Essays on the Emergence and Diffusion of Breakthroughs

A dissertation presented

by

Sen Chai

to

the Technology and Operations Management Unit

at the Harvard Business School

in partial fulfillment of the requirements
for the degree of
Doctor of Business Administration
in the subject of
Technology and Operations Management

Harvard University
Cambridge, Massachusetts
May 2013

Copyright

Thesis advisor: Lee Fleming Sen Chai

Essays on the Emergence and Diffusion of Breakthroughs

Abstract

This dissertation deals with the emergence and diffusion of creative breakthroughs. The first two chapters concentrate on breakthrough emergence to assess and expand extant theories of sources of breakthroughs, and employ a hybrid methodology. In particular, I study the discovery of RNA interference (RNAi) in the life sciences. By employing regressions, I find that the predictive power of current theories altogether is quite low: ranging from less than 1% for the Nobel Prize to 13% for productivity. These results prompted fieldwork and the use of interviews with scientist informants to address the gap and gain a deeper understanding of the phenomenon. My findings show that the seminal discovery was missed several times not only due to difficulties in solving a particular problem but also due to failures of identifying breakthrough opportunities and proposing them. I suggest a cognitive framework with institutional underpinnings at the basis of this failure stemming from three barriers: framing barriers, paradigmatic pressures and boundary barriers. In the problem identification stage, path dependence from established technologies and the quest toward normal science blinded scientists from recognizing a prospective breakthrough as they framed RNAi as a tool while ignoring its scientific merit for inquiry. In the problem-solving stage, scientists suffered from the socio-cognitive barrier of being constrained by current dogma. Due to reticence in challenging the dogma, they hesitated to propose solutions that significantly strayed away from the confines of established theory. Moreover, existing boundary barriers between communities of

scientists intensified the effect of barriers in both stages. It prevented recognition of links between several prior instances of odd observations thus heightening the difficulty in identifying the breakthrough by misrepresenting the problem's magnitude. While solving the problem, similar anti-dogmatic results stayed isolated and unsubstantiated which diminished confidence in proposing a radically new paradigm. The third chapter explores diffusion of discoveries beyond the boundary of the scientific institution. It focuses on academic-industry collaborations and assesses the effect of a mediated funding scheme on innovative performance – quantity, impact and collaborative nature of patents and papers – by comparing funded and unfunded firms.

Table of Contents

Dedications

To my beloved parents

Acknowledgements

Foremost, I would like to express my earnest gratitude to my committee chair, Prof. Lee Fleming, for his continued support of my doctoral studies and research, for his mentorship, encouragement, enthusiasm, and immense knowledge. His guidance and care helped me throughout my research and writing. I would also like to express my genuine appreciation to my committee members, Prof. Vicki Sato for keeping me grounded on the real world implications of my research and providing me first-hand opportunities to discover them, as well as Prof. Gary Pisano and Prof. Fiona Murray, for their motivation, insightful comments, and hard questions.

My sincere thanks goes to Prof. Willy Shih for providing indispensable data for this dissertation and industrial learning experiences, Prof. Richard Freeman for his honest and invaluable feedback and mentorship, Prof. Jan Hammond for teaching opportunities, Prof. Michael Toffel as well as members of my field exam committee Prof. Michael Tushman and Prof. Karim Lakhani. I also would like to thank Ronald Lai, Amy Yu, Alexander D'Amour and Edward Ye Sun for their excellent research assistance, Bill Simpson, Xiang Ao and Andrew Marder for their invaluable help with econometrics, and the doctoral office, Jennifer Muccarione, John Korn, Dianne Le, Janice McCormick, Marais Young and LuAnn Langan for supporting me throughout my time at HBS.

I am grateful for Prof. Janet Henderson for instilling in me a first taste for research while I interned in her lab in high school and throughout my undergraduate studies at McGill University, as well as Prof. Milica Popovich at McGill University and Prof. Jan Pietzsch at Stanford University for providing me further opportunities to develop my research skills.

I thank participants in the TOM DBA seminar, the CRAFT seminar and the Fleming lab group at UC Berkeley for their helpful comments on my research work. In particular, I thank Elizabeth Altman, Anil Doshi, Yun Jing Rose Gao, Sujin Jang, Chonnikarn Fern Jira, Lisa Kwan, Clarence Lee, Jeffrey Lee, Matthew Lee, Maria Loumioti, Hila Lifshitz-Assaf, Anoop Menon, Frank Nagle, Patricia Satterstrom, Luciana Silvestri, Christopher Small, Saundra Small, Haris Tabakovic, Chia-Jung Tsay, Tiona Žužul and my officemates, Jillian Berry and Melissa Valentine.

Enfin, je remercie ma famille: mes grands-parents ainsi que mes parents, Yihong Fu et Yuren Chai, pour leurs sacrifices, pour m'avoir donné toutes les opportunités possibles afin de réussir et pour avoir cru en moi depuis le début.

This page is intentionally left blank.

I. Introduction

While a considerable body of work has explored breakthroughs from historical and bibliometric standpoints, we still have limited understanding on how they emerge and diffuse. Consequently, this dissertation centers around gaining a deeper appreciation of this creative knowledge production process, more specifically, on how individual scientists and inventors discover or invent breakthroughs, and also on the counterfactual of what hinders discovery. In the next two chapters, I focus on breakthrough emergence and employ a hybrid methodology to assess and expand extant theories on sources of breakthrough in a specific scientific discovery in biotechnology, RNA interference. By employing quantitative regressions, I find that the combined predictive power of current theories is quite low. These results prompted me to go into the field, to use qualitative interviews of scientists to address the gap and gain an inductive understanding on the phenomenon of breakthrough emergence. The final chapter moves away from emergence and explores the diffusion and commercialization of scientific discoveries beyond the boundary of the scientific institution through a specific mechanism of academic-industry partnerships.

Chapter II in this dissertation (first paper), co-authored with Lee Fleming, builds on a large body of work that has correlated bibliometric measures from papers or patents to subsequent success, typically measured as the number of publications or citations, following widespread availability of computerized databases. We ask two simple questions: given available bibliometric knowledge at any point in time, how accurately can we predict who will discover a future breakthrough? Moreover, given numerous hypotheses from the literature, which factors should one focus on when predicting breakthroughs? After reviewing and synthesizing the (often competing) predictions from the literatures, we collectively test those hypotheses based on available data in the year before RNA interference was discovered. We operationalize breakthrough from the most stringent definition of authoring the Nobel winning paper and gradually relax it to an indicator of being in the elite (the top ten percent of citations), forward citation counts, and finally publication counts. Predictive power of current theories ranges from less than 1% for the Nobel Prize to 13% for productivity. Including prior publications and citations increases the latter number to 49%. We conclude with an agenda for future progress in the bibliometric study of creativity.

Chapter III (second paper) builds onto the findings in the initial quantitative work that there is a significant gap in bibliometric papers identifying sources of breakthrough that remains unexplained and answers the questions: Why are some scientists more successful than others at discovering breakthroughs? Why do people on the verge of breakthrough miss them? By interviewing scientists with breakthrough potential in a case historical analysis of a groundbreaking discovery in biology, RNA interference (RNAi), my findings show that the seminal discovery was missed several times not only due to difficulties in solving a particular problem as stipulated in current literature but also due to failures to identify breakthrough opportunities. I propose a cognitive framework with institutional underpinnings at the basis of this failure stemming from three barriers: framing barriers, boundary barriers and paradigmatic pressures. In the problem identification stage, path dependence from established technologies and the quest toward normal science blinded scientists from recognizing a breakthrough potential. Instead, they framed RNAi as a tool while ignoring it as a scientific concept worthy of study. Existing boundary barriers between communities of scientists aggravated this difficulty in identifying the breakthrough opportunity by misrepresenting the magnitude of the problem as it prevented recognition of links between several isolated prior instances of odd observations. In the problem-solving stage, scientists suffered from the socio-cognitive barrier of being constrained by current dogma. To avoid being wrong, they hesitated to propose solutions that significantly strayed away from the confines of established theory. Coupled with boundary barriers, similar anti-dogmatic observations stayed isolated and unsubstantiated, thus diminishing the confidence to identify a new revolutionary paradigm. Scientists offered remedial practices to circumvent these barriers from which I operationalize new sources of breakthrough emergence lacking in traditional measures of bibliometrics.

In chapter IV (third paper), co-authored with Willy Shih, I shift my focus from emergence to diffusion of scientific discoveries. Scientific research and its translation into commercialized technology is a driver of wealth creation and economic growth. Partnerships between public research organizations, such as universities and hospitals, and private firms are an established policy tool to foster the translation of basic science into commercial applications that has attracted increased interest. Yet questions about efficacy and the efficiency with which funds are used are subject of frequent debate. This final chapter examines empirical data from the Danish National Advanced Technology Foundation (DNATF), an agency that funds partnerships between universities and private companies to develop technologies important to Danish industry. We assess the effect of a particular mediated funding scheme which combines project grants with active facilitation and conflict management on innovative performance – quantity, impact and collaborative nature of patents and papers – by comparing funded and unfunded firms. Because randomization of the sample was not feasible, we address endogeneity around selection bias using a sample of qualitatively similar firms based on a funding decision score. This allows us to observe the local effect of samples in which we drop the best recipients and the worst non-recipients.

Going forward there are several possible extensions that can be built onto each chapter of this dissertation. From the second chapter where the lack of causal identification was identified as a major setback in the extant innovations literature, a follow-on project exploits the surprising and unexpected award of the 2006 Nobel Prize in Physiology and Medicine to the two discoverers of RNA interference as a quasi-natural experiment to study the impact of such high-profile awards on the patenting, licensing and commercialization trends in that particular technology sector. Especially with the recent establishment of several high-value prizes in physics (Fundamental Physics Prize) and the life sciences (Breakthrough Prize in Life Sciences), knowing their effects on subsequent knowledge creation is imperative. Similarly, this opportunity also affords the possibility to understand how these prizes change public interest in a field of technology.

The qualitative chapter on breakthrough emergence also provides many opportunities to further my work. For instance, although the role of conferences is multi-faceted, very few prior works have investigated the effect of conferences on subsequent collaborative behavior of attendees nor explored the impact of the collaborative outputs. Thus, we possess limited empirical evidence on how conferences foster collaboration and productivity, and know little about who are most likely to collaborate together. Not only do conferences facilitate information diffusion, they are also employed as a validation mechanism by scientists to substantiate the soundness of abnormal and unexpected results. Moreover, the physically and temporally condensed structure of conferences also facilitates rapport and network building amongst attendees, and acts as a platform that fosters new collaborative opportunities.

And finally, the results in the final chapter are the combined effect of mediation and funding. Continued research is needed to tease apart these effects, to pinpoint which characteristics and dimensions of these novel actively facilitated programs enhance collaboration across organizational and institutional boundaries, and how they affect the project’s final outcome independent of purely monetary provisions.

