principle, there is no reason to expect that this institution-of-origin effect is one-sided. Just as inventors are likely to pay more attention to publications "made in industry," one might expect that scientists pay more attention to publications "made in academia." After all, there is evidence that scientists working with firms produce papers that are poorer from a theoretical standpoint (Evans 2010) and that they are more guarded about information disclosure (Blumenthal et al. 1996). Though my analysis allows me to test only one side of the institution-of-origin effect, there is anecdotal evidence of the other side. Seventy years ago, J. R. Oppenheimer (1947: 175–76) stated an opinion that was echoed by some of my interviewees: "What you run into when you mix developmental activities, which are programmed, planned and focused, with activities which are of an inquiring and scientific nature, is the rivalry of two incompatible snobisms. The scientist is irritated by the practical preoccupations of the man concerned with development, and the man concerned with development thinks that the scientist is lazy and of no account and is not doing a real job anyway." One should also note that the institution-of-origin effect might take different forms at different institutional boundaries.

This study is only a first step toward a better understanding of the downsides of academia when it comes to disseminating scientific knowledge to inventors. Prior studies have primarily focused on its upsides; in particular, its traditional openness and emphasis on teaching (Bush 1945). However, phenomena such as retractions, fraud, and general lack of reproducibility in science (Azoulay et al. 2012; Furman, Jensen, and Murray 2012; *The Economist* 2013) have undermined the legitimacy of knowledge "made in academia." The challenge that inventors face in allocating their attention to academic science therefore raises the question of whether new institutions could be imagined that would systematically assess the shakiness of prominent scientific discoveries from the inventors' perspective. The importance of continuing research on this topic should not be understated. As individuals, firms, and nations continue to rely on economic spillovers from academic research (e.g., Arora, Belenzon, and Patacconi 2017), it is important to recognize that inventors' scarce attention affects the rate and direction of science-based invention.

REFERENCES

Aghion, Philippe, Mathias Dewatripont, and Jeremy C. Stein. 2008. "Academic Freedom, Private-Sector Focus, and the Process of Innovation." *The RAND Journal of Economics* 39 (3):617–35.

Agrawal, Ajay, Avi Goldfarb, and Florenta Teodoridis. 2016. "Understanding the Changing Structure of Scientific Inquiry." *American Economic Journal: Applied Economics* 8 (1):100–128.

Alberts, Bruce. 2010. "Is the Frontier Really Endless?" *Science* 330 (6011):1587–1587.

Arora, Ashish, Sharon Belenzon, and Andrea Patacconi. 2017. "The Decline of Science in Corporate R&D." *Strategic Management Journal*.

Arrow, K. 1962. "Economic Welfare and the Allocation of Resources for Invention." In *R.R. Nelson (Ed.), The Rate and Direction of Inventive Activity*, 609–25. Princeton, NJ: Princeton University Press.

Azoulay, Pierre. 2002. "Do Pharmaceutical Sales Respond to Scientific Evidence?" *Journal of Economics & Management Strategy* 11 (4):551–94.

Azoulay, Pierre, Waverly W. Ding, and Toby Stuart. 2009. "The Impact of Academic Patenting on the Rate, Quality, and Direction of (Public) Research Output." *The Journal of Industrial Economics* 57 (4):637--676.

Azoulay, Pierre, Jeffrey Furman, Joshua L. Krieger, and Fiona E. Murray. 2012. “Retractions.” Working Paper 18499. National Bureau of Economic Research. http://www.nber.org/papers/w18499.
Azoulay, Pierre, Toby Stuart, and Yanbo Wang. 2013. “Matthew: Effect or Fable?” *Management Science* 60 (1):92–109.
Begley, C. Glenn, and Lee M. Ellis. 2012. “Drug Development: Raise Standards for Preclinical Cancer Research.” *Nature* 483 (7391):531–33.
Begley, C. Glenn, and John P. A. Ioannidis. 2015. “Reproducibility in Science: Improving the Standard for Basic and Preclinical Research.” *Circulation Research* 116 (1):116–26.
Belenzon, Sharon, and Mark Schankerman. 2013. “Spreading the Word: Geography, Policy, and Knowledge Spillovers.” *Review of Economics and Statistics* 95 (3):884–903.
Bikard, Michaël. 2012. “Simultaneous Discoveries as a Research Tool: Method and Promise.” *MIT Sloan Working Paper*. Available at SSRN: https://ssrn.com/abstract=2329605
Bikard, Michaël, and Matt Marx. 2018. “Hubs As Lampposts: Academic Location and Firms’ Attention to Science.” SSRN Scholarly Paper ID 3006859. Rochester, NY: Social Science Research Network. https://papers.ssrn.com/abstract=3006859.
Blumenthal, David, Eric G. Campbell, Nancyanne Causino, and Karen Seashore Louis. 1996. “Participation of Life-Science Faculty in Research Relationships with Industry.” *N Engl J Med* 335 (23):1734–39.
Boudreau, Kevin J., Eva C. Guinan, Karim R. Lakhani, and Christoph Riedl. 2016. “Looking Across and Looking Beyond the Knowledge Frontier: Intellectual Distance, Novelty, and Resource Allocation in Science.” *Management Science* 62 (10):2765–83.
Boyack, Kevin W., Richard Klavans, Aaron A. Sorensen, and John P. A. Ioannidis. 2013. “A List of Highly Influential Biomedical Researchers, 1996-2011.” *European Journal of Clinical Investigation* 43 (12):1339–65.
Bush, Vannevar. 1945. “Science: The Endless Frontier.” *Transactions of the Kansas Academy of Science (1903-)* 48 (3):231–64.
Casadevall, Arturo, and Ferric C. Fang. 2012. “Reforming Science: Methodological and Cultural Reforms.” *Infection and Immunity* 80 (3):891–96.
Chai, Sen. 2017. “Near Misses in the Breakthrough Discovery Process.” *Organization Science* 28 (3):411–28.
Chamberlain, G. 1982. *Analysis of Covariance with Qualitative Data*. National Bureau of Economic Research Cambridge, Mass., USA.
Cockburn, Iain, and Rebecca Henderson. 1998. “Absorptive Capacity, Coauthoring Behavior, and the Organization of Research in Drug Discovery.” *The Journal of Industrial Economics* 46 (2):157–82.
Cohen, Wesley, and Daniel Levinthal. 1990. “Absorptive Capacity: A New Perspective on Learning and Innovation.” *Administrative Science Quarterly* 35 (1):128–52.
Cohen, Wesley, Richard Nelson, and John Walsh. 2002. “Links and Impacts: The Influence of Public Research on Industrial R&D.” *Management Science* 48 (1):1–23.
Cozzens, Susan E. 1989. *Social Control and Multiple Discovery in Science: The Opiate Receptor Case*. State University of New York Press.
Engelhardt, Hugo Tristram, and Arthur L. Caplan. 1987. *Scientific Controversies: Case Studies in the Resolution and Closure of Disputes in Science and Technology*. Cambridge University Press.
Evans, James A. 2010. “Industry Induces Academic Science to Know Less about More.” *The American Journal of Sociology* 116 (2):389–452.
Fini, Riccardo, Julien Jourdan, and Markus Perkmann. 2017. “Social Valuation across Multiple Audiences: The Interplay between Ability and Identity Judgments.” *Academy of Management Journal*, December, amj.2016.0661.
Fleming, Lee, and Olav Sorenson. 2004. “Science as a Map in Technological Search.” *Strategic Management Journal* 25:909–28.

