

## **Industry collaboration, scientific sharing and the dissemination of knowledge**

James A. Evans  
The University of Chicago

Corresponding author:

James A. Evans, The University of Chicago, Department of Sociology, 1126 E. 59<sup>th</sup> Street, Chicago, IL 60637, USA.

E-mail: [jevans@uchicago.edu](mailto:jevans@uchicago.edu)

### **Abstract**

Robert Merton famously characterized modern science as distinct from other social spheres by the importance of sharing. In contrast, secrecy is often claimed the most frequent method companies employ to benefit from their discoveries. This study interrogates these claims, and then uses fieldwork on academic research with the popular plant model *Arabidopsis thaliana* and the companies that support it to explore the nature of sharing in academy and industry. Using archival materials and panel models, the study then examines the consequences of industry collaboration, how it influences sharing between academic scientists and the reach of their ideas and materials.

Interviews with academic scientists and industrial research managers reveal differences in sharing. Academics are practiced at communicating discoveries and sharing materials, but occasionally withhold to secure credit or barter to maximize it. In contrast, companies manage their ideas and resources for longer-term control. The difference is not that academic scientists never keep secrets, but that many do so badly. Statistical findings suggest that industry sponsorship influences scientists to reduce their sharing of research materials and methods, but it increases the reception of scientists' early-stage manuscripts, likely as a substitute, enabling competing labs to infer a closed

lab's methods. Industry's influence also affects sharing indirectly by sponsoring research in less crowded areas. In this way, industry curbs the demand as well as the supply of sharing in science. As a result, industry sponsorship limits the social, organizational and geographic distance that sponsored ideas travel over time. Scientists find it difficult to enroll other academics in their research findings as they become enrolled in an industry project of avoiding and mitigating competing science.

### **Keywords**

sociology of science, industry collaboration, science policy, sharing, secrecy

In Robert Merton's famous characterization of science as a social system (1942), the free flow of scientific ideas and findings was central. Among the four institutional imperatives he claimed govern scientific activity, three directly mandate the elimination of restrictions to knowledge given and received: all discovered knowledge (the norm of disinterestedness) should be shared (the norm of 'communism') freely and without distinction (the norm of universalism). In Merton's system, the ultimate reward in science—credit—is obtained through recognition of priority (1957). Priority functions like an invisible hand: Scientists share knowledge universally by publishing in their self-interest (see also Dasgupta and David 1994 ; Polanyi 1962).

This vision of open science contrasts former eras in which 'nature's secrets' were not made public (David 2003). What distinguished Dark Age alchemy from Enlightenment chemistry, most notable among those like Robert Boyle and Isaac Newton who engaged in both, was neither experiment nor theory, but that results from the one

were secreted and the other published.<sup>i</sup> Variations of the word ‘secret’ riddle the mostly hand-written, closely circulated alchemical treatises: authors considered their knowledge too valuable, too powerful,<sup>ii</sup> too sacred, or, in the case of Boyle, too potentially embarrassing to be entrusted universally.<sup>iii</sup>

Today ethnographers, sociologists and historians dispute the image of modern science as fundamentally open (Shapin 2008). Just as Robert Boyle accused rivals of stealing his experimental notebooks (Shapin 1994), contemporary laboratories and research groups self-consciously secure and secret their results (Hackett 2005 ; Knorr-Cetina 1999 ; Mitroff 1974 ; Mulkay 1976 ; Mulkay and Edge 1973).<sup>iv</sup> Solutions to nature’s largest riddles promise the greatest glory and scientists compete for the same central answers. To maximize the likelihood they will publish first, scientists seek out others’ intermediate findings while concealing their own. Competition creates paranoia over being ‘scooped’ or beaten to publication by others, sometimes with one’s own disclosed data (Marshall 2002 ; Watson 1968). As a result, scientists dole out ideas and results selectively, even when they publish completed findings quickly.

Competition shapes communication and sharing in the world of industrial science, but in a different way and for a different reason. In business, the nature of ideas conspires against their marketing because to advertise an idea is to give it away (Arrow 1962). As a result, companies compete primarily over the sale of products and services, where discoveries represent only one step in development. Companies frequently conceal discoveries in order to develop innovative products and minimize competition. In a study of how manufacturing companies protect their ‘intellectual assets’, Cohen et al. (2000) found that companies safeguard innovations primarily through secrecy,

complementary manufacturing and lead time.<sup>v</sup> Complementary manufacturing and lead time are not alternatives to secrecy, but ways to extend and exploit it by keeping innovations in-house. Patenting, the only appropriation that requires publication of discoveries, is least used. Even in life science companies, where blockbuster discoveries are critical and patenting prominent, patenting trails direct and indirect secrecy in practice.

In sum, academic secrecy is characterized by strategic information exchange oriented toward the accrual of scientific credit through publication (Latour and Woolgar 1979). Industrial secrecy supports the ongoing profitability of products and has no clear expiration. Industry innovations often remain opaque in the product or service they enhance. Industry outputs must often be dismantled to recover or reverse engineer the central innovation.

Academic research has shifted in relation to the proprietary impulse of business and the military over the past two centuries. Technological advance in the second industrial revolution drew predominantly on much older science and occurred distant from most academic research (Rosenberg 1982). This changed in the first decades of the 20<sup>th</sup> Century, and especially following the successful U.S. development of antibiotics and nuclear weapons during WWII, which demonstrated the power of science while simultaneously bringing it under increased government control. Vannevar Bush, MIT's vice president and a prominent electrical engineer, in his 1945 report to the President, *Science, the Endless Frontier*, argued that for science to be most useful in the long-run, it must be insulated from the exigencies of immediate application (Calvert 2006).<sup>vi</sup> Following this direction, the U.S. government continued to fund science at wartime

levels, not only through military agencies (Sapolsky 1990) but in new institutions like the National Science Foundation (NSF) and National Institutes of Health (NIH) that allocated funding by peer review. These bodies were imitated around the world and contributed to the unprecedented growth and autonomy of academic science. Academic research in this period was never independent of military (Kaiser 2002; Kevles 1978; Kleinman and Solovey 1995; Leslie 1993; Masco 2006; Sapolsky 1990) or industry motivation (Grandin, Wormbs and Widmalm 2004; Kleinman 1998), but ongoing contact between many areas of research and either military or firms was intermittent.

Recent interactions between firms and universities have brought academic science closer to business. In the last three decades, discoveries in the academic life sciences have proven immediately relevant to problems in medicine and agriculture, spawning a biotechnology sector in which scientists at firms and universities converge around similar questions using many of the same scientific materials, techniques, and analyses. At the same time, intellectual property (IP) protection has broadened to allow the patenting of many basic discoveries, including genes, altered organisms and scientific research tools (Heller and Eisenberg 1998). Legislators championing the market-relevance of research have directly enabled university commercialization (Berman 2008), most notably through the 1980 Bayh-Dole Act in the U.S. and international copies of this legislation, which allowed universities to patent the results of government funding. This has lured academics to found life science firms and enticed established companies to maintain ongoing engagement with scientists (Powell 1996) and departments (Merrill and Cooper 1999 ; Mowery 1999).

The rise in relationships with industry has raised concern over the fate of

openness in public science (Nottenburg, et al. 2001, pp. 45-46; Press and Washburn 2000 ; Washburn 2005). These reservations, however, often caricature academic openness to accentuate the magnitude of conflict and call for action before understanding. This article represents an empirical exploration into the influence of industry collaboration on scientific sharing. Rather than simply test whether academics share more or less when in collaboration with industry, I explore the changed nature of how they share, and examine the range of possible consequences. I refuse to sensationalize these consequences, but neither do I understate them. If collaborating scientists hold back research materials and uncensored findings, others will more likely duplicate their work and fragmented discoveries will accumulate into a more fragile public knowledge. Scholars have speculated about these possibilities, but few use more than anecdotes for evidence (see Murray and Stern 2005 for a notable exception).

### **Sharing in science**

In the course of a scientific program, there are several moments at which scientists can share or withhold information or materials from others. At a workshop, seminar or conference, academics can speak or remain silent about research ideas—questions, study designs or hypotheses (Cambrosio, et al. 2004). Once experiments are prepared, scientists can impart or retain their research methods and materials. This involves the choice to grant lab visits, share algorithms, databases, DNA vectors, or cell-lines with others for whom they could prove useful (Hine 2006 ; McCain 1991). After early theoretical or experimental findings are obtained, researchers may present them or circulate them in manuscripts (Faberge 1982 ; Remington 1988). As research is refined,

scientists decide whether or not to publish, and whether to publish all of their results or a censored sample (Relman 1980 ; Relman 1984).

Recent research has fixed primarily on the final two stages—on whether and what to publish. This work comes predominantly from the domain of clinical medicine, where the commercial stakes are high and where sharing is understood to have ethical as well as efficiency implications: lives are at stake if a treatment's limitations remain hidden. Clinicians who feel ideologically committed to the openness of research have performed much of this research. Articles on industry-funded biomedical research reveal a lower publication rate for industry-funded studies (Dickersin, et al. 1992 ; Easterbrook, et al. 1991 ; Massie and Rothenberg 1993 ; but see Stern and Simes 1997 for a notable exception). Others demonstrate a positive correlation between company-funding and positive findings about company therapies (Davidson 1986 ; Djulbegovic, et al. 1999 ; Mandelkern 1999 ; Stelfox, et al. 1998 ; Vandembroucke, et al. 2000 ; Wahlbeck and Adams 1999 ; Yaphe, et al. 2001 ; Bekelman, et al. 2003 ; Sismondo 2008). Studies have not, however, demonstrated any inferiority in industry-sponsored methodology (Clifford, et al. 2003 ; Davidson 1986 ; see Jadad, et al. 2000 for an exception; Kjaergard, et al. 1999 ; Neumann, et al. 2000 ; Rochon, et al. 1994) and several studies find them technically more sophisticated (Cho and Bero 1996 ; Djulbegovic, et al. 1999 ; Leibeskind 2000 ; Mandelkern 1999 ; Massie and Rothenberg 1993). From these relationships, authors and meta-analysts (e.g., Lexchin, et al. 2003) infer that scientists censor findings harmful to a sponsoring company's products. Much less research delves more deeply into the informal sharing that takes place between scientists. The sharing of methods, materials and early manuscripts constitutes a greater portion of the diplomatic

and constructive communication linking scientists. Analyses of these deeper forms of information-sharing probe the degree to which industry collaboration influences the ongoing social life of science.

A few studies provide hints about the influence of industry collaboration on sharing. A 1995 survey conducted by Blumenthal et al. (1996) found that biological scientists reporting industry support classified their work as constituting ‘trade secrets’ a third more often than those reporting no such support. A study of industry funding in oncology research found that methods sections of industry-sponsored papers contained less detail (Knox, et al. 2000). The methods seemed no less valid, but there was more ambiguity about the exact procedures that had been followed. Finally, a study of articles published in *NatureBiotechnology*, matched with patents granted for the same research, found that once patents were issued, the companion papers were cited less frequently (Murray and Stern 2005). This effect was greatest for articles with all public-sector authors, smaller for those with some industry authors, and nonexistent for purely private-sector teams. This suggests scientists assume that materials and techniques associated with private-sector publications come with restrictive intellectual property (IP) rights and do not pursue them. IP-oriented or industry-affiliated researchers may *perform* ambiguity and an unwillingness to share, but this unwillingness may also be *imputed to them* as a result of their participation in private science.

For these reasons, I anticipate that materials and methods will less frequently flow from academic researchers with industry ties. A lack of sharing at this level, however, is unlikely to result in disinterest from competing labs if those labs perceive industry-sponsored research to be scientifically important. If competitors cannot directly access

methods and materials from what they believe to be a closed lab, they might seek indirect access through scrutiny of the lab's manuscripts—those associated with a research talk or poster presentation. Consequently, I predict shared manuscripts will *substitute* for shared methods: as labs collaborating with industry share fewer materials and methods, their shared manuscripts may be more requested and used.