II. Predicting Breakthroughs with Bibliometric Measures

i. Introduction

Since Schumpeter's observation (1942), the creative destruction of scientific and technological breakthroughs has held a central and recurrent prominence in the innovation literatures. The topic remains hugely important, as breakthroughs wreck havoc with extant industries and regions and at the same time, provide the renewing impetus for new industries and whole economies. Scholars of innovation have put forth several, sometimes conflicting, hypotheses identifying the sources of such breakthroughs. For instance, being more productive increases the number of creative trials thereby improving chances for breakthrough discovery (Simonton, 1999), whereas being less productive may also improve those chances by focusing and pursuing anomalies. Collaboration might increase the chances of breakthroughs (Singh & Fleming, 2010; Wuchty, Jones, & Uzzi, 2007), though working individually at some points in the process also appears beneficial (Girotra, Terwiesch, & Ulrich, 2010). Social brokers – those who are the sole connections between others – have been argued to be more (Burt, 2004) creative, less creative (Obstfeld, 2005; Uzzi, 1997), and more creative in particular circumstances and also hampered in their ability to diffuse their idea (Fleming et al. 2007). Individuals at the core of a community are a more likely source because they enjoy enhanced information and resource access from social ties (Collins, 1998; Gieryn & Hirsh, 1983); or at the periphery because they are not constrained by current approaches (Jeppesen & Lakhani, 2010). Specialists with deep technical knowledge are better equipped to see beyond the frontier, as opposed to generalists who can bring together disparate components (Dougherty, 1992; Leonard-Barton & Swap, 1999). Individuals realize breakthroughs earlier in their careers, because they are not constrained by the thinking of their field (Simonton, 1989) or later, because they must work through the accumulation of knowledge (Jones, 2009). Better scientists might prefer to stay in academia as they value the freedom in choosing their research direction (Stern, 2004), yet some corporate labs also do fundamental breakthrough work. Affiliation with a prestigious institution should increase breakthrough potential because of higher human capital and exposure to better ideas. And finally, mobility between multiple affiliations increases exposure to a greater diversity of ideas, but is associated with high setup costs and may also be an indicator of failed tenure attempts (McEvily & Zaheer, 1999).

Surprisingly little work has applied these ideas to predict future sources of creativity. But predicting such sources for various fields is important not only from a policy standpoint to inform public investment in science but also for managers and analysts in helping to identify key scientists. This motivates simple questions: given available bibliometric information in one year, along with the combined theories of the innovation literatures, how accurately can we predict the sources of breakthroughs in the following year? In particular, who is most likely in a field to publish, publish highly cited work, publish a highly cited outlier, even a Nobel prize-winning paper? Moreover, given the numerous hypotheses from the literature elaborated above, which factors should one focus on when predicting breakthroughs? After clarifying the definition of breakthrough used herein, we review the literatures on creativity and breakthrough, synthesize predictions from these literatures, compile a dataset on the discovery of RNA interference a breakthrough in molecular biology, and test the predictive power of our current theories. Combining all current bibliometric theories on creativity – lone scientists vs. teams, brokerage vs. cohesion, periphery vs. core, specialist vs. generalist, experience vs. youth, affiliation type, affiliation prestige and mobility – can explain only 0.2% of the variance in predicting the Nobel paper, 5.3% for authoring a paper in the top 10% of citations, 13.1%

for citations, and 13.4% for publications. If prior publications and citations are included in the models, these numbers increase to 0.8%, 20%, 37.8%, and 49.6%, respectively. We close with a discussion of how we might increase our rate of progress in understanding the sources of scientific and technical breakthroughs.

ii. Sources of Scientific and Technological Breakthroughs

Before trying to predict who discovers a breakthrough, one must define the breakthrough and those at risk of discovering it. We adopt Simonton's (1999) notion of impact that encompasses both creative novelty and success and implement the empirical convention of citations from future scientific publications. Our setting to study RNA interference (RNAi) fits within this depiction of breakthrough because of its broad research impact and therapeutic potential. The field generated several prestigious awards such as the Lasker Award for Basic Medical Research in 2008 and the Nobel Prize in Physiology and Medicine in 2006. Furthermore, small interfering RNA, a critical component involved in the RNAi pathway, was also named breakthrough of the year in 2002 by Science (Couzin, Enserink, & Service, 2002). While communities, organizations or scientists might be at risk of a breakthrough, we restrict our analysis to individual scientists within a single community of scientific researchers. We defer the very important question of how to define such a community to our methods section below, and begin by reviewing and synthesizing the literatures. Our intent is not to generate novel theory but rather comprehensiveness; we minimize comment on evidence that supports or refutes these theories in the interest of brevity.

Publication history and eminence – Prior history and eminence are important factors in determining sources of breakthroughs. In a sense, this controls for unobserved heterogeneity in the research ability of scientists using prior publications. While prior prominence is a well-accepted indicator of a scientist's research capabilities, the relationship between quality and quantity is more controversial. Limited time and resources imply a tradeoff between quality versus quantity. An obsession with quantity can cause researchers to miss breakthrough cues as they swiftly plow through their originally designed experiments. Thus the more productive scientists might make incremental rather than radical discoveries. On the other hand, since creative processes usually require many attempts to obtain a success, let alone a path-breaking success, the more productive a person the more trials they create and consequently the more likely their breakthrough potential. To use a sports analogy, the more at-bat opportunities available the higher the probability of hitting a homerun. We summarize this accordingly as: *Greater productivity increases the chances of a breakthrough by increasing the number of creative draws, whereas less productivity increases the chances by enabling focus and pursuit of anomalies.*

Collaborative vs. Individual researchers – Recent studies show a continuing and increasing trend for teams to contribute to the production of knowledge through paper and patent publications in all natural and social science domains (Wuchty et al., 2007). Alluding to the burden of knowledge theory, to compensate for an ever increasing body of knowledge inventors and scientists have to narrow their expertise (Jones, 2009), which translates to reduced individual capabilities and forces innovators to work more predominantly in teams. Furthermore, collaboration fosters breakthrough emergence as circling ideas for critique by co-inventors decreases the likelihood of poor outcomes, while multiple collaborators permits the recombination of more diverse components (Singh & Fleming, 2010).

Conversely, proponents of the lone superstar have argued that even though teams bring greater collective knowledge and effort, there remain significant costs to increased teamwork such as coordination losses (McFadyen & Cannella, 2004) and groupthink (Janis, 1971). Therefore a shift to teamwork may be a costly phenomenon that promotes low-impact science. Girotra, Terwiesch, & Ulrich (2010) provide a resolution of the controversy and demonstrate benefits to alternating between solitary and collaborative work (though such alternating is difficult to capture with bibliometric data). In short, the above theoretical arguments can be recapped as: *Collaboration increases the chances of breakthrough because it increases the diversity of search and efficiency of idea selection, whereas solitary work increases chances because it minimizes idea suppression and social loafing.*

Brokerage vs. Cohesion – Context and in particular, social structure, has long been thought to influence creativity, flow of knowledge and ideas. The optimal structures, however, remain an open theoretical and empirical question. The notion of brokerage or structural holes, implicit in Granovetter's paper on the strength of weak ties (1973), is a structure in which an individual is directly connected to collaborators who are themselves not connected. Brokerage has been argued to enhance innovative creativity and output given that brokers occupy a nexus position in which they have first access to diverse information. Assuming recombinant search as the process of innovation (Schumpeter 1939, Henderson and Clark 1990) brokers have control over these distinct pieces of information and are thus provided the best opportunity to generate new knowledge combinations (Burt, 2004).

In contrast, Coleman's model of social capital (Coleman, 1988) argues for the benefits of cohesion, with the increase of trust, redundant information paths that facilitate tacit knowledge transfer, shared risk taking, and easier mobilization (Obstfeld, 2005; Uzzi, 1997). A cohesive structure facilitates distributed understanding of all components of the new knowledge, fosters a greater sense of mutual ownership of the creative product and increases the likeliness of the creation from being used again (Fleming, Mingo, & Chen, 2007; Reagans & McEvily, 2003). These conflicting theories can be summarized as: *Brokerage increases the chances of breakthrough because it enables first access to and control of information, whereas cohesion increases the chances because it increases trust which allows richer lateral diffusion of information.*

Core vs. Periphery – The sociology of science literature supports a view in which the most successful problem-solvers may not necessarily lie at the core of the problem field. The main theoretical reasoning behind this line of work is that scientists situated at the periphery of their community possess focused naïveté – a useful ignorance of prevailing assumptions and theories. They draw from different knowledge pools than the actors at the core which translates into diverging perspectives and ultimately helps them in uncovering potentially novel and highly impactful breakthroughs (Jeppesen & Lakhani, 2010). These arguments are echoed in the organizational literature on innovation, which argues that breakthroughs come from outside an extant industry (Tushman and Anderson, 1986).

An opposing viewpoint believes that individuals situated at the core are social elites who benefit from better availability of resources and established relationships (Collins, 1998; Gieryn & Hirsh, 1983; Merton, 1949). Viewing knowledge creation as a recombinant search process, core scientists have better access to relevant information, more resources, and are less isolated, which in turn increases their likelihood of creating breakthrough work. Consequently, *a core position increases the chances of breakthrough because it provides better access to information and resources, and a peripheral position increases the chances because of useful ignorance of prevailing assumptions and theories.*

Specialist vs. Generalist – People that focus become specialists in their area of expertise, at the expense of breadth. Whether specialization enables breakthrough remains an open question. Specialists may be better positioned to solve a breakthrough because their deep knowledge in a field enables them to optimally evaluate and combine components at their disposal. They can predict outcomes better, due to their deep reservoir of experience.

According to advocates of marginality, however, specialists are deeply rooted in their respective scientific domains and may suffer from a curse of knowledge that limits exploration beyond their immediate knowledge neighborhoods. Generalists are not bound to the current thinking in the focal field and can therefore offer different perspectives and heuristics that enables them to sample a larger search space thus drastically increasing the probability of discovering a new and fruitful combination (Dougherty, 1992; Leonard-Barton & Swap, 1999). To summarize, *being a generalist increases the chances of breakthrough because a broader variety of components can be recombined, whereas being a specialist increases the chances because deeper understanding of the components enables more accurate prediction (and more effective winnowing of probably useless combinations).*

Lifecycle – On the one hand, the burden of knowledge literature (Jones, 2009; Wuchty et al., 2007) is based on the observation that innovators are not born at the cutting edge frontier of knowledge and must undertake significant education. Furthermore, significant increases in the total stock of knowledge over the past few centuries imply that the amount of education innovators must accumulate also increases proportionally. The implication of more learning is a delayed contribution to the stock of knowledge thus pushing back the average age of contribution (Jones, 2009).

On the other hand from a cognitive viewpoint, Simonton has studied the relationship between age and creativity in numerous artistic and scientific fields (Simonton, 1989). Although fields differ significantly across optimal creative age, younger scientists were found not to be afraid to tackle hard problems, and are not encultured with conventional wisdom. They have had less time to socialize into the norms of established institutions and can therefore freely think outside the box thereby increasing their propensity of generating breakthroughs. *Consequently, youth increases the chances of a breakthrough because it is not weighed down by established beliefs, whereas experience increases the chances because it enables contributions at the frontier of science.* Taking both sides of the argument into consideration, we can also posit a curvilinear inverted-U relationship where relatively junior scientists with some experience but not completely newcomers are more likely to discover breakthroughs. They have had enough time to surmount the burden of knowledge and but are still relatively new to the field.

Organizational affiliation – Whether a researcher works in academia or corporations affects the impact of scientific publications differently. Due to the institutional priority-based rewards system in science, higher-quality researchers may be willing to tradeoff more income in private firms to earn the higher expected prestige rewards in academia (Stern, 2004), especially when they are given the freedom and authority to direct their own research agendas into areas that they perceive as high-risk breakthrough areas. Because researchers in academia are allowed more flexibility in pursuing their individual research agendas than in for-profit organizations, higher-quality scientists tend to choose academia over private corporations. *Consequently, academic affiliation increases the chances of a breakthrough because of sorting of higher quality human capital.*

Prestige – Similarly, the ranking of institutions in which scientists have been affiliated with is also an indicator of the breakthrough potential of an individual. Most university rankings are based on several criteria, one of which being the quality of research publications it produces. More prestigious institutions have stricter selection processes for faculty and students, and are also better positioned to learn about recent research results, through seminars and other modes of scientific communication. *Affiliation with a prestigious institution increases the chances of a breakthrough because of higher human capital and exposure to better ideas.*

Mobility – Compared to those with low organizational mobility, as individuals move from one institution to the next they are exposed to more heterogeneous ideas, perspectives, assumptions, problem-solving techniques and thought processes (McEvily & Zaheer, 1999). However, there are significant setup costs associated with each move, and specifically in our academic setting moves may be an indication of failed attempts at obtaining tenure. *Mobility increases the chances of a breakthrough because of exposure to a greater diversity of ideas, whereas staying put increases the chances because it minimizes setup costs and indicates successful tenure application.*

iii. Methodology and Data

Setting

Our setting is a breakthrough in the biological sciences, RNA interference. It is a naturally occurring endogenous mechanism activated by double-stranded RNA (dsRNA) precursors that induce the silencing of specific genes. The phenomenon was initially observed by plant biologists in the early 1990s where an attempt to transgenically alter color pigmentation in petunia plants yielded unexpected outcomes. The trigger mechanism to this phenomenon was discovered in 1998 by Andrew Fire and Craig C. Mello (1998) for which the two scientists were awarded the Nobel Prize in Physiology and Medicine in 2006.