Freedman, Leonard P., Iain M. Cockburn, and Timothy S. Simcoe. 2015. “The Economics of Reproducibility in Preclinical Research.” *PLOS Biology* 13 (6):e1002165.
Furman, Jeffrey, Kyle Jensen, and Fiona Murray. 2012. “Governing Knowledge in the Scientific Community: Exploring the Role of Retractions in Biomedicine.” *Research Policy* 41 (2):276–90.
Furman, Jeffrey, and Megan J. MacGarvie. 2007. “Academic Science and the Birth of Industrial Research Laboratories in the U.S. Pharmaceutical Industry.” *Journal of Economic Behavior & Organization* 63 (4):756–76.
Furman, Jeffrey, and Scott Stern. 2011. “Climbing atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research.” *American Economic Review* 101 (5):1933–63.
Gambardella, Alfonso. 1992. “Competitive Advantages from In-House Scientific Research: The US Pharmaceutical Industry in the 1980s.” *Research Policy* 21 (5):391–407.
Gittelman, Michelle, and Bruce Kogut. 2003. “Does Good Science Lead to Valuable Knowledge? Biotechnology Firms and the Evolutionary Logic of Citation Patterns.” *Management Science* 49 (4):366–82.
Hartshorne, Joshua K., and Adena Schachner. 2012. “Tracking Replicability as a Method of Post-Publication Open Evaluation.” *Frontiers in Computational Neuroscience* 6 (March).
Henderson, Rebecca, Adam Jaffe, and Manuel Trajtenberg. 1998. “Universities as a Source of Commercial Technology: A Detailed Analysis of University Patenting, 1965-1988.” *Review of Economics and Statistics* 80 (1):119–27.
Hippel, Eric von. 1994. “‘Sticky Information’ and the Locus of Problem Solving: Implications for Innovation.” *Management Science* 40 (4):429–39.
Jones, Benjamin. 2009. “The Burden of Knowledge and the ‘Death of the Renaissance Man’: Is Innovation Getting Harder?” *Review of Economic Studies* 76 (1):283–317.
Kaplan, Sarah, and Mary Tripsas. 2008. “Thinking about Technology: Applying a Cognitive Lens to Technical Change.” *Research Policy* 37 (5):790–805.
Lampe, Ryan. 2012. “Strategic Citation.” *Review of Economics and Statistics* 94 (1):320–33. https://doi.org/10.1162/REST_a_00159.
Li, Guan-Cheng, Ronald Lai, Alexander D’Amour, David M. Doolin, Ye Sun, Vetle I. Torvik, Amy Z. Yu, and Lee Fleming. 2014. “Disambiguation and Co-Authorship Networks of the U.S. Patent Inventor Database (1975–2010).” *Research Policy* 43 (6):941–55.
Merton, Robert K. 1957. “Priorities in Scientific Discovery: A Chapter in the Sociology of Science.” *American Sociological Review* 22 (6):635–59.
———. 1968. “The Matthew Effect in Science The Reward and Communication Systems of Science Are Considered.” *Science* 159 (3810):56.
Mobley, Aaron, Suzanne K. Linder, Russell Braeuer, Lee M. Ellis, and Leonard Zwelling. 2013. “A Survey on Data Reproducibility in Cancer Research Provides Insights into Our Limited Ability to Translate Findings from the Laboratory to the Clinic.” *PLOS ONE* 8 (5):e63221.
Mokyr, Joel. 2002. *The Gifts of Athena: Historical Origins of the Knowledge Economy*. Princeton University Press.
Murmann, J. P. 2003. *Knowledge and Competitive Advantage: The Coevolution of Firms, Technology, and National Institutions*. Cambridge University Press.
Murray, Fiona. 2010. “The Oncomouse That Roared: Hybrid Exchange Strategies as a Source of Distinction at the Boundary of Overlapping Institutions1.” *The American Journal of Sociology* 116 (September):341–88.
Murray, Fiona, and Sarah Kaplan. 2010. “Entrepreneurship and the Construction of Value in Biotechnology.” In *Technology and Organization: Essays in Honour of Joan Woodward*, 29:107–47. Research in the Sociology of Organizations 29. Emerald Group Publishing Limited.
Nelson, Richard, and Sidney Winter. 1982. *An Evolutionary Theory of Economic Change*. Harvard University Press.
Ocasio, William. 1997. “Towards an Attention-Based View of the Firm.” *Strategic Management Journal* 18:187–206.