This pattern has clear implications for the dissemination of research. If industry-sponsored materials and methods are shared more selectively, others will less likely pursue their programs and expand on their experiments. Due to increased ambiguity, experimental duplication not only might be attempted less often, but when attempted it might less likely produce similar results.<sup>vii</sup> If so, ideas from industry-sponsored articles will less frequently spread beyond the scientists, research organizations, regions and scientific areas that spawned them, and their resulting networks of scientific influence will be smaller (Latour 1987). This is consistent with research on patenting and the dissemination of invention. Economists have found that patents initially are cited most within their geographic region (Jaffe, et al. 1993) and sponsoring firm (Adams and Jaffe 1996), although this concentration lessens over time. Sorensen and Fleming (2004) demonstrate how patents that do not cite academic literature—presumably produced by non-academics—are cited less rapidly than those that do, and the inventors who cite them are less geographically dispersed. In sum, I predict that industry collaboration and sponsorship will intentionally slow the social, geographic and scientific spread of ideas.

I explore the influence of industry ties on academic sharing, and evaluate these hypotheses about the implications of that influence in an expanding corner of molecular biology: research performed with the model system, *Arabidopsis thaliana*.

### **Evidence from *Arabidopsis thaliana***

*Arabidopsis thaliana* is the fruit fly of the plant world. The wild, purple-flowered relative of mustard was the first higher plant to have its genome sequenced and it has become the most popular genetic model organism in plant biology and agricultural biotechnology (Walbot 2000). Other science studies have used model organism communities as sites for research, including the fly (Kohler 1994), mouse (Rader 2004), and worm communities (Ankeny 2007).

Most major universities, many research institutes and government agencies, and all major plant biotech companies perform or fund research on *Arabidopsis*. Its short, nonredundant genome and rapid reproductive cycle make it easy to link DNA sequences to plant functions—to perform both classical and molecular genetics. Fundamental discoveries in *Arabidopsis* have fed back into basic science, revealing novel biological mechanisms like non-Mendelian inheritance (Lolle, et al. 2005), along with previously unknown similarities between plant and animal function. Although most early *Arabidopsis* investigators came from outside plant biology—from physics, chemistry and molecular biology labs studying *E. coli*, yeast, or *Drosophila*—this research is increasingly being transferred into crop plants with social and profitable implications, such as drought resistance corn, and the plant-manufacture of pharmaceuticals and industrial substances including oils and plastics.

In order to investigate industry influence on the academic *Arabidopsis* community, I began with historical and qualitative investigations that guided subsequent quantitative inquiry. I initially gathered hundreds of articles and unpublished papers using *Arabidopsis*, published histories of *Arabidopsis* research and the emergence of

plant biotechnology, government grant data in the U.S., Great Britain, Japan, the Netherlands and several other countries, and annual reports from the North American and International Steering Committees of the *Arabidopsis* community. Between 2003 and 2005, I interviewed *Arabidopsis* scientists (15), executives at technology companies (14), and government sponsors of *Arabidopsis* research (7). These led to participant observation in *Arabidopsis* labs and at the 14<sup>th</sup> and 16<sup>th</sup> International Conference on Arabidopsis Research, where I conducted dozens of additional informal interviews, some of them surrounding a poster-presentation of early findings from this research.

*Arabidopsis* research presents a conservative case through which to examine the effect of business collaboration on secrecy in academic science. An examination of the *Arabidopsis* community will likely *underestimate* the amount of strategic privacy in science for three primary reasons. First, because of its relatively small size compared with other model organism communities (e.g., *Drosophila*—fly, *Saccharomyces*—yeast, mouse, rat), *Arabidopsis* scientists attend conferences together: they know and talk to one another. As a result, they are keenly aware of reputation throughout the community and may sanction one another for inappropriate behavior.

Second, nearly all of the scientists who brought *Arabidopsis* into the molecular era remain central as statesmen in the field. For a generation, Chris Somerville, Martin Korneef, and Elliot Meyerowitz have actively promoted community building and scientific sharing through a range of institutions, from the *Arabidopsis* Information Service (AIS) newsletter to an annual *Arabidopsis* conference to North American and international research steering committees, the formation of biological stock centers, and most recently, The *Arabidopsis* Information Resource (TAIR), an NSF-funded

informatics website that hosts databases of *Arabidopsis* papers and information (Meyerowitz 2001 ; Somerville and Koornneef 2002 ; Somerville 2001). As consequence, postdocs and scientists who came to *Arabidopsis* from larger or more competitive communities, such as *Saccharomyces* and *Drosophila*, often make note of *Arabidopsis*' communal atmosphere and the relative openness of its scientists.<sup>viii</sup>

Finally, Monsanto, Novartis, DuPont, Dow and other companies in the agricultural, lumber, chemical, food, and pharmaceutical industries became interested in plant biotechnology in the early 1980s, at about the same time that *Arabidopsis* began to emerge as a research tool. A number of small, dedicated plant biotechnology companies were founded soon afterwards. Because of early commercial collaborations with academics, this setting provides an opportunity to examine the consequences of commercial influence over time rather than forcing focus on the evolution of that influence, and the changing legitimacy of industry ties. Despite industry interest, however, *Arabidopsis* biology remains only a model for economically and socially meaningful organisms. One step removed from application, findings established in *Arabidopsis* remain further from the marketplace than findings from crop field trials or their medical equivalent, human clinical trials.

In clinical trials, the commercial returns are much more assured—approximately 70% are funded by industry (Mello, et al. 2005). Clinical medicine is also where most research on industry influence has been performed. It is no surprise that such influence has been identified there. Clinicians have performed most of this research, however, and because they largely assume widespread industry influence on medical research, results have been loosely analyzed and interpreted as confirmation of a caricatured conflict

between openness and secrecy. *Arabidopsis* research provides a setting where corporate dominance is not assumed. This enables me to explore its implications for the quality of sharing and the spread of knowledge in academic science. Moreover, if corporate influence on sharing in the *Arabidopsis* community can be identified, we can be confident it is amplified elsewhere.

Despite its relatively small size, paternalistic leadership, and consistent but minor industry influence, the *Arabidopsis* community exhibits the same type of competitive pressure found in other research settings. An advanced graduate student from one plant pathology lab described strategic secrecy and the narrow conditions under which it was safe to share:

There is a lab in New York where [a student in our lab] has sent his mutants and they sent him theirs for different assays. They are doing different things, and the technologies are different so [its okay]. My advisor had this postdoc ... and they were collaborating with a group in England. It turns out that group eventually scooped my advisor. Not a nice story. So, there is a risk in sharing. Scoop means that somebody publishes your research before you do .... You are collaborating with a lab, and the other lab kind of either twists the truth ... or like stops sharing their part and just accepts the sharing that you are doing and then takes advantage of your information in order to publish theirs faster ... which happened to the postdoc in our lab. They were trying to clone the same gene, and they convinced our lab that they had cloned it already so the postdoc that was working on it here stopped his work—which had been a substantial amount of work—on that gene. It turns out they hadn't, and it took them another three years to actually publish. This postdoc could have finished his original project very easily. [BS05]<sup>ix</sup>

Sometimes researchers would share incorrect or incomplete information to scoop others or avoid being scooped. Most scientists, post-docs and students, however, reported receiving information and materials following a request, ‘though sometimes not as soon as you would like’ [BS04]. Most often this sharing was untethered, a form of generalized reciprocity (Sahlins 1972) associated with an ideology of openness, which cultivated the

expectation that borrowing scientists would share with others in the future. Sometimes, however, when a resource of particular value was created, the creator might engage in a more direct exchange. One student described:

One thing that got generated a while ago were these—a reverse genetic tool—it was a population of tDNA mutants...The guy that generated them, I think for a while was requiring to be a coauthor on any paper where people used them to clone their genes. [BS04]

In this case authorship was exchanged for materials, and in other situations materials were bartered.

While sharing was occasionally selective between academic labs, it tended to be much more restricted between industry ones. Agricultural biotechnology firms could select from a broad menu of legal protection to secure their knowledge, including patents, plant variety protection (PVP), plant patents (for asexually reproducing plants) and copyrights (for software or databases) (Buttel and Belsky 1987). As in other industries, the most effective protection is the trade secret (Cohen, et al. 2000). Company executives frequently noted the secrecy they demanded of their in-house scientific staff. One plant biotech president and founder with substantial academic credentials and sympathy described:

In my company, if there was anything of any value, I wouldn't let anyone talk about it—maybe a few years after we are making products. If there is anything without commercial value, then we let people publish it. Often big companies do that. But if it has value, it's not discussed outside. [BI02]

The rewards and barriers to sharing were higher between competing company labs and so, rather than scoop, they were occasionally known to engage in larceny. In 1983, scientists from Germany's Institute for Biological Crop Protection in Darmstadt, funded by the company Boehringer Mannheim, filed patents for a strain of the bacteria *Bacillus*

*thuringiensis* (*Bt*) and soon after published their research on it. *Bt* produces a protein that kills caterpillars when they ingest it. This new strain also killed other insect pests including many types of beetle. Agriculture firms wanted to clone the protein-producing gene and put it directly into plants, making them poisonous to their enemies. The discovering lab received about seventy-five requests to share the strain, but rejected them all. After the San Diego plant biotech firm Mycogen unsuccessfully sought to license this *Bt* from Boehringer, Corinna Herrstadt, a young German scientist at Mycogen phoned Alysius Krieg, one of the strain's discoverers, asked for a lab tour and was received warmly. According to rumor at Mycogen [BM14], during the tour Krieg left the room momentarily and Herrstadt 'slid her hand along the surface of a table in Krieg's laboratory, then wiped her fingers on the Petri dish hidden in her purse .... [Then] she brought that dish back to San Diego and there "discovered" *Bt san diego*.' Mycogen eventually settled for \$4 million and forfeited its U.S. patent on the 'novel' strain (Charles 2001).

Because of this industry climate, academic plant biologists could be imperfect partners for industry when they shared work too freely. The former director of a prominent British plant biology research institute, who had recently shifted to the private sector, described problems with these partnerships.

The frequency with which people want to put information into the public domain is just that much greater in the public sector, and you have to manage that—with the patent laws also being somewhat different on the two sides of the Atlantic. I mean, you can't disclose anything in the UK if you want to gain a patent. There's no year's grace or anything like that and so that's pretty tricky to manage. It takes a lot of sophistication on behalf of the institution and the PI to manage all of this .... Some PI's do it successfully, fairly readily. Others just don't get it and they don't do very well. [BM26]

Maintaining relationships with firms demanded discretion from public scientists that this entrepreneurial director had to coach and manage. Other corporate research managers noted with resignation that academic partners were less predictable than corporate ones—‘University students and post-docs are not in a company—they’re influenced by the campus environment, you know, humans next each other in the lab, talking to each other. What can you do? .... On secrets about new inventions coming from the lab, its much more professional in companies than universities’ [BM21].

Some *Arabidopsis* scientists spoke of receiving direct pressure from industry to safeguard materials and findings. One senior scientist detailed a skirmish with a major chemical company early in his career, before he had learned industry discretion, over sharing research materials.

I was a consultant for ChemCo. ... I had a very warm relationship with them ... and one day the head of research gave me a gene I could use as a probe. This gene was for an enzyme that they thought was involved in metabolizing their core herbicide. And I went back to the lab and isolated a mutant resistant to the herbicide and used the gene and cloned it. They had written patents covering 10 million herbicides and there were a couple of tiny little areas of chemistry that PetroCo and PharmCo patented that the ChemCo patents didn’t reach. This gene that I cloned conferred resistance to those herbicides. What this meant was that even though PetroCo probably didn’t have enough chemistry to find really useful compounds for every crop, they could take this gene and engineer a plant to be resistant to their tiny little area of chemistry .... so this made [ChemCo] extremely nervous. I was also rather idealistic so I just handed this gene out far and wide: No paperwork, no strings attached, no anything. I just sent it to every academic lab that asked for it, and this really enraged them and so ... I got into a legal rangle of the first magnitude .... This moved up to a level where it was alarming people: the president of the university went to another big chemical company, and they came and offered me a million dollars in legal aid the next day, and 500 thousand for a nonexclusive right to my patent rights for this thing. I could see what was going to happen. The university president had said, ‘They’ll have to file here, and no court in this state is going to find against a state professor who’s facing off a multi-national giant who wants hegemony’, but what I could see was with a million dollars in legal aid—and they promised more if I needed it—I’d spend the rest of my life in a court room with those guys .... so I phoned my

friends at ChemCo and I said, ‘I have a million dollars in legal aid. This is going to get extremely messy. Find a solution to this.’ And so suddenly they went from threatening us to being very agreeable. They bought my post-doc—they didn’t want him to go out and give talks about this work. He wanted to get a job. In the end, they basically gave him a down payment on a house and made some cash payments and made an agreement with the university and everybody was happy. [BIO1]

After payments were made, the scientist stopped sharing the clone. He worked with companies again in the future, eventually founding a successful start-up while retaining a prominent open lab, but he learned professionalism through his experience. He learned that some things in some contexts with some partners should not be shared.