RNAi is not only valuable as a research tool; it also opened the possibility for a whole new class of drugs in biotechnology. For instance in research, the selective and robust effect of RNAi can induce suppression of specific genes of interest both in vitro and in vivo that can be applied to large-scale screenings that systematically shut down each gene in the cell to identify components necessary for a particular cellular process or event. In drug development, the RNA interference pathway can be conceivably used to treat genetically based diseases by turning off, for example, Huntington's disease or certain genetically based cancers.

Sampling and Identifying a Community of Scientists

We must first define a scientific community, in other words identify the risk set of scientists, before answering whom in that community is most likely to discover a breakthrough. How a community is defined is crucial to understanding how scientific breakthroughs arise within it. Despite a fairly advanced literature on community detection, spearheaded by network physicists and applied mathematicians as evidenced by the extensiveness of reviews available in the literature (Fortunato, 2010), several unique and defining characteristics of our data and study still make it difficult to detect communities in a straight-forward fashion and, consequently, expose the limitations of current available methods. Indeed, one important drawback of these structurally detected communities, which follows from the starting definition of a community consisting of cohesive group of nodes or links, is that all members of a given community must be connected to one another. "Sub" communities can be identified within a larger connected component, however, the analysis begins with – and assumes – a connected component. Thus, these methods inherently preclude communities that share similar functional attributes yet have members who are unconnected. For instance, in the case of RNAi scientists, given our depiction of scientific collaboration in which individual scientists are connected by their co-authorship relationships, even though two scientists do not necessarily have a collaborative relationship reflected through co-authored publishing both take part in the same community that work in advancing understanding of RNAi. Consequently, we envision the community of RNAi scientists to be depicted as a collaborative network with several unconnected components. At this point, unfortunately, most theoretical research into community detection techniques have solely focused on structurally identifying communities, promptly ignoring nodal characteristics that define functional communities.

To incorporate functional characteristics in our definition of the RNAi community, we include content search of titles, abstracts and Medical Subject Headings (MeSH) keywords. Since our study looks to predict a future discovery event it is centered on the period prior to the breakthrough where a defined community of RNAi researchers has yet to emerge. It is, therefore, not surprising that MeSH keywords such as "RNA, Interference", and the same phenomenon in plants and fungi – respectively named "co-suppression" and "quelling" – did not enter the MeSH lexicon until 2002. To circumvent this issue, we review archival documents on the history of RNAi including the Nobel lectures, and find that scientists were attempting to explain gene expression regulation or gene silencing by experimenting with both dsRNA and antisense RNA as causal agents. Furthermore, they believed in the hypothesis that RNA plays a central role in gene silencing mechanisms which rests on the premise that RNA molecules are not only restricted to the passive role of carrying genetic information but also possess catalytic functions. Consequently, we define the community of researchers at risk of discovering a breakthrough from their published peer-reviewed articles using the MeSH search terms[1] "RNA, Double-Stranded", "RNA, Antisense", "RNA, Catalytic", "Gene Silencing" and "Gene Expression Regulation" in PubMed. We primarily make use of MeSH keywords to search for published papers since MeSH

[1] The exact search string used in PubMed query extracted on October 26, 2011: ((((gene silencing[MeSH Terms] OR gene expression regulation[MeSH Terms]) AND (RNA, double-stranded[MeSH Terms] OR rna, antisense[MeSH Terms] OR rna, catalytic[MeSH Terms])) AND "1980"[Publication Date] : "1999"[Publication Date]) AND English[Language]) NOT interferon[MeSH Terms]. We also found that dsRNA generated a lot of noise as it was heavily used by immunologists studying interferon responses. To minimize the noise from interferon we include in our MeSH search the "NOT interferon[MeSH Terms]" term.

keywords are believed to be a relatively objective classification scheme.[2] We augment the MeSH keyword searches with a title and abstract search so as to include scientists who initially observed the RNAi phenomenon in plants and fungi[3]. We include in this community papers that were published until 1999, as we believe that those who quickly published following the 1998 breakthrough paper are also in the risk set for breakthrough discovery. By extracting unique authors from the set of papers obtained above, we effectively identify a community of RNAi scientists topically based on their publication focus. This sample of scientists defining the pre-breakthrough RNAi community yields 1,551 papers and 3,959 unique authors. However, we are missing affiliation data for 49 of these individuals, so the majority of our regressions are run with a sample of 3,910 authors.

Out of the 3,959 unique authors present in our sample, 144 authors are completely new to the defined field of RNAi in 1998. These scientists do not have any prior publications either within our RNAi community or any other tangential field within the life sciences, in other words they only appear in the MedLine dataset we draw from after 1997. Only one author from these 144 newcomers was affiliated with a non-academic institution. Furthermore, only one newcomer published solely in 1998 while all others collaborated either as first author (n = 57), last author (n = 10) or appeared as a middle author (n = 93). These collaboration structures reflect the apprenticeship model for graduate studies in the life sciences. Because most of the publications in which new scientists partake are co-authored with the principal investigator and other members of a laboratory, the number of

[2] Instead of being assigned by authors themselves, MeSH is a comprehensive controlled vocabulary for the purpose of indexing journal articles and books in the life sciences and also serves as a thesaurus that facilitates searching. It is created and updated by the United States National Library of Medicine (NLM) and used by the MEDLINE/PubMed article database and by NLM's catalog of book holdings

[3] The exact search string used in augmented PubMed query extracted on October 26, 2011: ((((cosuppression[title/abstract] OR co-suppression[title/abstract] OR quelling[title/abstract] OR RNAi[title/abstract] OR RNA interference[title/abstract])) NOT interferons[MeSH Terms]) AND "1980"[Publication Date] : "1999"[Publication Date]) AND English[Language].

new authors getting cited at least once is relatively high (n = 127). Including these newcomers in our sample illustrates the true dynamic nature and evolution of such scientific communities. However, it complicates data collection since we have no historical information on these newcomers prior to 1998 therefore leaving several explanatory variables, such as brokerage and specialist, undefined. Undefined variables for newcomers are set to the mean.

Given the scientists' bibliometric attributes before 1998, we predict who would have a fruitful year in 1998 thereby identifying individual sources of breakthrough. We restrict the dataset used in the empirical analysis to the PubMed Author-ity database (Torvik & Smalheiser, 2009) due to the biological nature of the breakthrough, and organize each data point as unique author records. Moreover, as our quantitative analysis focuses on observing publication performance in 1998, the explanatory variables consist of measures calculated using each author's prior bibliometric data up to 1997 inclusively encompassing all papers available in the PubMed database, while publication data in 1998 are used to calculate outcome variables.

Regression Models

We first attempt to predict the breakthrough itself using rare events logistic (relogit) models on all authors of the Nobel paper because publishing such a paper is an extremely rare event with only six successes out of the entire sample of nearly four thousand scientists. The relogit procedure estimates the same model as a standard logistic regression but estimates are corrected for the bias that occurs when observed events are rare. We then predict using logistic models with cluster robust standard errors scientists who are at the top ten percent of the citation distribution. Finally we use count models with either the forward citation counts of 1998 papers or the number of publications in 1998 as outcome variables to operationalize respectively impact in another way and productivity. The count models are quasi-maximum likelihood Poisson (QML Poisson) with robust standard errors since publications and citations are non-negative counts and over-dispersed which prevents the use of standard Poisson models where it is assumed that the mean and variance of the variable distribution are equal. We also run OLS regression models so as to evaluate the predictive power of our measures of sources of breakthrough.

Dependent Variables

Nobel paper dummy – nobeldum is an indicator that equals one for the six authors on the Nobel winning RNAi paper in 1998.

Top 10% citations dummy – Measures of publication impact based on citations rest on a social definition of creative success, where scientists are only thought to be creative if they receive recognition from their community or society as a whole, and their work is used as a foundation for further advancements (Simonton, 1999). We therefore relax our definition of breakthrough from the Nobel paper to scientists with citations in the top ten percent of the citation distribution with the indicator top10cite. Top10cite_fl is an indicator for the robustness check, where only the scientist's citation count for their first or last author papers in 1998 is in the top ten percent of its distribution (as are all variables with a suffix of _fl). Dependent variables for the top 5% returned similar results.

Number of forward citations for 1998 publications – We further relax our operationalization of breakthrough using forward citation counts garnered until 2010 of 1998 publications (ncite98), which rests on the same premise of social construction of success. For the OLS models we take the natural logarithm plus 1 (lncite98) to match count explanatory variables that underwent the same transformation due to the Poisson models.

Number of 1998 publications – The final dependent variable is a measure of productivity (npub98) depicted by the number of 1998 publications. Similarly, we also take the natural logarithm plus 1 (lnpub98) for the OLS regressions.

Explanatory Variables

Publication history and eminence – Publication history is the count of one's total number of publications since first publishing until the year prior to the 1998 breakthrough (npub97) while publication eminence is the number of aggregated forward citations to these publications (ncite97). When npub97 is zero the scientist does not have any prior publications and is a newcomer. When ncite97 is zero the scientist could either be a newcomer with no prior publications hence no prior citations or could have produced prior publications that have not been subsequently cited. Since we employ the quasi-maximum likelihood Poisson count model in our regressions and both variables are counts, we take their natural logarithm and denote them respectively as lnpub97 and lncite97.

Collaborative vs. Individual researchers – We capture the number of co-authors (ncoauthor) each scientist has collaborated with for all publications prior to 1998 by calculating the degree network measure of each author node – number of directly linked neighboring nodes to a focal node. The network is portrayed by collaborative co-authorship ties for all publications prior to 1998 for researchers within the RNAi community as defined in the prior section. Lone scientists who do not collaborate and newcomers have no co-authors in the period prior to 1998.

Brokerage vs. Cohesion – Using the same network depiction, we measure cohesion (constraint) by calculating Burt's constraint (Burt, 2004). To calculate the constraint, $C_i = \sum_j c_{ij}$ where $c_{ij} = \left(p_{ij} + \sum_k p_{ik}p_{kj}\right)^2$ and p_{ij} is the fraction of *i*'s relation invested in contact j. p_{ij} translates to the degree of *i* if there is no prior weight to the social networks or, in other words, all connections are considered to be equal strength. Since cohesion is undefined for newcomers, we set newcomers' cohesion at the average cohesion value without taking into account newcomers.

Periphery vs. Core – We create two measures of core mirroring the topical and collaborative community. The first measure depicts collaborative core of the scientific community, where core is structurally operationalized by the indicator variable collabcore. We consider scientists situated in the largest connected component of the network of RNAi scientists with the most number of interlinked collaborations to be in the core and hence assign a value of one, while all other scientists including newcomers are considered to be in the periphery of the community and take on the value of zero.