Oppenheimer, J. Robert. 1947. “Functions of the International Agency in Research and Development.” *Bulletin of the Atomic Scientists* 3 (7):173–97.

Podolny, Joel M., and Toby E. Stuart. 1995. “A Role-Based Ecology of Technological Change.” *The American Journal of Sociology* 100 (5):1224–60.

Polanyi, Michael. 1958. *Personal Knowledge*. London: Routledge. http://www.press.uchicago.edu/ucp/books/book/chicago/P/bo19722848.html.

Polidoro, Francisco. 2013. “The Competitive Implications of Certifications: The Effects of Scientific and Regulatory Certifications on Entries into New Technical Fields.” *Academy of Management Journal* 56 (2):597–627.

Polidoro, Francisco, and Matt Theeke. 2011. “Getting Competition Down to a Science: The Effects of Technological Competition on Firms’ Scientific Publications.” *Organization Science* 23 (4):1135–53.

PubMed. 2016. PubMed Related Citations Algorithm. March 21, 2016. https://ii.nlm.nih.gov/MTI/Details/related.shtml.

Rao, Hayagreeva. 2004. “Institutional Activism in the Early American Automobile Industry.” *Journal of Business Venturing* 19 (3):359–84.

Rosenberg, Nathan. 1990. “Why Do Firms Do Basic Research (with Their Own Money)?” *Research Policy* 19 (2):165–74.

———. 1994. *Exploring the Black Box: Technology, Economics, and History*. Cambridge University Press.

Rosenberg, Nathan, and Richard Nelson. 1994. “American Universities and Technical Advance in Industry.” *Research Policy* 23 (3):323–48.

Rosenblatt, Michael. 2016. “An Incentive-Based Approach for Improving Data Reproducibility.” *Science Translational Medicine* 8 (336):336ed5-336ed5. https://doi.org/10.1126/scitranslmed.aaf5003.

Schmookler, J. 1966. *Invention and Economic Growth*. Harvard University Press Cambridge, MA.

Simcoe, Timothy S., and Dave M. Waguespack. 2011. “Status, Quality, and Attention: What’s in a (Missing) Name?” *Management Science* 57 (2):274–90. https://doi.org/10.1287/mnsc.1100.1270.

Simon, Herbert. 1947. *Administrative Behavior*. Free Press.

Stern, Scott. 2004. “Do Scientists Pay to Be Scientists?” *Management Science* 50 (6):835–53.

Stokes, Donald E. 1997. *Pasteur’s Quadrant: Basic Science and Technological Innovation*. Brookings Institution Press.

Teodoridis, Florenta. 2017. “Understanding Team Knowledge Production: The Interrelated Roles of Technology and Expertise.” *Management Science*, July.

Teodoridis, Florenta, Michaël Bikard, and Keyvan Vakili. 2017. “The Pace of Change and Creative Performance: Specialist and Generalist Mathematicians at the Fall of the Soviet Union.” SSRN Scholarly Paper ID 2911141. Rochester, NY: Social Science Research Network. https://papers.ssrn.com/abstract=2911141.

The Economist. 2013. “How Science Goes Wrong,” October 19, 2013. http://www.economist.com/news/leaders/21588069-scientific-research-has-changed-world-now-it-needs-change-itself-how-science-goes-wrong.

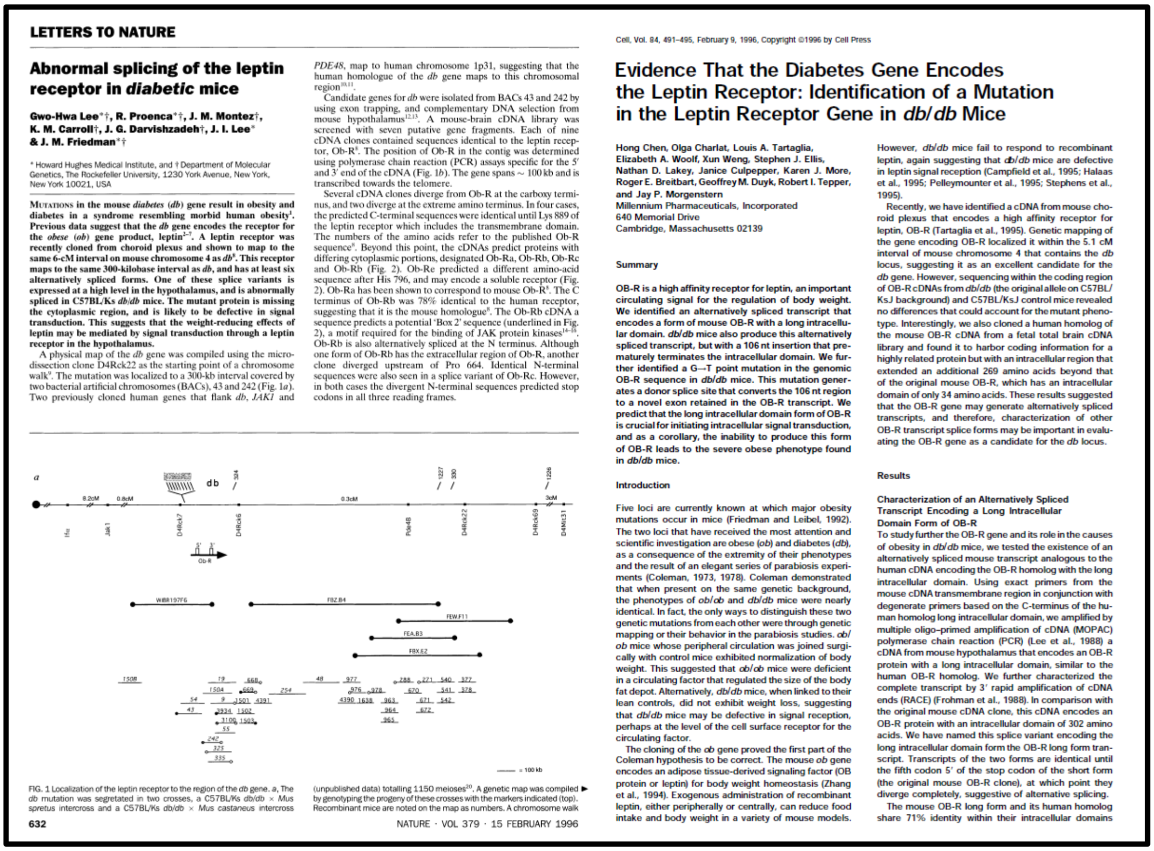
Verlegh, Peeter W. J., and Jan-Benedict E. M. Steenkamp. 1999. “A Review and Meta-Analysis of Country-of-Origin Research.” *Journal of Economic Psychology* 20 (5):521–46.