Once industry partners developed a research program into a product, such as a tool to insert foreign DNA into plants or a crop with genetically enhanced agronomics, they would obtain legal rights. They would then barter these with other companies through cross-licenses to obtain enough rights to market a product. Instead of exchanging resources for credit, as academics occasionally did, they exchanged rights to commercialize an innovation.

Interviews with *Arabidopsis* researchers and administrators at collaborating companies highlight differences in the nature of restricted sharing in academic and commercial settings. Secrecy, duplicity, and delayed and restricted sharing, when they occurred between academics, were deployed to procure and manage scientific credit. Industry withholdings, on the other hand, were performed to secure long-term control and profit from an invention. In this sense, it is not that academic scientists never kept secrets, but that many did so badly: they could be unpredictable, would routinely communicate with others across institutional boundaries, and despite nondisclosures, let things slip. The interviews also intimate that, as a result of these differences, industry

collaborations influenced academic sharing. Academics with industry partnerships became more discrete, selective and ‘professional’ in their management of ideas and resources. In an effort to understand the *average* effect of these partnerships on academic scientists and the dissemination of published science, I engage in a more systematic analysis of industry linkages and sharing among *Arabidopsis* scientists.

### **Quantitative analysis of publication data**

To evaluate sharing behavior within the *Arabidopsis* community, I collected abstracts and bylines from 18,359 published articles utilizing *Arabidopsis thaliana* between 1907 and 2002 (see methodological appendix for detail). I then linked these data with ThompsonReuter’s Science Citation Index (SCI) and assembled citation data on the 11,000 *Arabidopsis* articles present in both databases. Between 1975 and 2002, the number of papers grew exponentially, from 91 and 112 in 1980 and 1985, respectively, to 373 in 1990, 1070 in 1995 and 2192 in 2000. I also scanned and parsed the acknowledgement sections of *Arabidopsis* articles to identify mentions of scientists who shared advice and materials, and government agencies and companies that provided support.

*Arabidopsis* articles address a variety of fields: ThompsonReuters classifies them as speaking to the concerns of plant science, ecology and agriculture, but also biophysics, physiology, biochemistry, virology, cell biology and others. I extracted an extensive list of scientific terms from the titles and abstracts of *Arabidopsis* articles.

I then associated these articles with the scientists who produced them and the academic research organizations where they worked. With the acknowledgement and

citation data, I developed measures to capture the degree to which scientists shared materials and manuscripts for each year in which they published. I then linked these scientists to the companies with whom they collaborated. I use regression models to test my hunches about how industry collaboration in prior years affects sharing in the present one. Then I consider the spread of ideas through *Arabidopsis* papers by examining the degree to which article sponsorship influences the dissemination of citation attention beyond the scientists, organizations, regions and scientific areas in which they were created. I subsequently describe each of these steps regarding data organization, coding and analysis. (See the methodological appendix for technical details.)

#### *Principal investigators and contexts*

I examine those scientists who ran their own labs at any time within the period of study. Government grants title such scientists ‘principal investigators’ (PIs), for their role in managing proposed research programs. Between 1974 and 2003, 5,725 PIs, a third of whom resided in the U.S., used *Arabidopsis* in their research according to bylines in publications and conference directories. These scientists were associated with 3,163 distinct universities, hospitals, nonprofit and government research institutes.

#### *Materials sharing*

I measured materials sharing by counting the number of independent *Arabidopsis* articles in which PIs are mentioned as having shared materials, data, or technical assistance and on which they are not themselves listed as authors. These attributions are listed in acknowledgements sections, and uncovered 36,512 such mentions—9,021 to PIs who

were not authors of the article.

For example, an article drawn at random from the sample, published in *Nucleic Acids Research* (1997) by research groups from the Universities of Amsterdam and Utrecht, states ‘[a]nalysis of the chromatin domain organisation around the plastocyanin gene reveals an MAR-specific sequence element in *Arabidopsis thaliana*’. Matrix attachment regions or MARs are DNA sequences where the nuclear matrix—a network of fibers in the cell nucleus—attaches and appears to influence gene expression. The article’s acknowledgement notes, ‘The authors would like to thank ... Drs J. Mulligan and R. Davis (Stanford University) for their kind gift of the lambdaGEM-11 *A.thaliana* genomic library; Drs D. Bouchez and C. Camilleri (CEPH/INRA/CNRS) for their chromosome mapping data....’

Stanford biochemist Ron Davis directs the Stanford Genome Technology Center where John Mulligan, now chairman of Blue Heron Biotechnology, was a postdoc in the late 1990’s. David Bouchez (PI) heads the group ‘Organisation & dynamics of the cytoskeleton in *Arabidopsis thaliana*’ at the Genetics and Plant Breeding Laboratory within INRA Versailles where Christine Camilleri works as an engineer. Willingness to share materials, then, is calculated as the number of times in the subsequent five years that a PI is mentioned in acknowledgements. In this example Davis and Bouchez, *Arabidopsis* PIs, would each receive one sharing point for the previous five years.

### *Pre-publication sharing*

I measured manuscript sharing by proxy as the degree to which scientists’ prepublication research is utilized and referenced in others’ papers. Unlike physics, economics and

other fields with central preprint archives, manuscripts in molecular biology are held closely and shared only selectively prior to publication. Biologists will sometimes circulate papers to trusted colleagues for publishing advice, or by requests ensuing from a presentation, conference or workshop. As a result, I assess preprint sharing using traces of contemporaneous citations.

Because articles cite other articles from the past, and not the present or future, the graph defined by citations between scientific articles should represent a ‘tree’, or an acyclic graph where no article  $x_1$  both cites and is cited by another article  $x_2$ . In reality, however, scientific corpora contain many cycles between articles. In my set of *Arabidopsis* articles and those that cite them, cycles range from length 2 ( $x_1$  cites  $x_2$  and  $x_2$  cites  $x_1$ ) to length 8 ( $x_1$  cites  $x_2$ ,  $x_2$  cites  $x_3$ , ...  $x_7$  cites  $x_1$ ). Figure 1 graphically illustrates how articles A through G are implied in a chain of citation cycles. In the illustration, articles E, F and G are each cited in multiple cycles. Taking the union of those cycles, E is cited by one unique article (F), F is cited by two unique articles (E, G), and G is cited by two unique articles (E, F).

The existence of these cycles represents the degree to which authors share—and other scientist’s request and use—pre-publication manuscripts and discoveries. Through the publication process, citations to pre-published manuscripts are typically revised to reference forthcoming articles as both cited and citing articles approach publication. Following this logic, I calculate manuscript sharing by taking the number of distinct publications that cite each *Arabidopsis* article in a citation-cycle.<sup>x</sup>

*Dissemination*

In order to establish the way industry ties shape the dissemination of science beyond a publication's own authors and institutions, I associated each citing and cited paper with the authors and organizations that produced it. Next, for every cited-citing paper combination, it was established whether *any* of the authors or institutions associated with the cited *Arabidopsis* paper was also an author or institution associated with the article that cited it. Each cited-citing pair received an author-distance or organization-distance score of '0' if any of the authors or organizations on the citing paper had also coauthored the cited paper and a score of '1' if the two papers shared no authors or organizations.

I calculated a lag variable to indicate years since publication of each *Arabidopsis* article. For each lagged year, the author-distance and organization-distance were averaged across the collection of papers that cite the focal paper in that year. These averaged author- and organization-distances represent the proportion of an *Arabidopsis* paper's citations produced beyond the original article's authors and host institutions. For example, if Martin Koornneef and Chris Somerville, at Wageningen University and Michigan State University, respectively, coauthored a paper in 1994, I calculated lags for that paper in 1994 (0 years), 1995 (1 year), 1996 (2 years) and so on up to 2002. If their paper was cited twice in 1994, once in a paper by Koornneef from Wageningen and Dirk Inze from the State University of Ghent, and a second time by Fred Ausubel and one of his postdocs at Massachusetts General Hospital, then it would receive an author-distance and an organization-distance score of .5 in lag year 0—the year it was authored—because half of its citing articles went beyond the focal article's set of authors and organizations.

### *Geographic spread*

I measured the geographic distance *Arabidopsis* science covered by establishing the

geographic coordinates for each institution associated with an author on an *Arabidopsis* article in the SciSearch database. Each publication was assigned the set of longitude and latitude coordinates associated with each of its host institutions. Then, for every cited-citing paper combination, the distance between the two papers was established by taking the minimum distance between the set of institutions producing the cited *Arabidopsis* article and the set producing the citing articles. These distances between each cited and all citing articles were averaged in every year following publication.

Consider the 1994 *Arabidopsis* article described above, with authors at Wageningen University (latitude 51.9667; longitude 5.6667) and Michigan State University (latitude 42.7283; longitude 84.4488). The minimum distance between that paper's institutions and the citing paper authored by scientists from Wageningen and Ghent is 0, because the distance from Wageningen to Wageningen, the closest institutions in the set intersection of cited and citing institutions, is 0. The minimum distance between the focal paper's institutions and the citing paper authored by researchers at Mass General (latitude 42.3621; longitude 71.0236) is 684.67 miles, because the distance from Michigan State to Mass General, the closest two institutions in the set intersection of cited and citing institutions, is 684.67. Thus, the average minimum distance that the 1994 Koornneef and Somerville article has spread in year 0 is 342.34 miles, or the average of 0 and 684.67.

### *Scientific dissemination*

I assessed the scientific dissemination of published *Arabidopsis* science by measuring the degree to which citing articles went beyond the scientific issues explored in the focal *Arabidopsis* article. This involved calculating the proportion of citing articles that did not

share any of the scientific terms coded in the focal *Arabidopsis* article. This calculation was performed for each year following publication in the same way as author- and organization-distances described earlier. The resulting term-match scores represent the proportion of citing articles within a given lag period that traveled beyond the scientific terms of the focal *Arabidopsis* article.

### *Industry and government ties*

I documented industrial collaboration and funding in several ways. I *inferred* the association between firms and PI's from articles co-authored and patents co-invented by scientists at universities and firms. I also identified any company collaborations or funding mentioned in article acknowledgements.

Studies of commercialization in the life sciences suggest a distinction between partnerships that involve collaborative research relationships versus those that resemble market transactions (Powell 1990). It is not obvious whether these types of ties have distinct effects on sharing, but acknowledgements clearly distinguish them, allowing me to explore possible differences. I coded coauthorship, collaboration and the sharing of scientific materials with industry as *relational ties*, reflecting deep, ongoing affiliations. Alternately, I coded funding, licensing and other payments as *transactional*, suggesting discrete, proprietary exchanges.

With the complete list of co-authoring, inventing and acknowledged companies described earlier, I searched the news databases of Lexis-Nexis and related electronic archives for instances of company names co-occurring in the same article with 'institute' or 'university'. I then coded collaborative and funding relationships linking these

companies with the universities, departments and researchers in my sample.

Relationships suggesting materials sharing or joint efforts in research were coded as relational, while funding, licensing and marketing were treated as transactional. This ability to distinguish between industry ties to scientists and those to universities allows me to identify the differing influence of more and less immediate ties to a scientist's program. In total, the data comprise 3054 relational ties and 736 transactional partnerships between companies and principal investigators, and with 1584 relational partnerships and 693 transactional ties between companies and academic research organizations.

For my analysis of the dissemination of article ideas, industry ties were assessed directly within the article and also across all articles published by a PI, allowing funding indicated in one article to be associated with all articles the PI researched concurrently. The first approach assumes that scientists can build 'Chinese Walls' between separate projects within their labs; the second does not. Despite some claims by interviewed scientists that they were able to separate their projects, the same post-docs, students and technicians in the lab often worked across them. Moreover, resources garnered for one project were often utilized in others. As such, I anticipated the second approach to better capture the reality of industry's influence on the dissemination of academic science. Results proved this to be the case. Only when industry funds and collaborations were imputed to all of a lab's contemporaneous articles did they demonstrate a significant influence on the spread of attention and citations. The tables present only these results.