The second measure depicts core versus periphery from a technical standpoint (techcore). Following our topical construction using MeSH keywords of the scientific community working on suppressing gene expression, technical core is calculated by tabulating the frequency of MeSH keywords used in our definition of community "RNA, Double-Stranded", "RNA, Antisense", "Gene Expression Regulation" and "Gene Silencing", "RNA, catalytic" and all previous variants[4] in a scientist's publication history and normalizing by the total frequency of all her MeSH keywords, i.e.

$$\frac{Mesh\ freq\ of\ RNA,dsRNA+RNA,Antisense+Gene\ Expression\ Regulation+Gene\ Silencing+RNA,catalytic}{\sum_i freq\ of\ MeSH_i}.$$

The more a scientist's work is focused in the key antecedent fields to RNA interference as reflected by the frequency in which their published works are classified, the more they are embedded in the technical core of the community. Our dataset provides the top 20 most frequent MeSH keywords per author and so for many scientists in our sample

[4] Prior MeSH keywords for "Gene Expression Regulation" include "Gene Expression", "Genes" and "Phenotype". When tabulating frequency for "Gene Expression Regulation" we also incorporated counts of its prior keywords.

the majority of their work is not in precursor fields to RNAi. Hence those who's top 20 most frequent MeSH keywords do not match none the above five MeSH keywords take on the value of zero for techcore. We tested for the second order effect of core and periphery for evidence of the middle status conformity theory but found not substantiation. Moreover, as 55% of the values of this variable are zero, representing a non-core position, we dichotomize this variable. Any non-zero value of this variable takes on the indicator value.

Specialist vs. Generalist – It can be difficult to disentangle the notions of periphery and core with those of specialist and generalist. Some may even argue for the homology between periphery and generalist, as well as core and specialist. These concepts can be quite different, however. A researcher can be specialist in one field in which they possess deep expertise while simultaneously be at the periphery another. Similarly nothing prevents a generalist to be situated at the core of a given community.

We capture the degree of expertise of each individual scientist using a publication breadth measure implemented based on the breadth of MeSH keywords in a scientist's publications. This metric is a measure of the prominence of high-frequency peaks in the unique list of MeSH keyword distribution associated with every publishing author. We first identity the top most frequent number of MeSH terms for each scientist, k[5]. Again in our case since we have the top 20 MeSH keywords for each author, *k=5*, and we calculate publication depth as the ratio of the frequency sum of the top 2 to 6 most frequent MeSH keywords, i.e. the high frequency peaks, to the sum of the frequency of all MeSH keywords from range 2 to 20, $pubdepth = \frac{sum\ of\ MeSH\ freq\ in\ range\ 2\ to\ k+1}{(sum\ of\ MeSH\ freq\ in\ range\ 2\ to\ k+1)+sum\ of\ remaining\ MeSH\ freq}$ (Swanson, Smalheiser, & Torvik, 2006). According to the measure (pubdepth), a specialist

[5] $k = int(1.7ln(u)+0.5)$ where u=number of unique MeSH for individual i = 20, so k = 5.

with a narrow range of high frequency MeSH keywords has a high value in the numerator, and consequently has higher depth values; whereas a generalist tends to be characterized by a more uniform set of MeSH keyword frequency distribution with higher variance and less defined high-frequency peaks which translates into lower numerator and depth values. Since specialist is undefined for newcomers, we set newcomers' pubdepth at the average value without taking into account newcomers.

Lifecycle – Scientists' experience is proxied by the number of years since one's first publication. Newcomers have zero years of experience while seasoned scientists may have several decades under their belts. Due to the model specification and the count nature of our variable, we take the natural logarithm and denote as lexp. Non-parametric modeling of this variable supports use of the more parsimonious first degree logarithmic of the variable (its effect is increasingly and monotonically negative).

Organizational Affiliation – The proportion of a scientist's academic affiliations is stored in variable academic. Academic equals one for a pure academic scientist and zero for a scientist working strictly in industry. Therefore if a scientist has a total of 3 affiliations, 2 in academia and 1 in industry, their value for academic would be set to 2/3.

Prestige – The prestige (prestige) of a scientist's affiliated institution is a weighted average score with weights assigned according to the top 50 overall research universities as ranked by U.S. News in 1998. The best university is assigned a weighted score of 50, the second best a score of 49 decreasing to a score of 1 for the 50th ranked university, while institutions beyond 50 receive a score of 0. Prestige is calculated as follows $\frac{\sum_{i=0}^{n} \#\, publications_i \cdot university\ score_i}{total\ \#\, publications}$ where *n* is the total number of unique affiliations. For example, if a scientist has a total of 10 publications, 3 of which was published when affiliated with the second best research university as per the US News ranking, 2 of which was published when affiliated with a university ranked 30th and the remaining 5 was published with unranked institutions; her prestige score would amount to $\frac{\#pubs_{2nd\ best} \cdot score_{2nd\ best} + \#pubs_{30th\ best} \cdot score_{30th\ best} + \#pubs_{unranked} \cdot score_{unranked}}{total\ \#\ publications} = \frac{3 \cdot 49 + 2 \cdot 20 + 5 \cdot 0}{10} =$ 18.7. Furthermore, we add an indicator of being affiliated with a top 50 ranked institution at least once (prestiged) so as to correct the skewed distribution with the above weighted measure of prestige. For newcomer both measures of prestige are set to zero.

Mobility – An indicator measures mobility between organizations or institutions where the total number of one's affiliation is greater than one (affil1p). We also include a dummy variable for newcomer (newcomer) equal to one if a scientist appears in our sample only starting in 1998 and zero otherwise.

iv. Results

Table II-1 shows the summary statistics and correlation matrix for all dependent and exploratory variables used throughout the regression analyses. Table II-2 reports results for the four models we evaluate that predict authors of the Nobel winning paper, scientists in the top 10% of the citation distribution, citation count for 1998 papers and publication count in 1998. The rare events logit (King & Zeng, 1999a) provides an estimate that a particular scientist is most likely to discover the Nobel winning breakthrough. Thus in the case for the discovery of RNA interference's trigger our results show that specialized brokers with prior eminence but less publication history prevailed. The model drops academic since it is a perfect predictor of authoring in the Nobel paper. It also drops collabcore, prestige dummy and affil1p because these three variables are perfectly multicollinear to newcomer for the six authors of the Nobel paper. Figure II-1 illustrates effects sizes for all models.

The logit model in Table II-2 assesses the probability of a scientist publishing impactful papers with total citation counts in the top tenth of the distribution. As expected a top ten percent citation scientist is positively and significantly associated with their prior publications' impact as depicted in the form of the number of forward citations for pre-1998 papers. We also find evidence that increased prior productivity and brokerage significantly increase the likelihood of attaining this top tier in citation. We find no significant correlation between the number of co-authors, structural or social core, specialization, or multiple affiliations and the odds of being highly cited. With regard to the lifecycle theme, we observe a linear relationship where younger scientists correlate significantly with high citation likelihood. While the proportion of academic affiliations shows no significant relationship, the prestige of the affiliation is positively but weakly significant. Newcomers positively affect the likelihood of being on top of the citation distribution. Calculating the effect size for variables that are significant, we find that a one standard deviation increase from the mean yield increases of 95.1%[6] for the natural log of prior publication count, 338% for the natural log of prior citation count, and 72.7% for newcomer; and decreases of 24.9% for constraint and 83.1% for natural log of experience to the probability of being in the top 10% of citations. At the ten percent significance level, a one standard deviation increase in prestige increases the probability of being in the top tenth of the citation distribution by 16.4%. In sum, we find that younger scientists with more prior history and eminence, and situated in brokerage positions of their field have a higher probability to be in top tenth of the citation distribution. We also find that being a newcomer contributes positively.

[6] Effect size = $\frac{1+e^{-(\propto+\beta_i \cdot \mu_i)}}{1+e^{-(\propto+\beta_i \cdot (\mu_i+\sigma_i))}} - 1 = \frac{1+e^{-(\propto+\beta_{inpub97} \cdot \mu_{lnpub97})}}{1+e^{-(\propto+\beta_{lnpub97} \cdot \left(\mu_{lnpub97}+\sigma_{lnpub97}\right))}} - 1$

Table II-1 (following two pages) – Summary statistics and correlation matrix of dependent and explanatory variables

Variable	N. Obs	Mean	Std. Dev.	Min	Max
nobeldum	3959	0.002	0.039	0	1
top10cite	3959	0.1	0.3	0	1
lncite98	3959	1.761	1.756	0	8.009
lnpub98	3959	0.888	0.822	0	4.477
lnpub97	3959	2.585	1.371	0	7.073
lncite97	3959	4.178	2.058	0	9.692
lnpub97_fl	3959	1.934	1.426	0	6.836
lncite97_fl	3959	3.193	2.3	0	9.533
constraint	3959	0.657	0.315	0	1.932
lncoauthor	3959	2.279	1.15	0	5.855
collabcore	3959	0.821	0.383	0	1
techcore	3959	0.45	0.498	0	1
pubdepth	3959	0.281	0.058	0	1
lexp	3959	2.354	0.883	0	4.174
academic	3959	0.996	0.044	0	1
prestige	3910	5.889	10.683	0	50
prestiged	3910	0.347	0.476	0	1
newcomer	3959	0.036	0.187	0	1
affil1p	3910	0.591	0.492	0	1

	nobeldum	top10cite	lncite98	lnpub98
nobeldum	1			
top10cite	0.1167	1		
lncite98	0.1169	0.6044	1	
lnpub98	0.0168	0.4711	0.815	1

	lnpub97	lncite97	lnpub97_fl	lncite97_fl	constraint	lncoauthor	collabcore
lnpub97	1						
lncite97	0.815	1					
lnpub97_fl	0.939	0.771	1				
lncite97_fl	0.788	0.888	0.861	1			
constraint	-0.21	-0.225	-0.202	-0.216	1		
lncoauthor	0.649	0.568	0.529	0.455	-0.089	1	
collabcore	0.214	0.325	0.151	0.222	-0.361	0.263	1
techcore	-0.219	-0.076	-0.218	-0.104	0.033	0.017	0.063
pubdepth	0.318	0.165	0.302	0.202	-0.069	0.178	-0.005
lexp	0.847	0.724	0.784	0.674	-0.158	0.551	0.252
academic	-0.004	0.02	0.017	0.005	0.001	-0.003	0.008
prestige	-0.058	0.094	-0.044	0.065	-0.014	-0.059	0.066
prestiged	0.1	0.231	0.11	0.206	-0.062	0.016	0.087
newcomer	-0.371	-0.401	-0.268	-0.274	-0.001	-0.392	-0.42
affil1p	0.677	0.599	0.627	0.568	-0.173	0.394	0.166

	techcore	pubdepth	lexp	academic	prestige	prestiged	newcomer	affil1p
lnpub97								
lncite97								
lnpub97_fl								
lncite97_fl								
constraint								
lncoauthor								
collabcore								
techcore	1							
pubdepth	-0.122	1						
lexp	-0.12	0.234	1					
academic	0.011	-0.001	-0.006	1				
prestige	0.084	-0.063	-0.048	-0.027	1			
prestiged	0.024	0	0.108	-0.019	0.755	1		
newcomer	-0.178	0.001	-0.519	0.017	-0.108	-0.143	1	
affil1p	-0.165	0.188	0.698	0.002	-0.056	0.122	-0.235	1

Table II-2 (following page) – Predictive models of the Nobel winning paper with rare events logit, top 10% of citations with logit, number of forward citations of 98 papers and number of 98 papers both with quasi-maximum likelihood Poisson.