Zucker, Lynne G., and Michael R. Darby. 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 (23):12709–16.

Zucker, Lynne G., Michael R. Darby, and Jeff S. Armstrong. 2002. “Commercializing Knowledge: University Science, Knowledge Capture, and Firm Performance in Biotechnology.” *Management Science* 48 (1):138–53.

Zucker, Lynne G., Michael R. Darby, and Marilynn B. Brewer. 1998. “Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises.” *The American Economic Review* 88 (1):290–306.

Figure 1. A set of "paper twins"



LETTERS TO NATURE

Abnormal splicing of the leptin receptor in *diabetic* mice

Gwo-Hwa Lee*†, R. Proenca*†, J. M. Montez†, K. M. Carroll†, J. G. Darvishzadeh†, J. I. Lee* & J. M. Friedman*†

* Howard Hughes Medical Institute, and † Department of Molecular Genetics, The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA

Mutations in the mouse *diabetes* (*db*) gene result in obesity and diabetes in a syndrome resembling morbid human obesity[1]. Previous data suggest that the *db* gene encodes the receptor for the *obese* (*ob*) gene product, leptin[2–7]. A leptin receptor was recently cloned from choroid plexus and shown to map to the same 6-cM interval on mouse chromosome 4 as *db*[8]. This receptor maps to the same 300-kilobase interval as *db*, and has at least six alternatively spliced forms. One of these splice variants is expressed at a high level in the hypothalamus, and is abnormally spliced in C57BL/Ks *db*/*db* mice. The mutant protein is missing the cytoplasmic region, and is likely to be defective in signal transduction. This suggests that the weight-reducing effects of leptin may be mediated by signal transduction through a leptin receptor in the hypothalamus.

A physical map of the *db* gene was compiled using the micro-dissection clone D4Rck22 as the starting point of a chromosome walk[9]. The mutation was localized to a 300-kb interval covered by two bacterial artificial chromosomes (BACs), 43 and 242 (Fig. 1*a*). Two previously cloned human genes that flank *db*, *JAK1* and *PDE4B*, map to human chromosome 1p31, suggesting that the human homologue of the *db* gene maps to this chromosomal region[10,11].

Candidate genes for *db* were isolated from BACs 43 and 242 by using exon trapping, and complementary DNA selection from mouse hypothalamus[12,13]. A mouse-brain cDNA library was screened with seven putative gene fragments. Each of nine cDNA clones contained sequences identical to the leptin receptor, Ob-R[8]. The position of Ob-R in the contig was determined using polymerase chain reaction (PCR) assays specific for the 5′ and 3′ end of the cDNA (Fig. 1*b*). The gene spans ~ 100 kb and is transcribed towards the telomere.

Several cDNA clones diverge from Ob-R at the carboxy terminus, and two diverge at the extreme amino terminus. In four cases, the predicted C-terminal sequences were identical until Lys 889 of the leptin receptor which includes the transmembrane domain. The numbers of the amino acids refer to the published Ob-R sequence[8]. Beyond this point, the cDNAs predict proteins with differing cytoplasmic portions, designated Ob-Ra, Ob-Rb, Ob-Rc and Ob-Rb (Fig. 2). Ob-Re predicted a different amino-acid sequence after His 796, and may encode a soluble receptor (Fig. 2). Ob-Ra has been shown to correspond to mouse Ob-R[8]. The C terminus of Ob-Rb was 78% identical to the human receptor, suggesting that it is the mouse homologue[8]. The Ob-Rb cDNA a sequence predicts a potential 'Box 2' sequence (underlined in Fig. 2), a motif required for the binding of JAK protein kinases[14–16]. Ob-Rb is also alternatively spliced at the N terminus. Although one form of Ob-Rb has the extracellular region of Ob-R, another clone diverged upstream of Pro 664. Identical N-terminal sequences were also seen in a splice variant of Ob-Rc. However, in both cases the divergent N-terminal sequences predicted stop codons in all three reading frames.