### *Controls*

Many qualities of *Arabidopsis* scientists, their universities, and their scientific subfields likely alter the degree to which labs are open or closed, independent of industry ties. By including these features in the models, I attempt to isolate the influence of industry connection from other, confounding influences. These include the tenure with which a scientist has researched *Arabidopsis* (and its square); a scientist's social status within the *Arabidopsis* community; and the degree to which a scientist brokered relationships between others. I included indicators of the type of host department (e.g., agriculture, plant biology, medicine ... sociology), research institution (e.g., science and technology university, agricultural university, university with a medical school—general universities without medical or agricultural influences are the reference), and country hosting each scientist. In addition, models control for the diversity of topics studied using *Arabidopsis* and the diversity of departments studying *Arabidopsis* within the scientist's institution within each year of study, as well as whether that institution facilitated industry engagement by sponsoring an office to facilitate technology transfer.

Most importantly, all models control for features of the scientific subfield in which scientists published. My measures of materials and manuscript sharing depend both on a scientist being willing to share and a scientific public wanting to receive. In this sense, a scientist could be influenced to share less by either being unwilling to *supply* material when requested, or shifting to a less populated area of research where others are less likely to *demand* those materials. I control for this latter possibility by assessing the position of a scientist's research with respect to other work in the broader universe of *Arabidopsis* research. In academic science, researchers typically define themselves and organize their activities not by university but by scientific subfield. Subfields vary in

size—in the number of other researchers that could potentially ask for assistance. Subfields also trace different scientific practices, sources of funding, and research audiences. Consequently, I control for scientific subfield and some of its crucial attributes. I identified subfields within *Arabidopsis* science by clustering *Arabidopsis* articles into seventeen subfields based on the distribution of scientific terms across them.

The analysis controls for the proportion of a scientist's research performed in a commercially relevant subfield, as indicated by citations from patents to subfield articles.<sup>xi</sup> Subfields with higher-than-median patent citations per article were deemed more commercially relevant. Less commercial fields like 'photosynthesis' explore biologically significant and distinctive aspects of plants; while more commercial fields like 'terpenes' examine a class of biologically inessential protein substances that are used to produce economically important commodities like turpentine and rubber.<sup>xii</sup> Finally, all models control for the research citations accruing to all articles within the scientist's subfields, both as a flow—the flurry of citations received in the last year—and as a stock—the total citations received prior to the last year. The citation flow traces current excitement about research in the subfield, while citation stock indicates the historical importance of that research.

### *Modeling strategy*

Instances in which my measures identify materials and manuscript sharing are relatively infrequent with a few authors receiving several and most receiving none. Such rare events are mapped nicely by the Poisson distribution, and so I use Poisson panel regression models to evaluate whether scientists' industry ties are associated with their

subsequent sharing behavior. As a conservative test, I compare researchers in periods when they collaborate with industry to themselves in periods when they do not, a fixed-effects strategy. I also present similar estimates from random-effects models that compare the much larger set of all scientists with each other. Random-effect models allow me to include scientists whose relationships with industry do not change over the course of their *Arabidopsis* publishing careers.

The unit of analysis in these models is scientist-publication year. For example, if Martin Yanofsky from UC San Diego published one article in 1984, three articles in 1986 and two articles in 1987, he would receive three separate entries in the regression corresponding to 1984, 1986 and 1987. For 1986 and 1987, when he published more than one article, I take the average values of Yanofsky's materials and manuscript sharing along with other variables across all articles he published in each year. I then predict scientific sharing as a function of industry ties, holding other influences constant.

I use linear panel models to examine how industry ties influence the dissemination of ideas as captured by the widening circle of articles that reference them over time. The level of analysis in these models is article-citation year, but industry sponsorship cannot change for a given article over the years. As a result, I compare the distances covered in all article-years with one another. I predict the distance that an article's citations travel using industry ties of the article's authors, controlling for influences from other features of the social and scientific context.

Table 1 lists descriptive statistics for the variables used in scientist-level analysis. The means and standard deviations describe the 3,645 PIs and 7,424 scientist-publication years in the random effects models. Table 1 also includes variable means for the sub-

sample used in the fixed-effect model that deviates most from the overall mean, and the ratio of that sub-sample to the sample mean. The means for acknowledgement mentions of sharing and citation cycles are .65 and .03, respectively, but both are rare events—0 the vast majority of the time. Seven percent of the universities or research institutes in which scientists work are reported as maintaining industry collaboration, while 12% are purported to receive industry money. Sixty-six percent of the PI scientists report government funding, 10% industry collaboration and 7% company funding. These numbers suggest that industry funding in plant biology may be slightly lower than average for the broader bio-sciences, most of which have a health focus, where industry funding represents about 12% of the total (Blumenthal, et al. 1996). While scientist-level industry funding is nearly identical across analyses, government funding and company collaboration were somewhat lower in the fixed-effect subsamples.<sup>xiii</sup>

---

Table 1 about here

---

Table 2 describes variables in the article-level diffusion analysis, comprising 6,316 articles in 31,046 citation years. The average minimum distance between the institutions that host the research behind an article and those that cite it is 2,187 miles. Eighty percent of the articles that cite a focal paper do not share any authors, 83% do not share any organizations, and 60% do not share any scientific terms. While 75% and 7% of the articles mention government and industry funding, respectively, 78% and 12% were authored by PIs who mention such funding in at least one article published that year.

---

Table 2 about here

---

## Results

Coefficient estimates for the PI-level Poisson models of sharing, listed in Table 3, are exponentiated and should be interpreted as *incidence ratios*. In the case of industry funding, an incidence ratio represents the ratio of the materials or manuscripts shared in years when a PI received industry funding to years in which she did not. A value greater than one indicates that industry funding is associated with more sharing; a value below one indicates less sharing.<sup>xiv</sup>

Table 3 shows that the flow of citations to a scientist's subfield increases the likelihood that she will be mentioned as sharing materials, but the stock decreases it. This suggests that scientists in the most rapidly growing fields are most likely to have the opportunity to share materials—they will most likely be asked. Presence of a technology transfer office also increases the likelihood that a scientist will share. It is possible that these offices, by crafting and sometimes imposing material transfer agreements (MTAs), actually facilitate sharing through regulation. In the within-scientist (fixed-effect) model, organization-level industry funding and PI-level industry collaboration and funding all significantly associate with less materials sharing. If a PI's university enters a relationship with a company, that PI will be more than 17% (exponentiated coefficient =  $.828 \pm .051$ ,  $p < .01$ :  $100 \times (.828 - 1) = -17.2\%$ ) less likely to be mentioned for sharing materials in the subsequent five years. A scientist's own industry collaborations ( $.857 \pm .049$ ,  $p < .01$ ) and funding ( $.883 \pm .051$ ,  $p < .01$ ) reduce their likelihood of sharing much

further. These findings are consistent with random-effects models run on the same sample,<sup>xv</sup> and those run on the complete sample, nearly ten times as large (3645 versus 380 PI's). The only change between fixed- and random-effects models is that company collaboration no longer becomes significantly associated with sharing.

---

Table 3 about here

---

The determinates of manuscript sharing are somewhat different. Subfield citation stocks and flows both negatively relate to the likelihood that contemporaneous articles will cite a scientist's papers. In the fixed-effect model, organization-level industry collaboration is associated with twice as many traces of manuscript sharing ( $2.0 \pm .829$ ,  $p < .10$ ), although industry funding appears to make sharing about half as likely ( $.418 \pm .154$ ,  $p < .01$ ). At the scientist-level, government funding is associated with more than three times as much manuscript sharing, but industry collaboration and funding does not significantly alter it. In the full-sample random-effects model, garnering additional data from thousands more PIs who only shared or did not share in their *Arabidopsis* publishing careers, industry collaboration and funding do appear positively associated. Collaboration is associated with a doubling in the amount of sharing measured ( $2.172 \pm .454$ ,  $p < .01$ ), while industry funding is associated with a rise of 64% ( $1.643 \pm .334$ ,  $p < .05$ ). These results are broadly consistent with expectations that industry funding would be associated with less materials sharing and more manuscript sharing.

Table 4 lists the results for article-level regression models that predict the distances that cited ideas travel from their sources over time as a function of industry collaboration and sponsorship. In the models presented in Table 4, funding is measured

at the level of the scientist, assuming that PIs do not maintain walls between their various projects and that the government or industry funding claimed for one project may influence others. The between-case  $R^2$  for models ranges from .097 in the organization distance regression to .243 in the one for geographic distance. The estimated coefficient for citation lag, or the difference between the year of citation and publication, is positive and significant for author, organizational and geographic distance. Each year following an article's publication, its ideas are referenced by work produced, on average, by 2% more unique authors and organizations and 66 miles further from the source. As the number of citations to an article increases, the average distance between an article and its citers increases by 10 miles and .04% new authors and organizations each year. These effects contrast with those in the term distance model. As the years go by, a greater proportion of an article's citations come from science that is relatively close. Moreover, more citations do not necessarily mean that an idea is spreading to new areas—it is as likely to catch fire at home.

---

Table 4 about here

---

All four industry collaboration variables post negative influences on the social, organizational and geographic distance models, but not all are statistically significant. Scientist-level industry funding has the most substantial effect: it reduces the spread of an article's ideas to new scientists by 2% and to new organizations by 3% each year. Such funding is not individually significant in the geographic distance model, but along with PI-level industry collaboration ( $f$ -statistic = 6.03,  $p = .0141$ ), industry ties to an article's authors reduce the spread of that paper's ideas by 147 miles each year. When combined

with organization-level collaboration and funding ( $f = 5.49, p = .0191$ ), industry ties of any type reduce the spread by nearly 200 miles annually.

Unlike author, organization and geographic distance, the average distance an article traveled into new scientific territory, as calculated by the proportion of distinctive scientific keywords within its citing articles, falls rather than rises with each passing year. Moreover, government funding significantly reduces the spread of science to these new areas, likely by sponsoring work on canonical problems at the center of scientific subfields. Industry collaboration and funding posted no significant influence on the spread of a paper's ideas to new areas of science, although the coefficient for company funding was negative.

In alternate specifications of all distance models, I examined the interaction between years since article publication and funding variables, but the influence was not statistically significant. This suggests that funding effects were relatively stable over time.

## **Discussion**

Interviews with academic *Arabidopsis* scientists and with company researchers and executives suggest differences in their handling of discoveries and materials. Industry researchers *managed* insights—they sought to develop them into technologies and were discreet, rather than sharing them before rights were secure. Academics, on the other hand, were practiced at *communicating* ideas and *sharing* resources. Many were ideologically committed to the principle of open sharing. Some academics selectively restricted their sharing in order to barter skills and resources for credit. Nevertheless,

academics were seen, by the companies that partnered with them, as inconsistently professional and often unreliable in protecting research ideas and resources. Interviews suggested that collaborations with industry cautioned, instructed and assisted them to manage those resources with greater discretion.

Findings from the quantitative analysis reinforce these insights and expand on them by exploring the consequences of these differences. Statistical models based on article and patent data demonstrate that in academic *Arabidopsis* research, industry funding and to a lesser extent collaboration reduces the sharing of materials and increases its substitute, the interrogation of pre-publication manuscripts. By contrast, government funding tends to increase the sharing of both. Models of the spread of article citations suggest that these patterns have consequences for the cumulative dissemination of ideas and findings through scientific papers. Insofar as citations to an article are proxy for engagement with its ideas, industry funding reduces the distance that ideas travel from industry-sponsored labs. Figure 2 illustrates these patterns by graphing the influence of industry and government sponsorship on the spread of scientific citations. This figure shows how industry sponsorship deflects the dissemination of ideas away from their natural path to new authors, institutions and regions. In the idiom of Actor-Network Theory, by enrolling companies in their research, scientists appear less able to enroll other academics.

Industry-sponsored ideas do not, however, travel to a smaller proportion of new scientific areas. In fact, government funding helps to support scientific ideas within the areas in which they were first advanced, but industry funding does not. Scientific articles appear to be influenced not only by industry funding reported directly in them, but also

by industry funding reported in contemporaneous articles by the same authors. This implies that labs tend to behave as integrated entities, without hermetic separation between projects or multiple modes of openness.