	Relogit	Logit	QML impact	QML prod
	Nobel	Top10c		
DV	nobeldum	top10cite	ncite98	npub98
	b/se	b/se	b/se	b/se
lnpub97	-1.569**	0.800**	0.316**	1.134**
	(0.32)	(0.12)	(0.07)	(0.04)
lncite97	1.450**	1.153**	0.671**	-0.054**
	(0.27)	(0.08)	(0.04)	(0.02)
lncoauthor	0.536+	-0.094	-0.052	-0.015
	(0.30)	(0.08)	(0.04)	(0.02)
constraint	-4.073*	-0.910**	-0.530**	-0.230**
	(1.86)	(0.30)	(0.15)	(0.08)
collabcore		0.012	0.135	0.084
		(0.29)	(0.12)	(0.07)
techcore	-1.714	-0.036	-0.133	0.014
	(1.57)	(0.14)	(0.10)	(0.04)
pubdepth	12.731**	-0.149	2.075+	0.285
	(4.70)	(1.59)	(1.23)	(0.39)
lexp	-1.788+	-2.015**	-1.136**	-0.962**
	(0.95)	(0.21)	(0.15)	(0.05)
prestige	0.036+	0.014+	0.008+	0.001
	(0.02)	(0.01)	(0.00)	(0.00)
prestiged		-0.057	-0.116	-0.078
		(0.19)	(0.11)	(0.06)
academic		1.734	0.277	0.979**
		(2.18)	(0.74)	(0.38)
newcomer	3.019	2.922**	1.932**	0.321**
	(2.62)	(0.63)	(0.44)	(0.12)
affil1p		0.26	0.315+	0.071
		(0.24)	(0.19)	(0.08)
constant	-7.932*	-6.991**	0.919	-0.921*
	(3.29)	(2.28)	(0.81)	(0.41)
N.Obs	3910	3910	3910	3910
Log-Likelihood		-829.552	-94554.127	-7611.202

+ p<0.10, * p<0.05, ** p<0.01

The third model in Table II-2 shows results for the first QML Poisson regression with forward citation count as the outcome variable. Predictably, the results of this model are similar to those found from the logistic model, as both dependent variables depict a similar concept of impact. We further interpret the effect size of these results. A one standard deviation increase in the natural log of prior publication citations increases citation count by 198%[7], while a one standard deviation increase in the natural log of prior publication increases citation count by 54.3%. Similarly, the coefficients on constraint and log experience indicate that a one standard deviation increase in each of the two variables decreases citations by 15.4% and 63.3% respectively. Furthermore, a one standard deviation increase in newcomer increases citation count by 43.6%. At the ten percent significance level, a one standard deviation increase in publication depth, prestige and multiple affiliations increase citation count respectively by 12.8%, 8.6% and 16.8%. In summary, we find that younger scientists with more prior eminence and history, situated in brokerage positions are more likely to discover breakthroughs.

The fourth model in Table II-2 presents a regression with the measure of productivity proxied by the number of papers published in 1998 as our dependent variable. Interpreting the results of the model we find that prior publication quantity is positively and significantly associated with productivity, whereas contrary to the prior two models that depict impact prior publication quality contributes negatively to productivity. This result illustrates that scientists who prioritize quality over quantity may be less productive in order to ensure the quality of their work. Similar to impact models, we also find that brokers with their nexus positions are not only more likely to discover breakthroughs but also tend to be more productive, whereas the number of co-authors is still insignificant.

[7] Effect size = $\frac{e^{\beta_i \cdot (\mu_i + \sigma_i)}}{e^{\beta_i \cdot (\mu_i)}} - 1 = \frac{e^{\beta_{lncite97} \cdot (\mu_{lncite97} + \sigma_{lncite97})}}{e^{\beta_{lncite97} \cdot (\mu_{lncite97})}} - 1$

Again we observe no evidence of periphery or core nor specialization or generalization. We also find that younger scientists write higher impact papers and more papers, possibly because they are incentivized by the tenure system in place at most academic institutions. Expectedly, scientists in academic institutions are more productive. Finally, newcomer also positively affects publication count. Effect sizes are computed and shown in Figure II-1 for all four models.

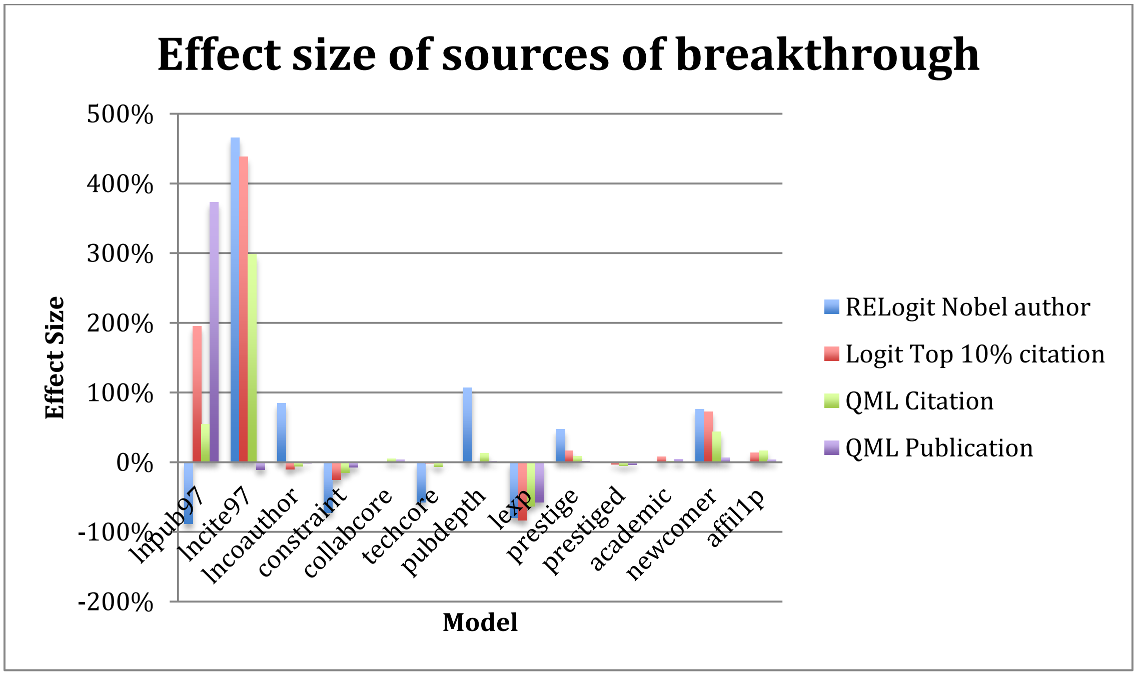


Figure II-1 – Effect size of each regression model by explanatory variable. The effect size is computed by increasing the variable under study by one standard deviation from the mean while holding all other explanatory variables at the mean.

Aware that ordinary least square models yield biased coefficient estimators for logistic and count models, we employ OLS primarily to shed light on the predictive power of our theoretical models while making sure to apply the natural logarithm to the two count dependent variables for this set of regressions. Figure II-2 shows the percent contribution to the variance of each explored theory for all four models. In these models we only interpret R^2 and delta R^2 measures even though coefficients are directionally consistent with the more appropriate non-linear models. Unsurprisingly, predicting who will discover the breakthrough per se is extremely hard due to the rare nature of such events. Any model trying to predict 6 successful events out of 4,000 would be challenged, and this is illustrated by the total R^2 of 0.8% when predicting authors of the Nobel winning paper (see Table II-3). However, the picture is less gloomy if we relax the predictive requirements and concede that any scientists who attain the top ten percent of the citation distribution has similar chances of discovering the breakthrough – still, anything beyond that point is mostly driven by chance and circumstance. Although the total explained variance jumps to almost 20%, we still only have one out of five chances of predicting the correct top 10% citation scientist. Furthermore, without prior publication and eminence, prediction using the theoretical themes only provides 5.6% explained variance. Further relaxing predictive requirements to the number of citations yield increased total explained variance to 37.8%, where together measures of prior eminence and productivity account for 24.7% of the variance and the remaining theoretical themes add another 13.1%. Finally when productivity becomes the dependent variable, explained variance increases further given that productivity is more consistent than extremely rare breakthrough events. Indeed with all exploratory variables included, the R^2 raises to 49.6%.

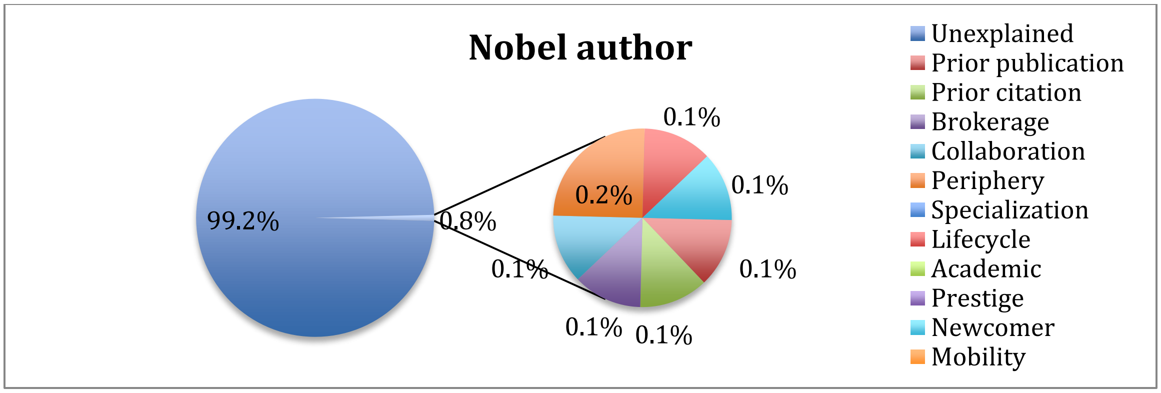
Robustness Checks

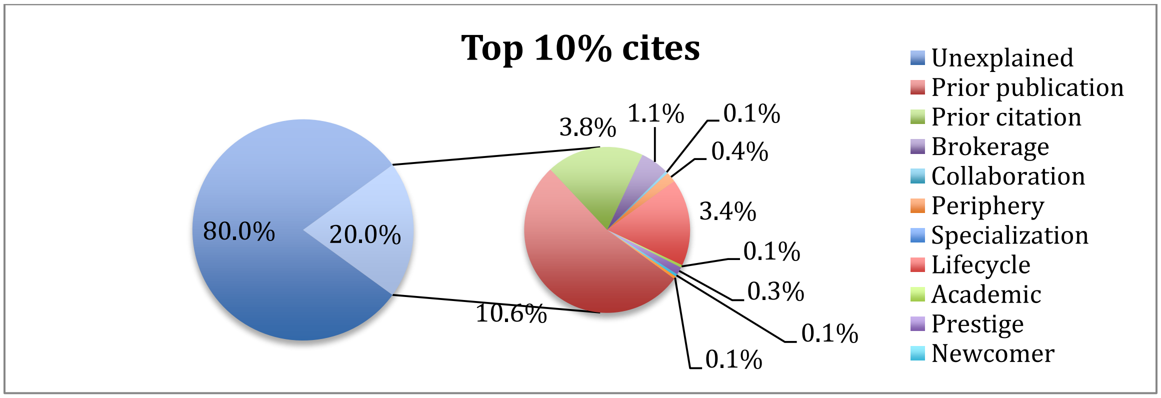
We run several sets of robustness checks to ensure stability of our results. We first run a split sample analysis whereby we randomly divide the initial sample in half. We then perform the same regressions on each of the two split samples and obtain very similar results between the split samples as well as compared to the initial full sample (available upon request from the authors). Similarly, we run a set of regressions while excluding the

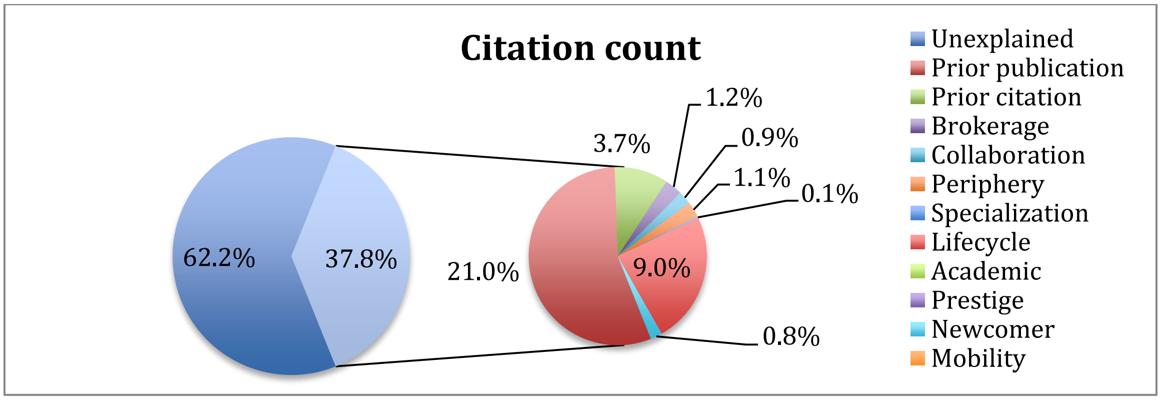
144 newcomers for whom we did not possess any information on the explanatory variables. Again the results are very stable and comparable to those obtained from the initial full sample.