FIG. 1 Localization of the leptin receptor to the region of the *db* gene. *a*, The *db* mutation was segretated in two crosses, a C57BL/Ks *db*/*db* × *Mus spretus* intercross and a C57BL/Ks *db*/*db* × *Mus castaneus* intercross (unpublished data) totalling 1150 meioses[20]. A genetic map was compiled by genotyping the progeny of these crosses with the markers indicated (top). Recombinant mice are noted on the map as numbers. A chromosome walk ▶

632 NATURE · VOL 379 · 15 FEBRUARY 1996

Cell, Vol. 84, 491–495, February 9, 1996, Copyright ©1996 by Cell Press

Evidence That the Diabetes Gene Encodes the Leptin Receptor: Identification of a Mutation in the Leptin Receptor Gene in *db*/*db* Mice

Hong Chen, Olga Charlat, Louis A. Tartaglia, Elizabeth A. Woolf, Xun Weng, Stephen J. Ellis, Nathan D. Lakey, Janice Culpepper, Karen J. More, Roger E. Breitbart, Geoffrey M. Duyk, Robert I. Tepper, and Jay P. Morgenstern
Millennium Pharmaceuticals, Incorporated
640 Memorial Drive
Cambridge, Massachusetts 02139

Summary

OB-R is a high affinity receptor for leptin, an important circulating signal for the regulation of body weight. We identified an alternatively spliced transcript that encodes a form of mouse OB-R with a long intracellular domain. *db*/*db* mice also produce this alternatively spliced transcript, but with a 106 nt insertion that prematurely terminates the intracellular domain. We further identified a G→T point mutation in the genomic OB-R sequence in *db*/*db* mice. This mutation generates a donor splice site that converts the 106 nt region to a novel exon retained in the OB-R transcript. We predict that the long intracellular domain form of OB-R is crucial for initiating intracellular signal transduction, and as a corollary, the inability to produce this form of OB-R leads to the severe obese phenotype found in *db*/*db* mice.

Introduction

Five loci are currently known at which major obesity mutations occur in mice (Friedman and Leibel, 1992). The two loci that have received the most attention and scientific investigation are obese (*ob*) and diabetes (*db*), as a consequence of the extremity of their phenotypes and the result of an elegant series of parabiosis experiments (Coleman, 1973, 1978). Coleman demonstrated that when present on the same genetic background, the phenotypes of *ob*/*ob* and *db*/*db* mice were nearly identical. In fact, the only ways to distinguish these two genetic mutations from each other were through genetic mapping or their behavior in the parabiosis studies. *ob*/*ob* mice whose peripheral circulation was joined surgically with control mice exhibited normalization of body weight. This suggested that *ob*/*ob* mice were deficient in a circulating factor that regulated the size of the body fat depot. Alternatively, *db*/*db* mice, when linked to their lean controls, did not exhibit weight loss, suggesting that *db*/*db* mice may be defective in signal reception, perhaps at the level of the cell surface receptor for the circulating factor.

The cloning of the *ob* gene proved the first part of the Coleman hypothesis to be correct. The mouse *ob* gene encodes an adipose tissue-derived signaling factor (OB protein or leptin) for body weight homeostasis (Zhang et al., 1994). Exogenous administration of recombinant leptin, either peripherally or centrally, can reduce food intake and body weight in a variety of mouse models. However, *db*/*db* mice fail to respond to recombinant leptin, again suggesting that *db*/*db* mice are defective in leptin signal reception (Campfield et al., 1995; Halaas et al., 1995; Pelleymounter et al., 1995; Stephens et al., 1995).

Recently, we have identified a cDNA from mouse choroid plexus that encodes a high affinity receptor for leptin, OB-R (Tartaglia et al., 1995). Genetic mapping of the gene encoding OB-R localized it within the 5.1 cM interval of mouse chromosome 4 that contains the *db* locus, suggesting it as an excellent candidate for the *db* gene. However, sequencing within the coding region of OB-R cDNAs from *db*/*db* (the original allele on C57BL/KsJ background) and C57BL/KsJ control mice revealed no differences that could account for the mutant phenotype. Interestingly, we also cloned a human homolog of the mouse OB-R cDNA from a fetal total brain cDNA library and found it to harbor coding information for a highly related protein but with an intracellular region that extended an additional 269 amino acids beyond that of the original mouse OB-R, which has an intracellular domain of only 34 amino acids. These results suggested that the OB-R gene may generate alternatively spliced transcripts, and therefore, characterization of other OB-R transcript splice forms may be important in evaluating the OB-R gene as a candidate for the *db* locus.

Results

Characterization of an Alternatively Spliced Transcript Encoding a Long Intracellular Domain Form of OB-R

To study further the OB-R gene and its role in the causes of obesity in *db*/*db* mice, we tested the existence of an alternatively spliced mouse transcript analogous to the human cDNA encoding the OB-R homolog with the long intracellular domain. Using exact primers from the mouse cDNA transmembrane region in conjunction with degenerate primers based on the C-terminus of the human homolog long intracellular domain, we amplified by multiple oligo–primed amplification of cDNA (MOPAC) polymerase chain reaction (PCR) (Lee et al., 1988) a cDNA from mouse hypothalamus that encodes an OB-R protein with a long intracellular domain, similar to the human OB-R homolog. We further characterized the complete transcript by 3′ rapid amplification of cDNA ends (RACE) (Frohman et al., 1988). In comparison with the original mouse cDNA clone, this cDNA encodes an OB-R protein with an intracellular domain of 302 amino acids. We have named this splice variant encoding the long intracellular domain form the OB-R long form transcript. Transcripts of the two forms are identical until the fifth codon 5′ of the stop codon of the short form (the original mouse OB-R clone), at which point they diverge completely, suggestive of alternative splicing.