---

Figure 2 about here

---

These models intimate a causal influence from industry, although subsequent testing suggested that this influence was mitigated by the self-selection of scientists into relations with industry.<sup>xvi</sup> Moreover, the nature of the causal influence that remained is not entirely clear. There were no substantial differences between collaborative and transactional industry ties—they behaved like substitutes in the models examined. The models do, however, control for movement between subfields, and so imply that industry sponsorship influences scientists' willingness to share requested materials independent of requests for those materials.

In a related paper, I show that government funding sponsors scientists to perform confirmatory experiments in well-trodden areas, while industry influences them to explore new territory with speculative techniques. In the network of all scientific terms, connected by co-presence in articles, government funding nudges sponsored terms toward the centers of dense hubs of duplication, while industry funding pushes them to the periphery (Evans, 2010). These parallel findings help to explain the pattern of effects in the final model predicting scientific spread: Governments fund incremental advance within fields; Companies funds speculation across and beyond them. Alongside the finding that smaller scientific subfields reduce the amount of scientific sharing possible, this suggests that industry alters academic sharing by influencing supply *and* demand:

Company partnerships heighten an academic lab's caution, but they also nudge it to engage in less crowded academic areas where few will request or find interest in their discoveries and materials.<sup>xvii</sup>

Note, however, that less sharing may not be so much a consequence of path-breaking research as its cause. Exploring new areas facilitates secrecy by avoiding competition. The chief technology officer of a plant genomics company made this clear when he noted: 'being a company ... you've got to worry about the competition and the competition is really the public sector, and increasingly so' [BM26]. That industry ties correlate with more manuscript sharing dispels the notion that industry funding only makes *Arabidopsis* research less academically appealing. Interview and field-level evidence together demonstrate that commercial relationships exert direct and indirect influence on the selectivity of sharing and the management of ideas in academic labs. Companies enroll scientists in their project of appropriating future rights to innovations by influencing them to avoid as well as deny their competitors.

This research ratifies the skepticism that social studies of science have long held for the caricature of open public science verses closed private research. In the world of plant science, both sectors allow and reward a measure of privacy. This article nevertheless demonstrates that academic science is less reliably closed—its research is more open—than industry research, and that it becomes more closed as the two collaborate.

In the name of scientific advance, governments have considered attempting to force academics to share, but the cost would be high. Not only would government oversight change the way many academics operate, regardless of industry ties, but fewer

relationships with industry and the development of fewer marketable technologies would likely result. This research also offers the comforting implication that scientific advance will continue if labs become secretive. Competitors can still reverse-engineer the manuscripts and articles of closed labs, but this will slow and fragment the accumulation of knowledge.

Because *Arabidopsis* research retains a community tradition and remains a large step away from commercial application, it represents a conservative case of industry influence. We would expect industry's sway on sharing to be larger in areas like biomedical, agricultural, or engineering research. In more applied areas like clinical medicine, scholars and participants have tried to demonstrate that industry keeps unfavorable academic work out of circulation and biases published results. This research suggests that industry collaborations reach through academic collaborators to alter academic interactions and exchanges, curbing the spread of scientific materials and ideas. In an age when Western governments have slowed spending on academic research, industry represents a higher proportion of academic support. This research encourages further study into not only how broadly, but how deeply markets penetrate science and alter it.

## Methodological Appendix

*Publication Data.* Articles using *Arabidopsis thaliana* were collected by combining publication data from BIOSIS, PubMed, Medline, AGRICOLA, and SciSearch where ‘Arabidopsis’ or ‘thaliana’ was mentioned in the title, abstract or author-provided-keywords of an article. With the help of the U.S. Department of Agriculture and the Arabidopsis Information Resource (TAIR), abstracts were added to this corpus from *Arabidopsis* papers presented in plant molecular biology conferences over the past several years. ThomsonReuter’s SCI links these to 45,577 non-*Arabidopsis* articles that cite them.<sup>xviii</sup> I annotated titles and abstracts with regard to methods and species used, genes and proteins examined, and biological processes, components and functions explored.<sup>xix</sup> I collected 8,400 ‘acknowledgement’ sections from *Arabidopsis* papers published between 1975 and 2000 that were also present in the SciSearch database.

*Principal Investigators and Contexts.* I established the existence and location of *Arabidopsis* principal investigators through a TAIR database of all individuals and organizations in the *Arabidopsis* community who had ever published *Arabidopsis* papers or used *Arabidopsis* data. Where labs weren’t specified in TAIR, I used bibliographical data from ISI and BIOSIS, which electronically list the institutional locations for the authors of each article in the database. In the field of molecular plant biology, the last author on a research article is almost always (with 90% accuracy) a PI—the head scientist in a research lab. For the universities in the sample, a number of characteristics were obtained from the World Education Database, such as the presence of a medical or agricultural school and the university’s broader identity as an agricultural, technical,

medical or general research university.

*Materials Sharing.* Attributions of materials sharing and technical assistance were isolated in acknowledgements through a rule-based parsing and matching procedure that, when tested against a sample of tagged author data, was able to identify approximately 90% of the other PIs mentioned.

*Pre-Publication Sharing.* My measure of pre-publication sharing—the number of citation cycles in which an article is implicated—is naturally sensitive to the lag between submission and acceptance in a particular journal. I scanned the submission and acceptance dates of 6,253 papers and found the lag relatively consistent within *Arabidopsis* publications at a mean of six months (191 days) and a standard deviation of about the same (196 days).

*Dissemination.* Measurement of article dissemination beyond authors and organizations required the precise identification of those entities. Both manual and automated methods were used to unify article authors and article institutions, despite idiosyncrasies in spelling and format. This insured that author- and organization-distance scores accurately reflected the spread of article influence and not merely idiosyncratic differences in the database record.

*Geographic Spread.* To associate each institution with geographic coordinates, I parsed each of the postal zip codes for U.S. institutions and then assigned latitude and longitude coordinates based on U.S. Postal Service information (Adams 2001 ; Sorenson and Fleming 2004 ; Sorenson and Stuart 2001). For non-U.S. institutions, fuzzy matching algorithms matched institution addresses against international city and region names from the National Geospatial-Intelligence Agency's (NGA) and U.S. Board on

Geographic Names' (US BGN) database of foreign area names.<sup>xx</sup> The distance between the closest institutions in each article set was established assuming a spherical earth, and calculating the following equation between every cited institution  $x$  and every citing institution  $y$ , such that the distance, in miles, between  $x$  and  $y$  =

$$3943.424 \times \ar \cos \left[ \left[ \cos \left( \left( \frac{2\pi}{360} \times x\_longitude \right) - \left( \frac{2\pi}{360} \times y\_longitude \right) \right) \right] \times \cos \left( \frac{2\pi}{360} \times x\_latitude \right) \times \cos \left( \frac{2\pi}{360} \times y\_latitude \right) \right] + \sin \left( \frac{2\pi}{360} \times x\_latitude \right) \times \sin \left( \frac{2\pi}{360} \times y\_latitude \right) \right]$$

where  $x$  and  $y$  latitudes and longitudes are measured in degrees.<sup>xxi</sup>

*Controls.* Scientist status was measured as a function of two properties: 1) the scientist's closeness centrality within the social network defined by article coauthorship, accumulating over time; 2) the centrality of a scientist's articles within the network of scientific references inscribed by the bibliographies of all *Arabidopsis* articles listed in ThomsonReuter's Science Citation Index. Because citations to articles atrophy over time, with a citation half-life of about five years, they were allowed to degrade following a cumulative normal distribution from a value of 1.0 in year 5 to one of .25 a decade later. The role of scientists as brokers was measured using Freeman's measure of betweenness centrality within coauthorship networks. The diversity of subjects studied using *Arabidopsis* is calculated utilizing a herfindahl index,  $1 - \sum_{i=1}^n (s_i^2)$ , where  $s_i$  represents the percentage of the university's *Arabidopsis* research performed within any one subfield. This captures the degree to which multiple subfields are represented within the university. Similarly, the diversity of institutional departments studying *Arabidopsis* is computed as one minus the herfindahl index of research from each type of department studying *Arabidopsis* within that organization.

To control for the presence of a technology transfer office, I ascertained data for North American universities through the annual survey of the Association for University

Technology Managers (AUTM). Information in this survey was collected on only North American institutions. Because firms generally considered the U.S. and Canada to be the most commercial academic environments in the world, I gave scientists from non-North American institutions a score of 0, even though some of the European institutions did create such offices in the later period of the study. As a result, this is an indicator of cross-sector facilitation in the most commercial environments.

*Modeling Strategy.* My metrics of materials and manuscript sharing occur infrequently, both have a relatively low dispersion<sup>xxii</sup> and so I assume

$$Sharing_{it} \sim \text{Poisson}(\mu_{it})$$

where  $Sharing_{it}$  is the measure of materials or manuscript sharing—the number of acknowledgement mentions scientist  $i$  receives from years  $t+1$  to  $t+5$  or the distinct prepublication cites his or her year  $t$  articles receive. Specifically, I model the following, specifying fixed effects:

$$\log(\mu_{it}) = \alpha + x_{it}\beta + \varepsilon_{it},$$

where  $\beta$  is a vector of coefficients that indicate the effect of covariates  $x_{it}$  on sharing. The fixed effect specification allows me to difference out the influence of unobserved heterogeneity over time. This substantially reduces the dataset because it requires that scientists demonstrate both periods of sharing and non-sharing. To test whether industry collaboration associates with sharing, not only within researchers but across the dataset, I also performed and present random effects models of the same.<sup>xxiii</sup> Moreover, Hausman tests comparing the coefficients of fixed and random-effects models using the fixed-effects sample do not favor the more conservative, fixed-effects estimation strategy.

The various distances across which scientific ideas moved were distributed normally and the following linear model was estimated with OLS:

$$Distance_{it} = \alpha_{it} + citelag_{it}\delta + x_{it}\beta + v_i + \varepsilon_{it}$$

where  $Distance_{it}$  refers to measures of social, organizational, geographic and scientific distance covered by ideas in article  $i$  by citation year  $t$ ,  $citelag_{it}$  represents the number of years between publication and  $t$ , and  $v_i + \varepsilon_{it}$  is the residual with  $v_i$  representing the article-specific component that differs only between and not within articles. The term  $x_{it}$  refers to variables utilized in the scientist-level of analysis, but averaged across a paper's PIs. Because the distance analysis is performed at the article-level, I first averaged all of the measures of context across that article's authoring scientists. For example, if a paper is authored by two scientists, one working at a MIT and the other Los Alamos National Lab, then the control variables *science-and-technology university* and *government lab* would each equal .5 for that article.