We run another set of similar regressions that reflects the publishing convention in our biological setting using the number of first or last author publications prior to 1998 (lnpub97_fl) and their forward citations (lncite97_fl) as explanatory variables to predict first or last authored 1998 papers (npub98_fl), their citations (ncite98_fl) and their citation distribution (top10cite_fl). The authorship order in biological publications stipulates that first authors are usually the scientists who perform most of the work and experiments, while the last author is typically the principal investigator who is the head of the laboratory in which the research is done. The role that middle authors play in a publication is more heterogeneous. While the contribution of some middle authors may be just as substantial as the first or last authors, others may have merely weighed in by providing a sample or an extract required for experiments or perhaps are technicians who are less involved in the intellectual process. Thus limiting to papers where the scientists are either first or last authors enable us to include researchers who arguably contributed intellectually the most to a given publication. The analysis format is comparable to the main results as we run the same analyses with similar models and corresponding dependent variables (Table II-4).

Figure II-2 (following page) – Explained variance for each model and source of breakthrough explored


Nobel author
99.2%
0.8%
0.2%
0.1%
0.1%
0.1%
0.1%
0.1%
0.1%
Unexplained
Prior publication
Prior citation
Brokerage
Collaboration
Periphery
Specialization
Lifecycle
Academic
Prestige
Newcomer
Mobility


Top 10% cites
80.0%
20.0%
3.8%
1.1%
0.1%
0.4%
3.4%
0.1%
0.3%
0.1%
0.1%
10.6%
Unexplained
Prior publication
Prior citation
Brokerage
Collaboration
Periphery
Specialization
Lifecycle
Academic
Prestige
Newcomer


Citation count
62.2%
37.8%
3.7%
1.2%
0.9%
1.1%
0.1%
21.0%
9.0%
0.8%
Unexplained
Prior publication
Prior citation
Brokerage
Collaboration
Periphery
Specialization
Lifecycle
Academic
Prestige
Newcomer
Mobility

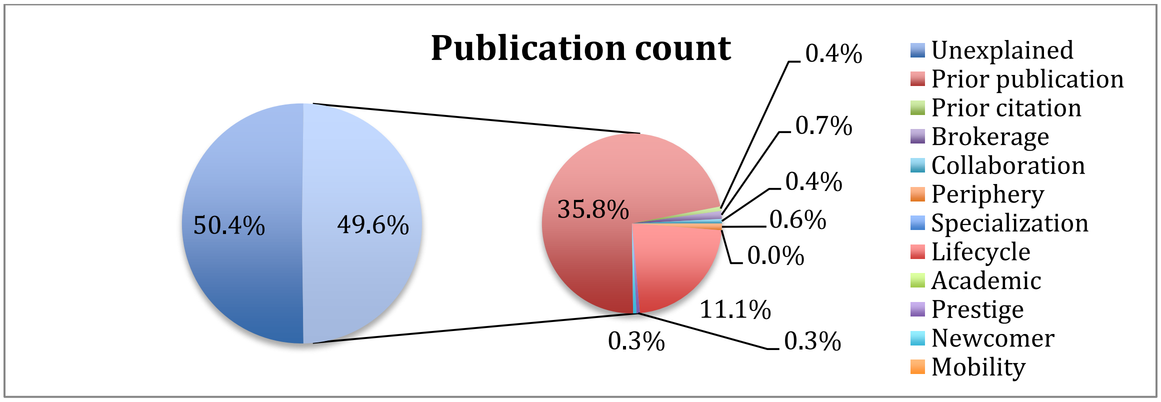

Publication count
50.4%
49.6%
35.8%
0.4%
0.7%
0.4%
0.6%
0.0%
11.1%
0.3%
0.3%
Unexplained
Prior publication
Prior citation
Brokerage
Collaboration
Periphery
Specialization
Lifecycle
Academic
Prestige
Newcomer
Mobility

Table II-3 (following two pages) – Predictive OLS model for the Nobel winning paper, top 10% of citations, number of forward citations of 98 papers and number of 98 papers

	OLS Nobel	OLS Nobel	OLS Top10c	OLS Top10c
DV	nobeldum	nobeldum	top10cite	top10cite
	b/se	b/se	b/se	b/se
lnpub97	-0.002*	-0.002*	0.012*	0.079**
	(0.00)	(0.00)	(0.01)	(0.01)
lncite97	0.001	0.001*	0.049**	0.054**
	(0.00)	(0.00)	(0.00)	(0.00)
lncoauthor		0.001+		-0.010+
		(0.00)		(0.01)
constraint		-0.002		-0.095**
		(0.00)		(0.01)
collabcore				-0.017
				(0.01)
techcore		-0.002		0.003
		(0.00)		(0.01)
pubdepth		0.012		-0.139*
		(0.01)		(0.07)
lexp		-0.002		-0.110**
		(0.00)		(0.01)
prestige		0		0.001
		(0.00)		(0.00)
prestiged		0.002		0.001
		(0.00)		(0.02)
academic				0.124
				(0.09)
newcomer		0.011		0.083**
		(0.01)		(0.03)
affil1p				-0.02
				(0.01)
constant	0.003	0.001	-0.135**	-0.057
	(0.00)	(0.00)	(0.01)	(0.09)
N.Obs	3959	3910	3959	3910
R2	0.002	0.008	0.144	0.2

+ p<0.10, * p<0.05, ** p<0.01

DV	OLS impact lncite98 b/se	OLS impact lncite98 b/se	OLS prod lnpub98 b/se	OLS prod lnpub98 b/se
lnpub97	0.245**	0.798**	0.413**	0.718**
	(0.03)	(0.04)	(0.01)	(0.02)
lncite97	0.281**	0.356**	-0.044**	-0.017+
	(0.02)	(0.02)	(0.01)	(0.01)
lncoauthor		-0.150**		-0.056**
		(0.03)		(0.01)
constraint		-0.441**		-0.121**
		(0.08)		(0.03)
collabcore		0.062		0.069*
		(0.07)		(0.03)
techcore		-0.117*		-0.037+
		(0.05)		(0.02)
pubdepth		0.487		-0.032
		(0.45)		(0.18)
lexp		-1.024**		-0.547**
		(0.05)		(0.02)
prestige		0.006+		0.003+
		(0.00)		(0.00)
prestiged		-0.091		-0.073*
		(0.08)		(0.03)
academic		0.296		0.342+
		(0.55)		(0.20)
newcomer		1.114**		0.343**
		(0.17)		(0.05)
affil1p		0.071		0.001
		(0.07)		(0.03)
constant	-0.043	0.734	0.006	0.224
	(0.06)	(0.58)	(0.02)	(0.21)
N.Obs	3959	3910	3959	3910
R2	0.247	0.378	0.362	0.496

+ p<0.10, * p<0.05, ** p<0.01

Due to lack of variance for the rare event logit when restricting the dependent variable to first or last authors of the Nobel paper, the model does not converge and therefore cannot be interpreted. However, all remaining three models in the robustness results are very similar, in terms of significance levels and directionality of beta coefficients, to their corresponding models that use papers with all authors. Similar to the base models, the consistent themes throughout the two robustness impact models remain that more prior publication and eminence, brokerage, younger scientists and newcomer tend to produce more impactful first or last author works. Moreover, for scientists in the top 10% of citations collaboration, technical periphery, prestige and multiple affiliations also have significant positive effects. For the productivity model, we find that similar to its baseline model prior productivity, brokerage, younger and scientists in academia tends to increase publication of first or last authored works. Moreover, specialization and multiple affiliations also have a positive effect on publication while newcomer exhibits a negative effect. The OLS results (available upon request from the authors) show less explained variance because we withhold information on middle-authored papers and restrict prior history and eminence to a sub sample.

Table II-4 (following page) – Robustness check where the dependent variables are derived from papers published as first or last author in 1998.

	Logit Top10c	QML impact	QML prod
DV	top10cite	ncite98	npub98
	b/se	b/se	b/se
lnpub97_fl	0.365**	0.217+	1.047**
	(0.12)	(0.12)	(0.05)
lncite97_fl	0.863**	0.716**	-0.021
	(0.07)	(0.05)	(0.02)
lncoauthor	0.199*	-0.045	0.004
	(0.08)	(0.06)	(0.03)
constraint	-1.064**	-0.583*	-0.257**
	(0.29)	(0.24)	(0.10)
collabcore	0.107	0.19	0.031
	(0.27)	(0.16)	(0.10)
techcore	-0.276*	0.024	0.027
	(0.14)	(0.11)	(0.05)
pubdepth	0.555	0.6	1.000*
	(1.43)	(1.12)	(0.50)
lexp	-1.565**	-0.987**	-0.885**
	(0.20)	(0.16)	(0.07)
prestige	0.013+	0.008	0.003
	(0.01)	(0.01)	(0.00)
prestiged	-0.011	-0.083	-0.149*
	(0.17)	(0.11)	(0.06)
academic	0.893	1.624	1.429*
	(1.98)	(1.00)	(0.72)
newcomer	1.104+	0.852*	-0.354+
	(0.59)	(0.39)	(0.18)
affil1p	0.518*	0.174	0.181+
	(0.24)	(0.20)	(0.10)
constant	-4.428*	-0.438	-1.840*
	(2.08)	(1.10)	(0.75)
N.Obs	3910	3910	3910
Log-Likelihood	-884.014	-53243.556	-5313.814

+ p<0.10, * p<0.05, ** p<0.01

Weaknesses and Limitations

These above findings should be interpreted with caution for a number of reasons. First, the processes of scientific discovery are cognitive and remain extremely difficult to capture with purely bibliometric measures. Most of the bibliometric innovations literature has ignored this cognitive aspect though some have recently attempted to implement such measures (Kaplan & Vakili, 2012). Second, our variables that attempt to capture interactions between scientists only partially seize these exchanges. Formal co-authorship collaborative links are captured through our network measures while all other social interactions such as seminars, conferences, and hallway chats are inevitably left out.

Third, any study that attempts to predict the source of an innovation will be sensitive to the definition of those at risk of innovating. Our definition of the community of scientists attempting to solve the puzzling mechanism of gene silencing is mainly functional, but because the same phenomenon was named differently by plant, fungi and animal scientists our definition is by force of association also organism-based. This definition, however, is noisy given that the MeSH keywords we use are also assigned to other fields studying various biological phenomena, such as the interferon community. Other definitions of the community could have taken a purely model organism view rather than our phenomenon-based angle, whereby scientists in the plant, fungi and worm fields – the three communities that initially observed gene silencing – would make up the sample. However, this alternative would result in an even noisier set with the combination of three large organism-based communities regardless of the biological phenomenon each scientist focuses on.

Finally, despite a sample of nearly four thousand scientists this paper is a case study of one particular breakthrough. For more generalizability, we would have to replicate this study across many communities, essentially moving up the level of analysis. This is now possible given increasing computing and analytical power and we expect to see this research direction. A related and interesting question would be to ask which communities are more likely – or perhaps more ripe – to discover breakthroughs. The analysis at the community level would entail aggregating individual scientists into systematically parsed communities, perhaps using techniques developed by applied physicists (Fortunato, 2010), and exploring research questions that assess the probability of a given field in making a revolutionary discovery.

v. Discussion

Aside from assessing the ability of current theories to predict future breakthroughs, the present work also identifies where such significant discoveries or inventions arise from within a community and what characteristics make a particular scientist more likely to discover them. Our data also sheds light on the number of researchers that came from within the field of RNAi and those who are completely new. Thus this work informs the micro-foundations of the innovations literature by bringing individual level data to a question typically focused on the publication, patent or organization as the unit of analysis, or remained mainly theoretical. Furthermore, to the best of our knowledge this work is the first to bring together several theories of creativity into one single predictive study thereby taking a comprehensive look at the phenomenon and the research field. Referring back to the deliberations within the extant literature, our results show that prior productivity, prior eminence, brokerage, youth and newcomer consistently contribute to a researcher's subsequent impact (measured using the top tenth cites and citation count) and, consequently, add weight to the scale of evidence on particular sides of each debate.