The mouse OB-R long form and its human homolog share 71% identity within their intracellular domains

Figure 2. The structure of the dataset

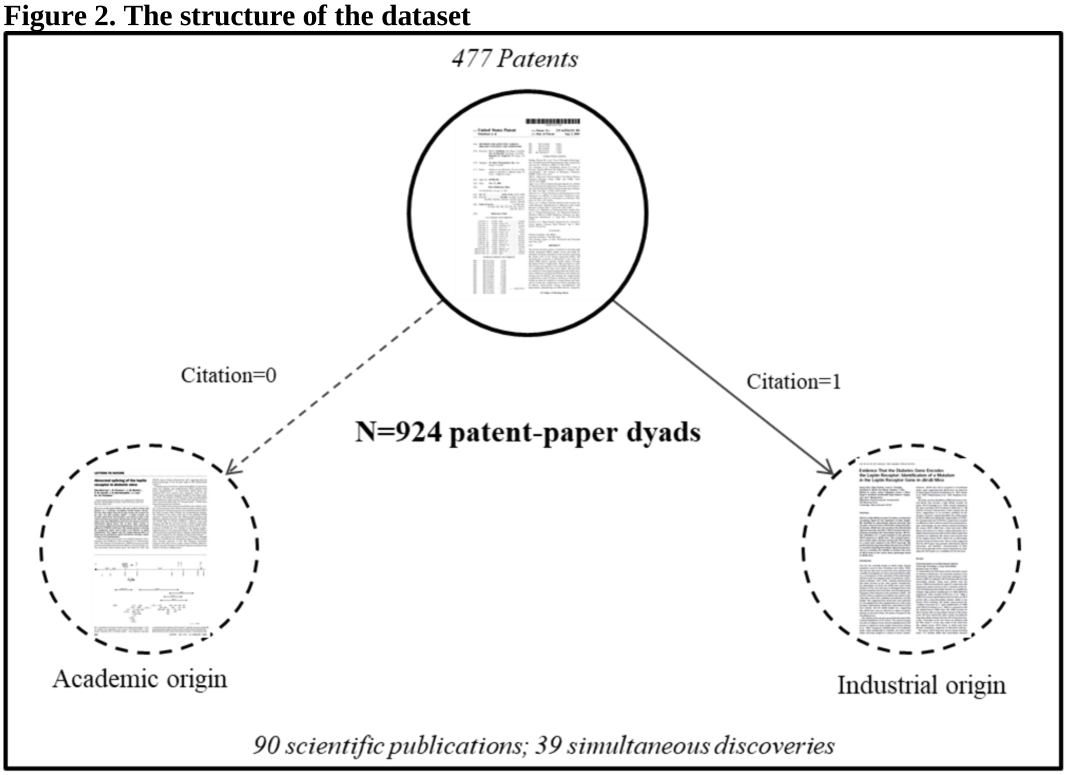


Table 1. Examples of sets of paper twins emerging from industry and academia

Discovery	Twin papers	Discovery organization
Identification of a molecule (FADD) that promotes apoptosis	Kischkel, F. et al. "Apo2L/TRAIL-dependent recruitment of endogenous FADD and caspase-8 to death receptors 4 and 5" *Immunity* 12, 611-620 (2000)	Genentech
	Sprick, M. et al. "FADD/MORT1 and caspase-8 are recruited to TRAIL receptors 1 and 2 and are essential for apoptosis mediated by TRAIL receptor 2" *Immunity* 12, 599-609 (2000)	German Cancer Research Center
Identification of a mutation in the leptin receptor gene Obr in db/db mice linking it to diabetes	Chen, H. et al. "Evidence That the Diabetes Gene Encodes the Leptin Receptor: Identification of a Mutation in the Leptin Receptor Gene in db/db Mice" *Cell* 84, 491-495 (1996)	Millennium Pharmaceuticals
	Lee, G. et al. "Abnormal splicing of the leptin receptor in diabetic mice" *Nature* 379, 632-635 (1996)	Rockefeller University
Antagonists to the cIAP protein trigger programmed cell death and activate important general signalling pathways	Varfolomeev, E. et al. "IAP Antagonists Induce Autoubiquitination of c-IAPs, NF-κB Activation, and TNFα-Dependent Apoptosis" *Cell* 131, 669-681 (2007)	Genentech
	Vince, J. et al. "IAP antagonists target cIAP1 to induce TNFalpha-dependent apoptosis" *Cell* 131, 682-693 (2007)	La Trobe University
Novel magnetic tunnel junctions based on Fe alloys and MgO that result in much higher magneto resistance values than ever reported at room temperature	Parkin, S. et al. "Giant tunnelling magnetoresistance at room temperature with MgO (100) tunnel barriers" *Nature Materials* 3, 862-867 (2004)	IBM
	Yuasa, S. et al. "Giant room-temperature magnetoresistance in single-crystal Fe/MgO/Fe magnetic tunnel junctions" *Nature Materials* 3, 868-871 (2004)	National Institute of Advanced Industrial Science and Technology
The protein TRPV3 is a cation channel protein sensitive to warm and hot temperatures	Smith, G. et al. "TRPV3 is a temperature-sensitive vanilloid receptor-like protein" *Nature* 418, 186-190 (2002)	GlaxoSmithKline
	Xu, H. et al. "TRPV3 is a calcium-permeable temperature-sensitive cation channel" *Nature* 418, 181-186 (2002)	Harvard University
	Peier, A. et al. "A heat-sensitive TRP channel expressed in keratinocytes" *Science* 296, 2046-2049 (2002)	Scripps

Table 2. Main variables and definitions

Variable	Definition	Source
Citation	Dummy variable equal to 1 if article *i* is cited in patent *j*; 0 otherwise	Harvard IQSS patent database
Academic origin	Dummy variable equal to 1 if paper *i*'s corresponding author is based in academia; 0 if he or she is based in a firm	Publication itself
Paper is more detailed	Dummy variable equal to 1 if at least one postdoctoral researcher believed that paper *i* is more detailed than any of its twin(s); 0 otherwise	Expert panel (see section 3.3)
Paper has richer theory	Dummy variable equal to 1 if at least one postdoctoral researcher believed that paper *i* has a richer theory than any of its twin(s); 0 otherwise	Expert panel (see section 3.3)