## References

- Adams, James D. (2001) 'Comparative Localization of Academic and Industrial Spillovers', *NBER Working Paper* 7843.
- Adams, James D. and Jaffe, Adam B. (1996) 'Bounding the Effects of R&D: An Investigation Using Matched Establishment-Firm Data', *NBER Working Paper* 5544.
- Ankeny, Rachel (2007) 'Model Organisms as Case-Based Reasoning: Worms in Contemporary Biomedical Science', in Angela N. H. Creager, Elizabeth Lunbeck and M. Norton Wise (eds), *Science without laws : model systems, cases, exemplary narratives* (Durham: Duke University Press): 287 p.
- Arrow, Kenneth J. (1962) 'Economic Welfare and the Allocation of Resources for Inventions', in Richard R. Nelson (ed), *The Rate and Direction of Inventive Activity: Economic and social factors* (Princeton, NJ: Princeton University Press).
- Bekelman, J. E., Li, Y. and Gross, C. P. (2003) 'Scope and impact of financial conflicts of interest in biomedical research: a systematic review', *JAMA* 289/4 (Jan 22-29): 454-65.
- Berman, E. P. (2008) 'Why Did Universities Start Patenting? Institution-Building and the Road to the Bayh-Dole Act', *Social Studies of Science* 38/6 (Dec): 835-71.
- Blumenthal, David, Campbell, Eric G., Causino, Nanyanne and Louis, Karen S. (1996) 'Participation of Life-Science Faculty in Research Relationships with Industry', *New England Journal of Medicine* 355: 1734-39.
- Buttel, Frederick H. and Belsky, Jill (1987) 'Biotechnology, Plant Breeding, and Intellectual Property: Social and ethical dimensions', *Science, Technology, and Human Values* 12/1: 31-49.
- Calvert, J. (2006) 'What's special about basic research?', *Science Technology & Human Values* 31/2 (Mar): 199-220.
- Cambrosio, Alberto, Keating, Peter and Mogoutov, Andrei (2004) 'Mapping Collaborative Work and Innovation in Bioscience', *Social Studies of Science* 34/3: 325-64.
- Campbell, Eric G., Clarridge, B. R., Gokhale, M., Birenbaum, L., Hilgartner, S., Holtzman, N.A. and Blumenthal, David (2002) 'Data Withholding in Academic Genetics: Evidence from a National Survey', *Journal of the American Medical Association* 287/4: 473-80.
- Chang, Jeffrey, Schutz, Hinrich and Altman, Russ (2003) 'GAPSCORE: Finding gene and protein names one word at a time', *Stanford Working Paper*.
- Charles, Daniel (2001) *Lords of the Harvest: Biotech, big money, and the future of food* (Cambridge, MA: Perseus publishing).
- Cho, M. K. and Bero, L. A. (1996) 'The quality of drug studies published in symposium proceedings', *Ann Intern Med* 124/5 (Mar 1): 485-9.
- Clifford, T. J., Moher, D. and Barrowman, N. (2003) 'Funding source, trial outcome and reporting quality: Are they related? Results of a pilot study.', in.
- Cohen, Wesley M., Nelson, Richard R. and Walsh, John P. (2000) 'Protecting their Intellectual Assets: Appropriability conditions and why U.S. manufacturing firms patent (or not)', *NBER Working Paper* 7552.
- Collins, H. M. (1974) 'Tea Set - Tacit Knowledge and Scientific Networks', *Science Studies* 4/2: 165-85.
- Collins, H. M. (1985) *Changing order : replication and induction in scientific practice* (London ; Beverly Hills: Sage Publications).
- Dasgupta, Partha and David, Paul A. (1994) 'Toward a New Economics of Science', *Research Policy* 23: 487-521.
- David, Paul A. (2003) 'From Keeping 'Nature's Secrets' to the Institutionalization of 'Open Science'', *Working Paper* <http://www-econ.stanford.edu/faculty/workp/swp04006.pdf>.
- Davidson, R. A. (1986) 'Source of funding and outcome of clinical trials', *J Gen Intern Med* 1/3 (May-Jun): 155-8.
- Dickersin, K., Min, Y. I. and Meinert, C. L. (1992) 'Factors influencing publication of research results. Follow-up of applications submitted to two institutional review boards', *JAMA* 267/3 (Jan 15): 374-8.
- Djulgovic, B., Bennet, C. L. and Lyman, G. H. (1999) 'Violation of the uncertainty principle in the conduct of randomized controlled trials (RCTs) of erythropoietin (EPO)', *Blood* 94/supplement 1.
- Easterbrook, P. J., Berlin, J. A., Gopalan, R. and Matthews, D. R. (1991) 'Publication bias in clinical research', *Lancet* 337/8746 (Apr 13): 867-72.
- Faberge, A. C. (1982) 'Open Information and Secrecy in Research', *Perspectives in Biology and Medicine* 25: 263-78.
- Godin, B. (2006) 'The linear model of innovation - The historical construction of an analytical framework', *Science Technology & Human Values* 31/6 (Nov): 639-67.
- Grandin, Karl, Wormbs, Nina and Widmalm, Sven (2004) *The science-industry nexus : history, policy, implications : Nobel Symposium 123* (Sagamore Beach, MA: Science History Publications/USA).
- Hackett, Edward J. (2005) 'Essential Tensions: Identity, Control and Risk in Research', *Social Studies of Science* 35/5: 787-826.

- Heller, Michael A. and Eisenberg, Rebecca S. (1998) 'Can Patents Deter Innovation? The anticommens in biomedical research', *Science* 280: 698.
- Hine, Christine (2006) 'Databases as Scientific Instruments and Their Role in the Ordering of Scientific Work', *Social Studies of Science* 36/2: 269-98.
- Jadad, A. R., Moher, M., Browman, G. P., Booker, L., Sigouin, C., Fuentes, M. and Stevens, R. (2000) 'Systematic reviews and meta-analyses on treatment of asthma: critical evaluation', *BMJ* 320/7234 (Feb 26): 537-40.
- Jaffe, A. B., Trajtenberg, M. and Henderson, R. (1993) 'Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations', *Quarterly Journal of Economics* 108/3 (Aug): 577-98.
- Kaiser, D. (2002) 'Cold War requisitions, scientific manpower, and the production of American physicists after World War II', *Historical Studies in the Physical and Biological Sciences* 33: 131-59.
- Kevles, Daniel J. (1978) *The physicists : the history of a scientific community in modern America* (New York: Knopf : distributed by Random House).
- Kjaergard, L. L., Nikolova, D. and Gluud, C. (1999) 'Randomized clinical trials in HEPATOLOGY: predictors of quality', *Hepatology* 30/5 (Nov): 1134-8.
- Kleinman, D. L. and Solovey, M. (1995) 'Hot Science/Cold War: the National Science Foundation after World War II', *Radical History Review* /63 (Fal): 111-&.
- Kleinman, Daniel Lee (1998) 'Pervasive Influence: Intellectual Property, Industrial History, and University Science', *Science and Public Policy* 25/2: 95-102.
- Knorr-Cetina, K. (1999) *Epistemic cultures : how the sciences make knowledge* (Cambridge, Mass.: Harvard University Press).
- Knox, K. S., Adams, J. R., Djulbegovic, B., Stinson, T. J., Tomor, C. and Bennet, C. L. (2000) 'Reporting and dissemination of industry versus non-profit sponsored economic analyses of six novel drugs used in oncology', *Ann Oncol* 11/12 (Dec): 1591-5.
- Kohler, Robert E. (1994) *Lords of the fly : Drosophila genetics and the experimental life* (Chicago: University of Chicago Press).
- Latour, Bruno (1987) *Science in action : how to follow scientists and engineers through society* (Cambridge, Mass.: Harvard University Press).
- Latour, Bruno and Woolgar, Steve (1979) *Laboratory life : the social construction of scientific facts* (Beverly Hills: Sage Publications).
- Leibeskind, Julia Porter (2000) 'Property Rights, Secrecy, and the Boundaries of the Firm: Theory and evidence', *Working Paper*.
- Leslie, Stuart W. (1993) *The Cold War and American science : the military-industrial-academic complex at MIT and Stanford* (New York: Columbia University Press).
- Lexchin, J., Bero, L. A., Djulbegovic, B. and Clark, O. (2003) 'Pharmaceutical industry sponsorship and research outcome and quality: systematic review', *BMJ* 326/7400 (May 31): 1167-70.
- Lolle, S. J., Victor, J. L., Young, J. M. and Pruitt, R. E. (2005) 'Genome-wide non-mendelian inheritance of extra-genomic information in Arabidopsis', *Nature* 434/7032 (Mar 24): 505-9.
- Mandelkern, M. (1999) 'Manufacturer support and outcome', *J Clin Psychiatry* 60/2 (Feb): 122-3.
- Marshall, E. (2002) 'Data sharing. DNA sequencer protests being scooped with his own data', *Science* 295/5558 (Feb 15): 1206-7.
- Masco, Joseph (2006) *The nuclear borderlands : the Manhattan Project in post-Cold War New Mexico* (Princeton, N.J.: Princeton University Press).
- Massie, B. M. and Rothenberg, D. (1993) 'Publication of sponsored symposiums in medical journals', *N Engl J Med* 328/16 (Apr 22): 1196-7; author reply 97-8.
- McCain, Katherine W. (1991) 'Communication, Competition, and Secrecy: The Production and Dissemination of Research-Related Information in Genetics', *Science, Technology and Human Values* 16/4: 491-516.
- Mello, M. M., Clarridge, B. R. and Studdert, D. M. (2005) 'Academic medical centers' standards for clinical-trial agreements with industry', *N Engl J Med* 352/21 (May 26): 2202-10.
- Merrill, S. A. and Cooper, R. S. (1999) 'Trends in Industrial Research and Development: Evidence from national data sources', in *Securing America's Industrial Strength, National Research Council Board on Science, Technology and Economic Policy* (Washington, D. C.: National Academy Press): 99-116.
- Merton, Robert K. (1942) 'Science and Technology in a Democratic Order', *Journal of Legal and Political Sociology* 1: 115-26.
- Merton, Robert K. (1957) 'Priorities in Scientific Discovery: A Chapter in the sociology of science', *American Sociological Review* 22: 635-59.

- Merton, Robert King (1976) 'The Ambivalence of Scientists', in Robert King Merton (ed), *Sociological ambivalence and other essays* (New York: Free Press): 32-65.
- Meyerowitz, Elliot M. (2001) 'Prehistory and History of *Arabidopsis* Research', *Plant Physiology* 125/January: 15-19.
- Mitroff, I. I. (1974) 'Norms and Counter-Norms in a Select Group of Apollo Moon Scientists - Case Study of Ambivalence of Scientists', *American Sociological Review* 39/4: 579-95.
- Mowery, David C. (1999) 'America's Industrial Resurgence?: An Overview', in, *U.S. Industry in 2000: Studies in Competitive Performance* (Washington, D.C.: National Academy Press): 1-16.
- Mulkay, M. J. (1976) 'Norms and Ideology in Science', *Social Science Information Sur Les Sciences Sociales* 15/4-5: 637-56.
- Mulkay, M. J. and Edge, David O. (1973) 'Social, Cognitive and Technical Factors in the Growth of Radio Astronomy'', *Social Science Information* 12/6: 36.
- Murray, Fiona and Stern, Scott (2005) 'Do Formal Intellectual Property Rights Hinder the Free Flow of Scientific Knowledge? An empirical test of the anti-commons hypothesis', in, *NBER Working Paper* (Cambridge, MA).
- Neumann, P. J., Sandberg, E. A., Bell, C. M., Stone, P. W. and Chapman, R. H. (2000) 'Are pharmaceuticals cost-effective? A review of the evidence', *Health Aff (Millwood)* 19/2 (Mar-Apr): 92-109.
- Nottenburg, Carol, Pardey, Philip G. and Wright, Brian D. (2001) 'Accessing Other People's Technologies: Do non-profit research agencies need it? How to obtain it.', *International Food Policy Research Institute's (IFPRI) EPTD Discussion Paper* 79.
- Polanyi, Michael (1962) 'The Republic of Science: Its Political and Economic Theory', *Minerva* 1: 54-74.
- Powell, Walter W. (1990) 'Neither Market Nor Hierarchy: Network forms of organization', *Research in Organizational Behavior*.
- Powell, Walter W. (1996) 'Interorganizational Collaboration in the Biotechnology Industry', *Journal of Institutional and Theoretical Economics* 120: 197-215.
- Press, Eyal and Washburn, Jennifer (2000) 'The Kept University', *Atlantic Monthly* 285: 39-54.
- Rader, Karen A. (2004) *Making mice : standardizing animals for American biomedical research, 1900-1955* (Princeton, N.J.: Princeton University Press).
- Relman, Arnold S. (1980) 'The new medical-industrial complex', *The New England Journal of Medicine* 303/17: 963-70.
- Relman, Arnold S. (1984) 'Dealing with Conflict of Interest', *The New England Journal of Medicine* 310/18: 1182-3.
- Remington, John A. (1988) 'Beyond Big Science in America: The Binding of Inquiry', *Social Studies of Science* 18: 45-72.
- Rochon, P. A., Gurwitz, J. H., Cheung, C. M., Hayes, J. A. and Chalmers, T. C. (1994) 'Evaluating the quality of articles published in journal supplements compared with the quality of those published in the parent journal', *JAMA* 272/2 (Jul 13): 108-13.
- Rosenberg, Nathan (1982) 'The Growing Role of Science in the Innovation Process', in Carl Gustaf Bernhard, Elisabeth Crawford and Per Sorbom (eds), *Science, Technology and Society in the Time of Alfred Nobel* (Oxford: Pergamon Press): 231-46.
- Sahlins, Marshall David (1972) *Stone age economics* (Chicago,: Aldine-Atherton).
- Sapolsky, Harvey M. (1990) *Science and the Navy : the history of the Office of Naval Research* (Princeton, N.J.: Princeton University Press).
- Shapin, Steven (1994) *A social history of truth : civility and science in seventeenth-century England* (Chicago: University of Chicago Press).
- Shapin, Steven (2008) *The scientific life : a moral history of a late modern vocation* (Chicago: University of Chicago Press).
- Sismondo, S. (2008) 'Pharmaceutical company funding and its consequences: a qualitative systematic review', *Contemp Clin Trials* 29/2 (Mar): 109-13.
- Somerville, C. and Koornneef, M. (2002) 'A fortunate choice: the history of *Arabidopsis* as a model plant', *Nat Rev Genet* 3/11 (Nov): 883-9.
- Somerville, C. R. (2001) 'An early *Arabidopsis* demonstration. Resolving a few issues concerning photorespiration', *Plant Physiol* 125/1 (Jan): 20-4.
- Sorenson, Olav and Fleming, Lee (2004) 'Science and the Diffusion of Knowledge', *Research Policy* 33/10: 1615-34.
- Sorenson, Olav and Stuart, Toby (2001) 'Syndication Networks and the Spatial Distribution of Venture Capital Investments', *American Journal of Sociology* 106: 1546-88.
- Stelfox, H. T., Chua, G., O'Rourke, K. and Detsky, A. S. (1998) 'Conflict of interest in the debate over calcium-channel antagonists', *N Engl J Med* 338/2 (Jan 8): 101-6.
- Stern, J. M. and Simes, R. J. (1997) 'Publication bias: evidence of delayed publication in a cohort study of clinical research projects', *BMJ* 315/7109 (Sep 13): 640-5.