Looking across the four baseline models in Table II-2, brokerage consistently affects both impact and productivity of future publications. Brokers with their nexus position are able to seize and control the information that flows to them and recombine in such a way to produce highly impactful research (Burt, 2004). Since our study concentrates on determining the production of revolutionary discoveries rather than describing diffusion mechanisms of such discoveries, which the opposing cohesiveness camp lends more evidence to, the prevalence of brokerage in this setting is expected.

We also find consistency throughout the baseline models that younger researchers are more prone to breakthrough, lending empirical evidence to Simonton's (1989) argument that younger scientists are not weighed down by the conventional thinking of a given field enabling them to take bigger intellectual leaps. They are also more productive; a manifestation of the academic tenure process that demands higher productivity from younger scientists while after tenure is granted the incentive for heightened productivity diminishes. Furthermore, we also observe a constant and positive newcomer effect except for the Nobel author model.

Both models depicting impact (logit predicting scientist top 10% of citation and QML Poisson predicting citation count) show that breakthroughs require many prior attempts and trials before hitting one that becomes a great success. These results corroborates Mowery and Ziedonis's (2002) finding with patents showing that inexperienced academic patenters tend to have less significant patents than those with more experience. This suggests that the road to breakthrough discovery involves a sizeable learning process and many trials. The productivity model, however, illustrates a fundamental tradeoff between publication productivity and quality where better productivity is positively associated with the quantity of prior works and negatively correlated with prior eminence. In sum, scientists who produce highly impactful works tend to produce fewer publications. Given that scientists are constrained with limited time, attention and resources, producing better work on the quality dimension takes away from one's ability to generate a greater amount of work.

Bibliometric empirical evidence for the collaborative teams vs. individual researchers debate is plentiful, but our main regression results remain insignificant. This lack of significance may be explained by the exceeding number of theories we test in our models compared the papers that we draw from. For instance, using patents Singh and Fleming (2010) only control for the number of claims, while Wuchty, Jones and Uzzi (2007) show correlational plots between teams or average team size and relative team impact. Specifically, the measures of prior publication and eminence implicitly contain information on collaboration as they include all prior works no matter the number and order of author publishing. Additionally, these studies are performed at the patent or paper level, whereas our study is centered at the individual level and aggregates all prior co-author into a single measure.

Position in the core versus periphery demonstrates no significance. Jeppesen and Lakhani (2010) specifically explore social and technical marginality without controlling for the numerous number of theories we include in the models of this work. More broadly, these problems also suggest that the literature needs to go beyond existing correlational studies and establish causality between sources of breakthroughs and subsequent impact.

With regard to the debate between generalist vs. specialist and their effect on impact, we are also unable to find consistent evidence in our results which may be explained by the fact that compared to prior literature our study is either at a different level of analysis or in a different setting. While numerous studies have studied diversity and specialization on innovative outcome at the team (Dougherty, 1992; Leonard-Barton & Swap, 1999), firm or industry levels (Brusoni, Prencipe, & Pavitt, 2001; Romer, 1987), exploration of the effect of specialization at the individual level, not to mention using bibliometric data, is surprisingly thin. An exception lies in the medical literature where many studies have considered physician specialization on patient outcomes in various diseases. However, even though both scientists and clinicians are called upon to solve a particular problem, the underlying process and the outcome measure of success are undoubtedly different.

One noteworthy (though frustrating) result is the weak predictive power of current theories of breakthrough emergence from OLS regressions, thus illustrating the difficulty of forecasting such rare events. While the predictive power for future citation counts is relatively high (a little less than 40%), that for being in the top 10% of the citation distribution is only at half (~20%), not to mention the less than 1% explained variance for predicting the actual breakthrough. Ignoring prior history and eminence, the combination of all theoretical themes shrink significantly and explain only 13.2% of the variance for future citation count and 5.3% for top 10% citation. In other words, it is still very hard to predict the impact of a scientist's work without data on their prior publication history and eminence. We fully acknowledge that the rare nature of breakthroughs make them particularly hard to predict as there are significant elements of chance and serendipity. However, we believe that the rate of progress in understanding the sources of scientific and technical breakthroughs can be further increased if we could get around issues of convenience sampling associated with historical accounts and the lack of causal identification in correlational bibliometric studies.

Lack of causal inference, especially in the network literature, remains a major critique of the breakthrough and bibliometric literature. Not only are most network measures endogenous, such as brokerage and collaborative core; it is also hard to disentangle the effect of experience with many of the explanatory variables we employ. Consequently, one should be cautious in not over interpreting and over relying on results from purely correlational bibliometric papers. Several techniques that construct clean treatment and control groups can be employed to address this issue besides the simple use of instrumental variables. Identifying exogenous shocks builds the treatment group, while econometric techniques such as matching on covariates (Furman & Stern, 2011) or regression discontinuity (Kerr, Lerner, & Schoar, 2011) are used to create the control group. For instance, Azoulay et al. (2010) explore the role that superstar scientists play in the generation of knowledge by exploiting the sudden exogenous death of such scientists. Similarly, the role of institutions on cumulative research and knowledge diffusion is investigated by using exogenous shifts of biomaterials across institutional settings (Furman & Stern, 2011). Identifying exogenous shocks, such as unforeseen policy changes, sudden closures of institutions or major firms etc., requires in-depth understanding of the phenomenon under study. Qualitative fieldwork, such as interviews with informant stakeholders and observations, is usually the necessary precursor to pinpoint such natural experiments and subsequently isolate causal mechanisms. Moreover, innovation contests (Terwiesch & Ulrich, 2009), such as TopCoder, enable researchers to design experiments in which to test various causal mechanisms that affect creativity and innovative capability using designated control and treatment groups.

With qualitative fieldwork, one needs to be cautious that it does not suffer from the known shortcomings of the method. For instance in historical accounts, the number of participants included is usually limited to those in the immediate proximity of the winners, such as their mentors, collaborators and eminent fellow scientists racing for the same discovery. Consequently, these individual case studies tend to sample *ex post* by convenience and, therefore, it remains hard to ensure the extensiveness of the study. They lack the macro view enabled by large archival quantitative methods. Moreover, qualitative research does not always control for confounding factors in their narratives (King, Keohane,

& Verba, 1994), which increases the difficulty of identifying the sources that causally enhance breakthrough discovery and synthesizing individual findings from each work. Comprehensive datasets should be leveraged to ensure exhaustiveness of historical studies. For instance, qualitative papers could build on regression models and sample scientists who were incorrectly predicted from the error terms, thus systematically identifying scientists who discovered a particular breakthrough but also those at risk. Moreover, with the increased availability of electronic contents of large corpuses, historical accounts can comprehensively analyze the content of entire bodies of knowledge (El Ghaoui et al., 2011) not only ensuring completeness in sampling but also finding all subsequent occurrence of concurrent ideas with their topic of interest.

vi. Conclusion

Using a sample of scientists at risk of discovering RNA interference in 1998, a collection of predictions from the bibliometric creativity literature, and all available bibliometric data from PubMed up to 1997, we attempted to predict the breakthrough creativity and productivity of those scientists in 1998. Most theoretical predictions came up insignificant, including past collaboration, core position in technological or social space, specialist versus generalist or mobility across institutions. A few theoretical predictions provided limited explanatory power: prior publication made a scientist more likely to publish many papers, though at the expense of quality; prior publication of highly cited papers made a scientist more likely to discover breakthroughs, at the expense of productivity; social brokerage in past publications made a scientist much more likely to discover a breakthrough. Age surprisingly, once all control variables were included, had a monotonic negative correlation with subsequent productivity and breakthrough discovery.

The results are sobering, especially in attempting to predict the authorship of the actual Nobel Prize winning paper, as a collection of current theories of creativity, and scientists' prior productivity and quality, together can explain less than 1% of the model variance. Less stringent definitions of breakthrough were more successful, with almost 50% of the variance being explained. Such results can be seen pessimistically – we have made essentially no bibliometric progress in predicting breakthroughs – or optimistically – we can predict half the variance in simple publishing productivity within a given field, given the history of the field.

From a policy standpoint, this work should give pause to the current efforts to use big data and computation to understand, justify, and optimize public investment in science. Policy makers and corporate lab managers should absolutely not apply bibliometric results blindly, given the large unexplained variance in our predictive regressions. Automated tools could certainly support a process run by domain experts. For example, as part of the peer review of grant applications, the predictive number could be calculated, but hopefully not over-interpreted, lest we kill (what still appears by all accounts to be the) golden goose of science. The current and typical peer-reviewed grant process may be inefficient and frustrating, but it is probably the least-worst method; it would be foolish to abandon it in favor of purely bibliometric criteria.

III. Moving Beyond Bibliometrics

i. Introduction

Starting with Schumpeter's notion of creative destruction (1942), breakthroughs have intrigued scholars and practitioners alike. The literature is rife with works that attempt to identify, quantify and describe sources of breakthroughs, especially when radical discoveries and inventions have shown to be an important foundation of scientific and technological advancement, and have been at least weakly linked to wealth creation and economic growth (Mueller, 2006). However, despite widespread scholarship on factors and circumstances that enhance breakthrough discovery, our ability to predict from whom breakthroughs are most likely to emerge is still relatively weak (Chai & Fleming, 2012). Aside from the evident explanation that breakthroughs are inherently rare and serendipitous events, several characteristics of the current literature contribute to this limited predictability.

Not only has extant literature emphasized the positive outcome of breakthroughs emerging without exploring as much why breakthroughs are missed and delayed, but because of this lapse in understanding, it has also concentrated most of its efforts on enhancing the problem solving process without taking into account that failures also exist at the problem identification stage. Furthermore, bibliometric and archival methods have limited the researcher's ability to make inferences beyond those from observable and measurable proxies. I address these shortcomings, and take a different approach from the usual outcome driven studies by digging deeper into the counterfactual process of why and how breakthroughs are missed and delayed. I draw a clear distinction between barriers that hinder breakthroughs from being discovered at the problem identification phase versus the problem solving stage. Moreover, by employing a case historical analysis of RNA interference (RNAi), a seminal finding in molecular biology, and interviewing scientists with the potential to make groundbreaking discoveries, I shed light on the phenomenon of breakthrough using a cognitive lens with institutional underpinnings.

I find that scientists on the verge of breakthrough missed the seminal discovery not only due to difficulties with solving a particular problem but also because of failures to identify the breakthrough opportunity before and while solving for the problem. My findings suggest that at the basis of this failure underlies a cognitive mechanism stemming from three barriers with institutional underpinnings: framing barriers, boundary barriers and paradigmatic pressures. In the problem identification stage, paths dependence from established technologies and the quest toward normal science blinded scientists from recognizing a breakthrough potential. Instead they framed RNAi as a tool useful in uncovering answers to their initial experiments while ignoring it as a scientific concept worthy of study in and of itself. Furthermore, existing boundary barriers between communities of scientists prevented recognition of links between several instances of odd observations in prior works, and, in turn, aggravated this difficulty in identifying the breakthrough opportunity by misrepresenting the magnitude of the problem. In the problem-solving stage, scientists also suffered from socio-cognitive paradigmatic pressures of being constrained by current dogma. To avoid being wrong, they hesitated to propose solutions that significantly strayed away from the confines of established theory. Again coupled with the boundary barrier that prevented connecting the dots, similar anti-dogmatic observations and results stayed isolated and diminished scientists' confidence in identifying and proposing a new revolutionary paradigm.