Paper is more sophisticated	Dummy variable equal to 1 if at least one postdoctoral researcher believed that paper *i* is more sophisticated than any of its twin(s); 0 otherwise	Expert panel (see section 3.3)
Paper has more practical emphasis	Dummy variable equal to 1 if at least one postdoctoral researcher believed that paper *i* has more emphasis on practical applications than its twin(s); 0 otherwise	Expert panel (see section 3.3)
Paper is clearer	Dummy variable equal to 1 if at least one postdoctoral researcher believed that paper *i* is clearer than its twin(s); 0 otherwise	Expert panel (see section 3.3)
US paper	Dummy variable equal to 1 if the corresponding author of article *i* is in the US; 0 otherwise	Publication itself
Journal impact factor	Impact factor (2009 value) of the journal that published article *i*	International Scientific Institute (ISI)
Patent-paper pair	Dummy variable equal to 1 if paper *i* has a patent-pair; 0 otherwise. Paper considered to have patent pair if there is at least one patent (of all USPTO patents) on which at least two of the paper's authors are listed as inventors and whose application date is the year of paper publication or the prior year.	Algorithm using Harvard IQSS patent database
Number of authors	Count of the number of authors on the paper (logged)	Publication itself
Authors' joint publication stock	Sum of the publication counts of each author^{+} listed on article *i* by the year of patent *j* application (logged)	Scopus
Authors' joint publication with industry	Sum of the counts of publications that included a firm in the address field for each author^{+} listed on article *i* by the year of patent *j* application (logged)	Scopus
Authors' joint patent stock	Sum of the patent counts awarded to each author^{+} listed on article *i* by the year of patent *j* application (logged)	Harvard IQSS patent database
Mixed academia-industry team	Dummy variable equal to 1 if article *i* includes authors from both academia and industry; 0 otherwise	Publication itself
Prior co-authorship	Dummy variable equal to 1 if any of the authors listed on article *i* has coauthored with an inventor on patent *j* by the year of patent application	Harvard IQSS patent database; Scopus
Time lag	Application year of patent *j* – Publication year of paper *i*	Harvard IQSS patent database; Web of Science
Geographic distance	Minimum geographic distance between all possible dyads between all authors of article *i* on the one hand and all the inventors of patent *j* on the other (logged)	Scopus; Harvard IQSS patent database; latlong.net
Same country	Dummy variable equal to 1 if the corresponding author of article *i* is located in the same country as the first inventor of patent *j*; 0 otherwise	Harvard IQSS patent database; Web of Science

$^{+}$ Based on the full publication and patent history of every author of every potentially cited paper with one exception—a paper that has 231 authors. For that paper, I used information about the first five authors, the last five authors, and the corresponding author. Results are robust to excluding that simultaneous discovery.

Table 3. Within-twin heterogeneity

Variable	Total (N=90) Mean	Std. dev.	Industrial origin (N=41) Mean	Std. dev.	Academic origin (N=49) Mean	Std. dev.	Mean diff. - Acad. vs. Indus. p<
Paper is more detailed	0.32	0.47	0.17	0.38	0.45	0.50	0.01
Paper has richer theory	0.08	0.27	0.07	0.26	0.08	0.28	0.88
Paper is more sophisticated	0.07	0.25	0.07	0.26	0.06	0.24	0.82
Paper has more practical emphasis	0.19	0.39	0.22	0.42	0.16	0.37	0.50
Paper is clearer	0.04	0.21	0.05	0.22	0.04	0.20	0.86

Notes: Mean difference tested using a Wilcoxon-Mann-Whitney test. Of the 90 papers, 49 are of academic origin and 41 are of industrial origin. The 39 sets of paper twins in my data include 10 triplets and 1 quadruplet and those sets tend to have more academic than industry papers.

Table 4. Summary statistics (dyad level)

Variable	All dyads (N=924) Mean	Std. dev.	Industrial origin (N=399) Mean	Std. dev.	Academic origin (N=525) Mean	Std. dev.
Citation	0.68	0.47	0.74	0.44	0.62	0.49
Academic origin	0.57	0.50				
Paper is more detailed	0.30	0.46	0.25	0.43	0.34	0.47
Paper has richer theory	0.06	0.25	0.08	0.28	0.05	0.22
Paper is more sophisticated	0.10	0.30	0.05	0.21	0.14	0.34
Paper has more practical emphasis	0.22	0.42	0.33	0.47	0.15	0.35
Paper is clearer	0.00	0.06	0.01	0.09	0.00	0.00
US paper	0.76	0.43	0.68	0.47	0.81	0.39
Journal impact factor	23.17	12.65	25.32	11.53	21.55	13.22
Publication year	2000.0	2.69	1999.9	2.82	2000.2	2.58
Patent application year	2002.4	2.79	2002.1	2.78	2002.6	2.78
Patent-paper pair	0.61	0.49	0.71	0.45	0.53	0.50
Number of authors (logged)	2.35	0.75	2.55	0.55	2.20	0.84
Authors' joint publication stock (logged)	5.61	0.77	5.79	0.67	5.48	0.81
Authors' joint publication with industry (logged)	3.29	1.54	4.10	1.08	2.67	1.55
Authors' joint patent stock (logged)	3.04	1.63	3.38	1.65	2.79	1.57
Mixed academia-industry team	0.37	0.48	0.34	0.47	0.40	0.49
Prior co-authorship	0.22	0.62	0.27	0.75	0.18	0.50
Time lag	2.42	1.61	2.32	1.63	2.49	1.59
Geographic distance (logged)	6.73	2.21	6.97	2.26	6.55	2.15
Same country	0.64	0.48	0.58	0.49	0.68	0.47

Notes: There are more dyads with papers of academic origin (N=525) than dyads with papers of industrial origin (N=399). The 39 sets of paper twins in my data include 10 triplets and 1 quadruplet and those sets tend to have more academic than industry papers.