- Vandenbroucke, J. P., Helmerhorst, F. M. and Rosendaal, F. R. (2000) 'Competing interests and controversy about third generation oral contraceptives. BMJ readers should know whose words they read', *BMJ* 320/7231 (Feb 5): 381-2.
- Wahlbeck, K. and Adams, C. (1999) 'Beyond conflict of interest. Sponsored drug trials show more-favourable outcomes', *BMJ* 318/7181 (Feb 13): 465.
- Walbot, V. (2000) 'Arabidopsis thaliana genome. A green chapter in the book of life', *Nature* 408/6814 (Dec 14): 794-5.
- Washburn, Jennifer (2005) *University, Inc. : the corporate corruption of American higher education* (New York: Basic Books).
- Watson, James D. (1968) *The double helix: a personal account of the discovery of the structure of DNA* (London,: Weidenfeld & Nicolson).
- Westfall, Richard S. (1980) *Never at rest : a biography of Isaac Newton* (Cambridge [Eng]. ; New York: Cambridge University Press).
- Yaphe, J., Edman, R., Knishkowsky, B. and Herman, J. (2001) 'The association between funding by commercial interests and study outcome in randomized controlled drug trials', *Fam Pract* 18/6 (Dec): 565-8.

Table 1. Variables in scientist-level analysis

variables	mean <sup>a</sup>	st.dev	max <sup>b</sup>	largest deviating mean <sup>c</sup>	ratio of deviating to overall mean
<i><u>Dependent variables</u></i>					
Acknowledgement mentions	0.654	2.389	45	1.736	2.654
Citation cycles <sup>d</sup>	0.029	0.211	5	0.210	7.139
<i><u>Subfield-level</u></i>					
Flow of article citations (number in last year)	1480.496	761.082	3348	21.445	0.014
Stock of article citations (number in years prior to last)	9589.816	6782.155	25616	1116.864	0.116
Commercial cluster (high patent citations to clust.: 0/1)	0.343		1	0.027	0.080
<i><u>Organization-level</u></i>					
Subfield diversity (1-Herfindahl of subfields)	0.797	0.167	0.910	0.045	0.057
Department diversity (1-Herfindahl of departments)	0.284	0.232	0.686	0.002	0.008
Transfer Office (U.S.) (0/1)	0.319		1	0.212	0.666
Company collaboration	0.072		1	0.036	0.497
Company funding/joint commercializing	0.119		1	0.078	0.658
<i><u>PI-level</u></i>					
Time since first <i>Arabidopsis</i> publication	2.200	3.130	26	2.364	1.075
Time <sup>2</sup>	14.634	38.503	676	20.127	1.375
Coauthor closeness centrality	0.126	0.074	0.290	0.001	0.011
Coauthor betweenness centrality	0.001	0.003	0.060	0.002	2.147
Article citation centrality (eroding over time)	0.275	0.179	0.637	0.113	0.413
Government funding (0/1)	0.656		1	0.178	0.271
Company collaboration (0/1)	0.095		1	0.046	0.482
Company funding (0/1)	0.074		1	0.078	1.051

<sup>a</sup>N=7424 for random effects analysis;

<sup>b</sup>All minimums are 0.

<sup>c</sup>Mean deviating most from overall mean for fixed effects analysis of acknowledgement mentions or citation cycles. N=1923 for fixed effects analysis of acknowledgement mentions; N=682 for fixed effect analysis of citation cycles.

<sup>d</sup>When only citation cycles involving no shared authors were considered, the mean drops to .011(.12) and the max to 4.

Table 2. Variables in article-level Analysis

Variables	Mean <sup>a</sup>	St.dev.	Min	Max
<i>Dependent Variables</i>				
Average minimum distance	2187.742	1681.449	0	12205.71
Author distance (proportion with no match)	0.800	0.288	0	1
Org. distance (proportion with no match)	0.828	0.272	0	1
Term distance (proportion with no match)	0.594	0.349	0	1
<i>Subfield-level</i>				
Flow of article citations (number in last year)	1399.647	742.659	0	3348
Stock of article citations (number in years prior to last)	7412.135	5774.514	0	25616
Commercial cluster (high patent citations to clust.: 0/1)	0.332		0	1
<i>Organization-level</i>				
Subfield diversity (1-Herfindahl of subfields)	0.820	0.132	0	0.912714
Department diversity (1-Herfindahl of departments)	0.288	0.217	0	0.688017
Transfer Office (U.S.) (0/1)	0.373		0	1
Company collaboration (0/1)	0.066		0	1
Company funding/joint commercializing (0/1)	0.106		0	1
<i>PI-level</i>				
Time since first <i>Arabidopsis</i> publication	2.857	3.118	0	25
Time <sup>2</sup>	20.246	39.176	0	625
Coauthor closeness centrality	0.099	0.075	0	0.284027
Coauthor betweenness centrality	0.002	0.004	0	0.060486
Article citation centrality (eroding over time)	0.298	0.169	0	0.637255
Government funding (0/1)	0.781		0	1
Company collaboration (0/1)	0.103		0	1
Company funding (0/1)	0.120		0	1
<i>Article-level</i>				
Publication minus citation year	3.689	2.984	0	27
Number cites	6.001	6.933	1	100
Government funding (0/1)	0.753		0	1
Company funding (0/1)	0.072		0	1

<sup>a</sup>N=31046 article-years (since publication), 6316 articles

Table 3. Estimated Incidence Ratios from Poisson Models  
Predicting Sharing with Industry Collaboration

Variables	Materials Sharing		Manuscript Sharing	
	Fixed-effects	Random-effects	Fixed-effects	Random-effects
<i>Subfield-level</i>				
Flow of article citations	1.001 ( $4 \times 10^{-5}$ )**	1.001 ( $4 \times 10^{-5}$ )**	0.999 ( $3 \times 10^{-4}$ )*	0.999 ( $2 \times 10^{-4}$ )**
Stock of article citations	0.999 ( $1 \times 10^{-5}$ )**	0.999 ( $9 \times 10^{-6}$ )**	0.999 ( $7 \times 10^{-5}$ )*	0.999 ( $3 \times 10^{-5}$ )**
Commercial cluster	1.239 (0.433)	1.521 (0.503)	0.376 (0.643)	0.002 (0.004)**
<i>Organization-level</i>				
Subfield diversity	1.427 (0.761)	6.800 (2.613)**	0.157 (0.497)	0.347 (0.208)†
Department diversity	0.801 (0.191)	0.776 (0.153)	0.515 (0.734)	1.005 (0.477)
Transfer Office (U.S.)	1.27 (0.113)**	1.298 (0.108)**	0.406 (0.253)	1.619 (0.424)†
Company collaboration	1.041 (0.069)	1.094 (0.072)	2.000 (0.829)†	1.245 (0.337)
Company funding	0.828 (0.051)**	0.881 (0.053)*	0.418 (0.154)*	0.793 (0.168)
<i>Scientist-level</i>				
Time since first pub.	0.94 (0.027)*	1.086 (0.024)**	0.703 (0.144)†	0.837 (0.048)**
Time <sup>2</sup>	1.001 (0.001)	0.998 (0.001)†	1.010 (0.009)	1.006 (0.004)
Coauthor close. Cent.	0.267 (0.242)	$6 \times 10^{-6}$ (0.000)**	$6 \times 10^5$ ( $3 \times 10^6$ )*	$1 \times 10^6$ (33575)**
Coauthor between. cent.	1.066 (6.278)	$4 \times 10^{17}$ ( $2 \times 10^{18}$ )**	$1 \times 10^{40}$ ( $7 \times 10^{41}$ )†	$7 \times 10^5$ ( $2 \times 10^4$ )
Article citation cent.	3.482 (0.886)**	9.314 (2.177)**	495.015 (1102.)**	370.234 (376.800)**
Government funding	1.068 (0.053)	1.055 (0.051)	3.165 (1.415)*	3.325 (0.851)**
Company collaboration	0.857 (0.049)**	0.951 (0.053)	1.461 (0.569)	2.172‡ (0.454)**
Company funding	0.883 (0.051)*	0.852 (0.049)**	0.980 (0.398)	1.643‡ (0.334)*
Observations	1923	7424	682	7424
Number of pi_id	380	3645	115	3645

Standard errors in parentheses

† significant at 10%; \* significant at 5%; \*\* significant at 1%

‡ When only citation cycles involving no shared authors are predicted, the company collaboration and funding effects increase to 2.456(.844)\*\* and 2.02(.638)\*\* , respectively

Table 4. Estimated Coefficients from OLS Panel Models Predicting Article Citation Spread with Company Collaboration and Funding

Variable	Distance variable			
	Social	Organizational	Geographic	Scientific
Concept	Prop. new authors	Prop. new orgs	Average miles	Prop. new terms
<i>Paper-level</i>				
Publication-cite year	0.021 (0.001)**	0.021 (0.001)**	65.56 (3.177)**	-0.008 (0.001)**
Number cites	0.004 (2.919×10 <sup>-4</sup> )**	0.004 (2.750×10 <sup>-4</sup> )**	10.454 (1.600)**	-2.250×10 <sup>-4</sup> (3.388×10 <sup>-4</sup> )
<i>Subfield-level</i>				
Flow article cites	8.06×10 <sup>-6</sup> (5.63×10 <sup>-6</sup> )	2.430×10 <sup>-6</sup> (5.290×10 <sup>-6</sup> )	0.022 (0.031)	-6.83×10 <sup>-6</sup> (6.80×10 <sup>-7</sup> )
Stock article cites	2.31×10 <sup>-6</sup> (8.20×10 <sup>-7</sup> )**	2.50×10 <sup>-6</sup> (7.27×10 <sup>-7</sup> )**	0.008 (0.005) <sup>†</sup>	7.23×10 <sup>-7</sup> (9.85×10 <sup>-7</sup> )
Commercial cluster	0.027 (0.021)	0.032 (0.019)	235.858 (113.235)*	0.01 (0.025)
<i>Organization-level</i>				
Subfield diversity	-0.019 (0.022)	-0.153 (0.021)**	119.053 (122.185)	-0.126 (0.027)**
Department diversity	0.032 (0.015)*	0.051 (0.015)**	-183.756 (85.036)*	0.033 (0.019) <sup>†</sup>
Transfer Office (U.S.)	0.032 (0.009)**	0.024 (0.009)**	242.62 (50.720)**	0.014 (0.011)
Company collaboration	-0.002 (0.013)	-0.003 (0.012)	-21.115 (69.313)	0.005 (0.015)
Company funding	-0.002 (0.009)	-0.001 (0.009)	-29.242 (52.215)	0.009 (0.011)
<i>Scientist-level</i>				
Time since first pub.	-0.002 (0.002)	-0.002 (0.002)	13.806 (12.888)	-0.012 (0.003)**
Time <sup>2</sup>	3.223×10 <sup>-4</sup> (1.587×10 <sup>-4</sup> )*	3.435×10 <sup>-4</sup> (1.492×10 <sup>-4</sup> )*	0.081 (0.875)	4.817×10 <sup>-4</sup> (1.914×10 <sup>-4</sup> )*
Coauthor close. Cent.	-0.117 (0.063) <sup>†</sup>	-0.147 (0.059)*	-1,288.05 (345.073)**	-0.025 (0.075)
Coauthor between. cent.	1.302 (0.787) <sup>†</sup>	1.632 (0.740)*	2,639.32 (4,339.36)	1.525 (0.948)
Article citation cent.	0.029 (0.025)	-0.014 (0.024)	90.734 (140.547)	-0.36 (0.031)**
Government funding	0.002 (0.007)	-0.001 (0.007)	23.284 (40.81)	-0.028 (0.009)**
Company collaboration	-0.011 (0.009)	-0.005 (0.009)	-73.976 (52.003)	0.009 (0.011)
Company funding	-0.021 (0.009)*	-0.028 (0.009)**	-74.348 (51.439)	-0.018 (0.011)
Constant	0.677 (0.023)**	0.836 (0.021)**	1,425.61 (124.781)**	0.819 (0.027)**
R <sup>2</sup> (within/betw/overall)	.049/.109/.074	.054/.097/.077	.018/.243/.157	.005/.141/.089