TABLE 5. Impact of academic (vs. industrial) origin: Main results

	Dependent variable = citation (1/0)		
	Conditional logit; controls only	Conditional logit; main effect	LPM; main effect
	(5-1)	(5-2)	(5-3)
Academic origin		-0.673***	-0.238***
		(0.25)	(0.07)
Paper is more detailed	0.623*	0.755**	0.184**
	(0.36)	(0.30)	(0.08)
Paper has richer theory	-1.073	-1.096	-0.287
	(0.87)	(0.72)	(0.18)
Paper is more sophisticated	-1.177	-0.99	-0.249
	(0.92)	(0.80)	(0.19)
Paper has more practical emphasis	0.593**	0.467	0.0843
	(0.29)	(0.32)	(0.09)
Paper is clearer	14.55***	14.58***	0.696***
	(1.41)	(1.35)	(0.19)
US paper	1.544**	1.701***	0.408***
	(0.67)	(0.66)	(0.15)
Journal impact factor	0.0148	0.00264	0.00118
	(0.03)	(0.03)	(0.01)
Patent-paper pair	0.204	-0.0079	-0.0478
	(0.51)	(0.45)	(0.10)
Number of authors	0.0745	0.0773	-0.0588
	(0.49)	(0.39)	(0.11)
Authors' publication stock	0.221	0.349	0.124
	(0.34)	(0.32)	(0.08)
Authors' patent stock	-0.0401	-0.117	-0.0367
	(0.12)	(0.17)	(0.04)
Time lag	0.313	0.506	0.0641
	(0.43)	(0.41)	(0.09)
Geographic distance	-0.115	-0.0949	-0.029
	(0.10)	(0.10)	(0.03)
Same country	-0.426	-0.402	-0.124
	(0.53)	(0.55)	(0.14)
Constant			0.145
			(0.36)
Observations	523	523	924
# simultaneous discovery/patent dyads	225	225	480
Pseudo-R2	0.119	0.153	0.149
Log-likelihood	-163.3	-157.1	-310.1
Simultaneous discovery/patent FE	yes	yes	yes

Notes: Observations are paper/patent dyads. Column (3) contains more observations as LPM does not drop cases in which the patent cites all the paper twins. Fixed effects are at the set-of-paper-twins/patent dyad level. Standard errors are clustered throughout at the level of the set of paper twins; *** p<0.01; ** p<.0.05; * p<0.1.

TABLE 6. Impact of academic (vs. industrial) origin: Robustness analysis

	Conditional logit; dependent variable = citation (1/0)					
	Higher-order twins (triplets, quadruplets) excluded (6-1)	5% most-highly-cited discoveries excluded (6-2)	Twins with PubMed related citation rank difference >10 excluded (6-3)	Twins published >6 months apart excluded (6-4)	Large collaborations (>20 authors) excluded (6-5)	Dual industrial and academic affiliations excluded (6-6)
Academic origin	-0.797***	-0.999***	-0.491***	-0.566**	-0.487**	-0.467**
	(0.219)	(0.215)	(0.147)	(0.254)	-0.199	-0.215
Observations	332	387	393	475	423	445
# simultaneous discovery/patent dyads	166	186	186	201	190	187
Pseudo-R2	0.420	0.32	0.267	0.139	0.126	0.096
Log-likelihood	-66.7	-91.6	-100.7	-145.2	-130.3	-143.1
Controls included	Yes	Yes	Yes	Yes	Yes	Yes
Simultaneous discovery/ patent FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: All models include controls for characteristics of the paper, author, institution, and paper/patent dyad. Standard errors are clustered throughout at the level of the set of paper twins; *** p<0.01; ** p<.0.05; * p<0.1.

TABLE 7. Impact of academic (vs. industrial) origin: Mechanism

	Conditional logit; dependent variable = citation (1/0)				
	Dyad excluded if prior tie between authors and inventors (7-1)	Controls for authors' involvment with industry (7-2)	Academic vs. corporate patents (7-3)	Main effect over time (7-4)	High-status vs. low-status academic origin (7-5)
Academic origin	-0.575**	-1.148***			
	(0.274)	(0.397)			
Mixed academia-industry team		0.538			
		(0.63)			
Authors' publication with industry		-0.614*			
		(0.35)			
Academic origin – Academic inventor			-0.633*		
			(0.36)		
Academic origin – Corporate inventor			-0.624**		
			(0.28)		
Academic origin – 0 to 1 years after publication				-0.912**	
				(0.40)	
Academic origin – 2 to 3 years after publication				-0.621**	
				(0.30)	
Academic origin – 4 to 5 years after publication				-0.531*	
				(0.31)	
Academic origin – High-status institution					-0.691*
					(0.36)
Academic origin – Low-status institution					-0.657**
					(0.30)
Observations	404	523	523	523	523
# simultaneous discovery/patent dyads	179	225	225	225	225
Pseudo-R2	0.175	0.172	0.147	0.155	0.153
Log-likelihood	-117.6	-153.4	-158.2	-156.7	-157.1
Controls included	Yes	Yes	Yes	Yes	Yes
Simultaneous discovery/patent FE	Yes	Yes	Yes	Yes	Yes

Notes: All models include controls for characteristics of the paper, author, institution, and paper/patent dyad. Standard errors are clustered throughout at the level of the set of paper twins; *** p<0.01; ** p<.0.05; * p<0.1.