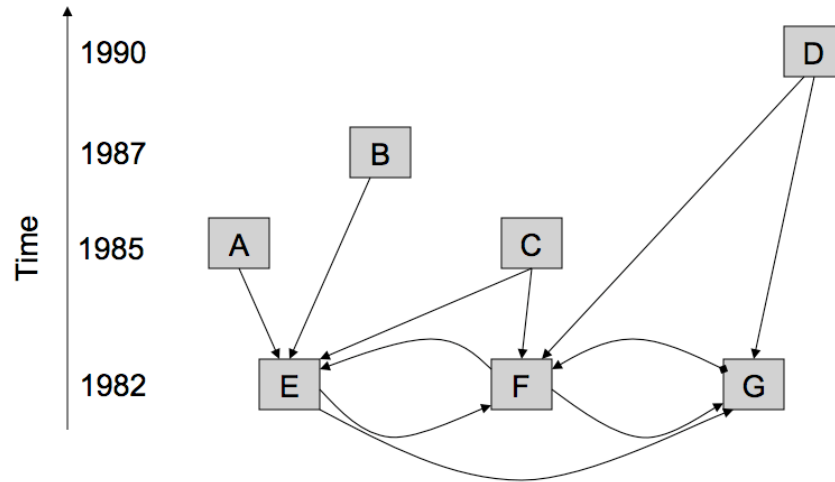
31046 observations on 6316 papers

Standard errors in parentheses

<sup>†</sup> significant at 10%; \* significant at 5%; \*\* significant at 1%

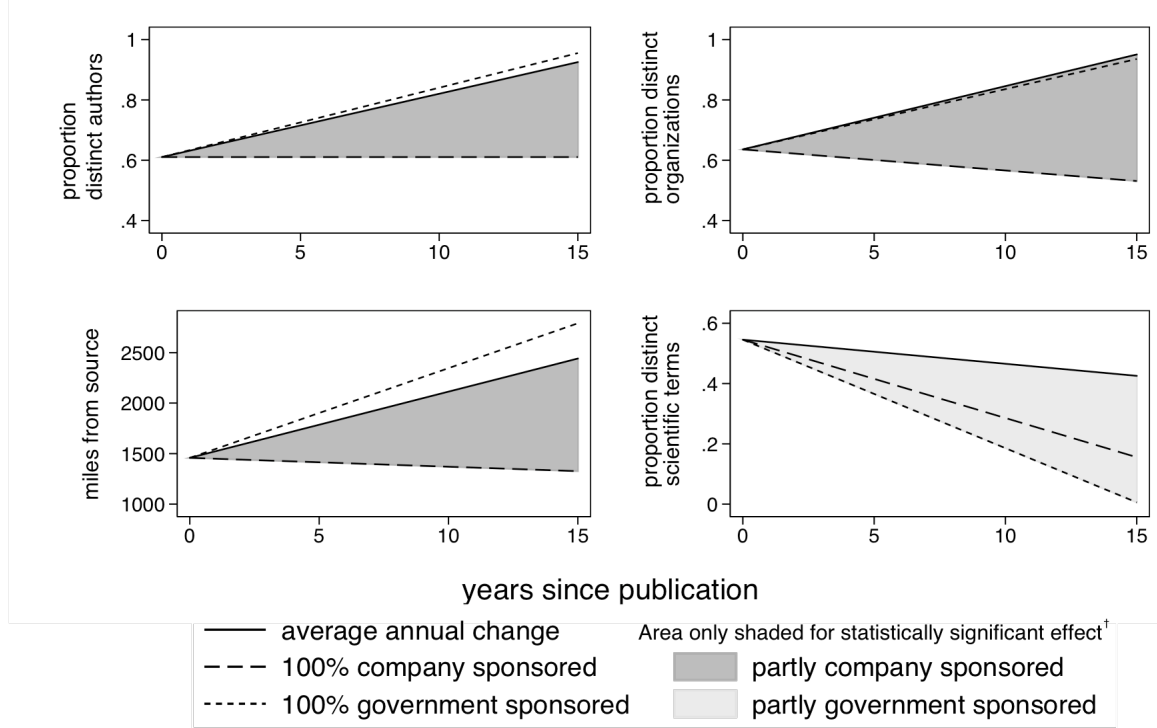
## Figure 1. Manuscript Sharing

Measured by the number of articles citing a scientist's articles in citation cycles



Cycles: E=1; F=2; G=2

Figure 2. Dissemination of ideas (inferred by citations) from scientific articles sponsored by industry and government



<sup>†</sup> In the author, organization, and geographic distance models, only company sponsorship was significant, and in the scientific term model, only government sponsorship was significant, corresponding to the shading in the figure.

---

## Notes

I am grateful for the financial support provided by the NSF (grant #0242971 to Walter Powell and James Evans), the Social Sciences Research Council (Program on the Corporation as a Social Institution and Program on Philanthropy and the Nonprofit Sector), the MacArthur Foundation (Fellowship at the Stanford Center for International Security and Cooperation), and the U.S. Department of Education (Jacob K. Javits Fellowship). I am also grateful for data provided by Thompson Scientific, The *Arabidopsis* Information Resource and Los Alamos National Lab Library. I have benefited from greatly from the comments of Toby Stuart and from participants in the University of Chicago Conference on the Emergence of Organization and the Harvard Organizational Behavior seminar and from two anonymous reviewers at *SSS*.

<sup>i</sup> In corresponding with the Swiss mathematician and alchemist Nicholas Fatio de Duillier in 1689, Newton observed of Boyle, the experimentalist and bankroller of the first scientific society and journal—the Royal Society and its *Philosophical Transactions*, ‘Mr Boyle has divers times offered to communicate & correspond with me in these matters but I ever declined it because of his conversing with all sorts of people & being in my opinion too open & too desirous of fame’ (Westfall 1980: pp. 492-3) Though willing to countenance alchemical research, Boyle also was willing to inappropriately leak it to others.

<sup>ii</sup> ‘Naturalists, who spake their words in hid language, should they have spoken out plainly, they would have done very ill for divers reason, for all men would have used this art and the whole world would have been spoiled and all agriculture perisht ...’ (Arnoldus de Nova Villa, circa 1300 in *De secretis naturae*, printed in 1611).

<sup>iii</sup> Boyle’s *Sceptical Chymist* (1661), a dialogue between chemist, Aristotelian and alchemist before the Royal Society, concludes with the chemist winning the debate, criticizing confidence in the four-element system of the Aristotelians and the three-principle system of the alchemists. Boyle’s unpublished appendix to the work allows the alchemist (Eleutherius) to have the last word and the chemist (Carneades) to concede it.

<sup>iv</sup> See also Campbell et al. (2002) for survey results that suggest the competitive secrecy of academic labs; also see Merton’s own discussion of counter-norms in science (Merton 1976).

<sup>v</sup> See also Liebeskind (2000) for a suggestive depiction of how high technology industries prefer secrecy over patents.

<sup>vi</sup> The notion that basic research drives technology but should be insulated from it did not originate with Bush, nor was it fully developed in his report (Godin 2006), but is often attributed to him because his was one of the most public and politically influential statements of it.

<sup>vii</sup> Collins demonstrates that experimental replication is difficult under the best of circumstances—when materials are widely available—because tacit knowledge on which experiments are built may not be widely shared (Collins 1974) and because the failure of complex experiments does not distinguish between flaws in the tested theory or the testing experiment (Collins 1985).

<sup>viii</sup> When I asked one postdoc the reason for her move from *C. elegans* research at Colorado to *Arabidopsis* at Stanford, she replied, ‘Because it was getting so competitive there ... there is *much* more sharing in the *Arabidopsis* community.’ By contrast, scientists entering *Arabidopsis* from the hinterlands of plant biology—e.g., using fern or *Medicago*—find *Arabidopsis* both larger and more competitive.

<sup>ix</sup> Codes uniquely identify each interview: the first character represents the informant’s training (B=biology); the second represents current status (S=student; I=investigator; M=manager); and the third and fourth represent an interview-specific unique identifier. All interviews were performed between September 2003 and April 2005.

<sup>x</sup> These cycles initially included all distinct citing papers, but when I reanalyze the data using only article citations from research groups that did not share an author with the cited paper, the effects were slightly stronger. See the <sup>‡</sup> note below Table 3.

<sup>xi</sup> Data for all of the publications cited by patents after 1992 were purchased from CHI Research, Inc.

<sup>xii</sup> Interviews confirmed that the subfields flagged as more and less commercial correspond with assessments of firms.

<sup>xiii</sup> Pairwise correlations between these variables are available from the author.

<sup>xiv</sup> This assumes that all other covariates are held at their means.

<sup>xv</sup> These estimates are not reported here, but are available from the author.

<sup>xvi</sup> See note xxiii.

<sup>xvii</sup> In research areas closer to application, industry funding may have the opposite influence, but with similar effect. If industry pushes scientists toward increasingly applied research, they may move away from the general, system-crossing findings that appeal to academics.

<sup>xviii</sup> SciSearch is produced by ThompsonReuters’ Institute for Scientific Information (ISI), which generously gave me use of the entire Science Citation Index for the purpose of this project.

<sup>xix</sup> Codes for molecular function, biological process and cellular component (13,976), which were used to code the *Arabidopsis* genes, were developed in a joint effort among curators of several model organisms to characterize genes independently of their host-organisms, entitled the Gene Ontology (GO) project. Development and anatomical annotations were developed by TAIR specifically for *Arabidopsis*. Curators from TAIR with Ph.D.s in molecular biology hand-coded all *Arabidopsis* genes in the publications. I contributed by using BLAST (Basic Local Alignment Search Tool)-like algorithms and more linguistically sophisticated parsing approaches (Chang, et al. 2003) to match all terms into article titles and abstracts. This process resulted in 28,350 unique terms linked by nearly 400,000 publication-term matches to *Arabidopsis* articles.

---

<sup>xx</sup> The NGA and US BGN data contain the names of over three million city and populated regions from around the world, including virtually every small city and many towns; 20,000 features of the database are updated monthly. More than 99% of the institutions in the SciSearch database had a geographic match with a city or region feature of this database with a confidence of near certainty.

<sup>xxi</sup> Alternately, the distance between the geographic centroid of the two institution sets was measured, which produced similar results in the final analysis.

<sup>xxii</sup> For materials sharing, which has a higher dispersion than manuscript sharing, negative binomial models were also performed with a similar pattern of results.

<sup>xxiii</sup> For this type of estimation to be valid, it must be the case that scientists with and without industry funding would have followed parallel paths in the absence of industry sponsorship. In an attempt to correct for violations of this assumption, I produced consistent findings combining both fixed and random effect specification with Inverse Probability of Treatment Weighted (IPTW) estimation (Robins et al., 2000; Hernán et al., 2001; Azoulay et al., 2006). In these results, the negative effect of industry ties on scientists' acknowledgement mentions, for example, is somewhat attenuated, but remains negative, and all of the industry funding and collaboration variables post collective negative significant ( $p < 0.1$ ). This suggests that some, but not all, of the individual-level industry effects listed in Table 6 could be attributable to the differential propensity for certain types of scientists to enter relations with industry.