Breakthroughs can be depicted by various measures but their definition is ultimately linked to the notion of impact (Simonton, 1999), defined herein as encompassing dimensions of both creative novelty and success. As opposed to some discoveries or inventions that become technological dead ends, breakthroughs are advances that disturb the previous understanding of a particular phenomenon in a fundamental manner and are foundationally at the basis of further enhancements. RNA interference fits the above definition. Not only did it disturb the previous conceptualization of the central dogma of life, it also gave rise to a new research technique of knocking down genes. Furthermore, it has had major commercial implications through the introduction of new molecular modalities in therapeutics by moving away from the traditional small molecular formulations based on chemistry. Several prestigious awards also recognized the field, most notably the Nobel Prize in Physiology and Medicine in 2006 and the Lasker Award for Basic Medical Research in 2008. Additionally, the naming of small interfering RNA (siRNA), a class of double-stranded RNA (dsRNA) involved in the RNAi pathway, as breakthrough of the year in 2002 by *Science* (Couzin, Enserink, & Service, 2002) also supports my decision to research RNA interference as a creative breakthrough.

The organization of this work is as follows: After reviewing the literature on sources and processes of breakthrough emergence, I place the RNAi discovery in historical and scientific contexts and describe the methods I employed to gather and analyze the data. I elaborate on the themes from my findings and propose a cognitive framework with institutional hinges that describes why breakthroughs are missed at various stages of the discovery process. I, then, move away from the inductive front end to deductively propose additional sources that enhance the likelihood for scientists to find a revolutionary discovery, by operationalizing using traditional bibliometric measures remedial practices that scientists offered as ways to circumvent the barriers. Finally, I conclude with a discussion on the implications of my results to extant literature.

ii. Literature Review

Sources of Creativity and Breakthrough

Scholars of innovation have put forth many hypotheses identifying sources of creativity employing diverse research designs at multiple levels of analyses in various settings. Despite constant attention, consensus is still relatively scarce. Evidence identifying sources of creativity at the organizational level starts with the age-long debate between whether small entrepreneurial entrants (Schumpeter, 1934) or major incumbents (Schumpeter, 1942) are the basis of creative inventions. And, expands into identifying capabilities that are required of firms to stay inventive, such as absorptive capacity (Cohen & Levinthal, 1990), dynamic capabilities (Teece, Pisano, & Shuen, 1997), experimenting early and often (Thomke, 2003) and sampling a large landscape for multiple trials (Rivkin & Siggelkow, 2002) and recombination (Fleming, 2001). Although seminal and having spawned off entire streams of literatures, these works have mainly concentrated on the firm's ability to build problem-solving skills as a way to sustain creativity, and have largely ignored the significance of problem identification in creating breakthroughs that I stress herein. Moreover, these studies cannot be readily applied to the current context because the locus of decision-making in scientific research is centered at the principal investigator level where heads of labs are responsible for providing funding, hiring personnel, deciding the research direction of their laboratory, etc. This predominant structure in science, thus, prevents analysis at the organizational level and requires studies to be performed either at the team/laboratory or individual level. At the laboratory level, the dynamic nature of the boundaries of these groups also complicates analysis. PhD students or postdoctoral fellows initially working under a single principal investigator in the same lab eventually take on professorship positions and start their own individual labs. Consequently, I perform the analysis herein at the individual scientist level.

At the individual level, the literature has mainly focused on factors or characteristics that contribute to breakthrough discovery. Despite a wide body of work, very little consensus has emerged. From a network analysis perspective, brokers (Burt, 2004) versus those situated in more cohesive positions (Obstfeld, 2005; Uzzi, 1997) have both shown to be more creative, although studies have also identified particular circumstances to tease apart their conflicting effects (Fleming et al. 2007). Similarly, whether individuals at the core of a community (Collins, 1998; Gieryn & Hirsh, 1983) or those sitting at its periphery are more creative (Jeppesen & Lakhani, 2010) is still debated. Specialists with deep technical knowledge are better equipped to predict outcomes beyond the frontier, as opposed to generalists who can bring diverse components together to recombine (Dougherty, 1992; Leonard-Barton & Swap, 1999). Arguments for both younger individuals to realize breakthroughs, because they are less entrenched in established beliefs (Simonton, 1989), or more experienced individuals, because they must work through the accumulation of knowledge (Jones, 2009) also exist. Finally, mobility between multiple affiliations affords exposure to diverse ideas, but is also associated with high transition and setup costs (McEvily & Zaheer, 1999). At the team level, agreement has emerged that collaboration increases chances of breakthrough (Singh & Fleming, 2010; Wuchty, Jones, & Uzzi, 2007) though collaboration effects are not homogeneous throughout the entire process (Girotra, Terwiesch, & Ulrich, 2010).

Despite the topic's perpetual allure, breakthroughs are difficult to study because creativity involves cognitive mechanisms extremely difficult to capture using purely archival and bibliometric data. Although some scholars have recently attempted to create cognitive measures using semantic analysis (Kaplan & Vakili, 2012), these newer techniques have yet to be widely tested. Using qualitative methods such as field interviews, however, offers the researcher a rare glimpse into the informant's train of thought and sense making throughout the process of discovery; thus, unfolding a richer and more complete picture of the process of breakthrough beyond those captured by purely archival methods that mainly focus on how various factors impact breakthrough outcomes. Coupled with the rarity of breakthroughs, many empirical challenges arise as most statistical tools available to social scientists find the average effect while ignoring or even dropping the exact outliers that breakthroughs consist of. For instance, even though preceding works have determined various structures and qualities that enhance creativity, they fail to explicitly address the rare nature of breakthroughs (though exceptions do exist such as (Singh & Fleming, 2010)).

Emergence of Scientific and Technological Breakthroughs

Rather than focusing on factors that improve discovery, the literature that explores *how* breakthroughs emerge has described the process by which both technological and scientific inventions or discoveries are made. In the technological realm, many works have investigated the evolution of technologies and the emergence of a standardized technological form from multiple paths. Works on the social construction of technology emphasize that the use and development of technologies are heavily ingrained within a social context (Bijker, Hughes, & Pinch, 1987). While a tangential stream is the socio-cognitive model of technological evolution that stresses the interaction between the social construction of technology and the cognitive aspect of belief (Garud & Rappa, 1994). Although both explore how technologies emerge, they primarily deal with the issue of achieving standardization amongst multiple prospective technological forms whereas in my scientific setting the process of breakthrough emergence is a process where scientists strive to find a single truth.

Within the scientific institution, Thomas Kuhn introduced the notion of paradigm shifts as the underlying structure for scientific revolutions through accumulated anomalies that amount to crisis (1962). Although seminal and highly influential, Kuhn's book is mainly theoretical and describes the process by which scientific revolutions emerge rather than focusing on what barriers impede them. Thus, I build onto this work by delving into the counterfactuals and providing detailed mechanisms of failures that cause delayed and missed scientific revolutions, complementing the thin existing articles that study missed breakthroughs (Berson, 1992; Dyson, 1972).

Science vs. Technology

To gain a full understanding of how scientific breakthroughs emerge, a clear distinction must first be drawn between science and technology. Although works have either theoretically discussed differences between knowledge created in the scientific versus technological realms (Dosi, 1982; Merton, 1957), independently studied how technological inventions (Fleming, 2002) and scientific discoveries (McFadyen & Cannella, 2004; McFadyen, Semadeni, & Cannella, 2009) arise, or described how they co-evolve (Cockburn & Henderson, 1998; Murray, 2002), to the best of my knowledge very few have explored how this difference is manifested throughout the knowledge production process.

Science and technology have been defined following two main streams in extant literature. One stream, classified under the new economics of science, takes an institutional stance (Dasgupta & David, 1994; Merton, 1957) while the other gets to the nature of knowledge generated. Under the institutional view, science is a distinctive incentive system from technology. Science is primarily characterized by openly sharing knowledge through academic publications produced mainly from research universities and institutes, and supported by a priority-based reward system (Merton, 1957). The technology institution, in contrast, aims at protecting their inventions using patents and trademarks, amongst other methods, for economic ends in order to facilitate extraction of rents through appropriation and commercialization (Dasgupta & David, 1994). The second stream depicts the relationship between science and technology by the nature of knowledge each creates. Science concentrates on demonstrating the *why* through a process of posing hypotheses that are empirically tested so as to refine theory, while technology searches for recipes of *how* by developing practical and useful techniques. In other words, science is preoccupied by the search for truth and mastering the underlying mechanism of action; whereas in technology, as long as an invention works, why and what happens within the black box between input and output is not always relevant.

Looking back at the literature, not only is the predictive power of extant theories of sources of creativity and breakthrough limited, many studies have also confounded their effect between scientific discoveries and technological inventions. Moreover, most of these studies focus on the outcome of attaining revolutions through problem solving rather than the counterfactual of missing breakthroughs. Consequently, I address these gaps herein using qualitative methods by understanding why revolutions are missed or delayed through the exploration of barriers to breakthrough using a cognitive lens with institutional underpinnings, and suggest a framework that encompasses both problem identification and problem-solving failures of identifying and proposing breakthrough opportunities. I also explore how differences in science and technology are exhibited throughout the knowledge production process.

iii. Methods

It is perhaps not surprising that many theories identifying sources of breakthroughs are conflicting and limited because various characteristics of breakthrough make them inherently difficult to study. These include, and are not limited to, the cognitive nature of path breaking discoveries coupled with their scarcity where only successes are easily observed. Yet following prevailing assumptions in the innovation literature that highly uncertain creativity is a path dependent process of recombinant search rather than a single radical event (Fleming, 2001; Henderson & Clark, 1990), I conceptualize breakthrough as marked by multiple failures before eventual success. Therefore using a case history method to study breakthrough emergence is appropriate as it unearths the nuances of multiple trials along the path of discovery irrespective of whether these were failures or successes (Corbin & Strauss, 2008; Miles & Huberman, 1984). Since I study the breakthrough *ex post*, understanding the circumstances scientists faced *ex ante* is critical. Although interviews potentially suffer from hindsight bias, they are useful in inquiring about cause of failures that are not always easy to obtain using purely archival methods. To minimize such retrospective sense making, I triangulate my findings from the interview data with archival sources such as the Nobel lectures, transcriptions of the Nobel interviews, and RNAi paper publications from each identified interviewee (Golden, 1992).

Many historical case studies of breakthroughs exist. Although extremely rich and incredibly descriptive when characterizing the invention or discovery, the number of stakeholders included in such historical accounts is usually limited to those in the immediate proximity of the winners, such as their mentors, collaborators, and eminent fellow scientists racing for the same discovery. Consequently, these individual historical accounts may suffer from convenience sampling and lack the macro and systematic view enabled by large archival quantitative methods. To ensure exhaustiveness of my case history and avoid the same pitfall, I first identified a community of scientists prior to the discovery of RNAi who were working in precursor fields to RNAi using the comprehensive Author-ity database of disambiguated authors derived from MedLine (Torvik & Smalheiser, 2009). Correctly defining this community of scientists was crucial to understand how scientific breakthroughs arose within it. I, then, developed a selection method through residual analysis to systematically determine not only the researchers who emerged but also ones with the highest potential. I concentrated my interviews on those who did not ultimately discover the breakthrough – the counterfactuals – to gain a different and understudied perspective on the phenomenon.

Historical, Scientific and Technological Context of RNA Interference

This work inducts sources of scientific breakthroughs based on the discovery of RNA interference in molecular biology. RNA interference is a naturally occurring endogenous mechanism triggered by dsRNA precursors. These long strands of dsRNA are processed into small interfering RNAs or microRNAs that bind to other types of RNAs which in turn increase or decrease their activity thereby turning genes on and off (Meister & Tuschl, 2004). RNA interference is valuable as a research tool as well as in biotechnology therapeutic development. For instance, in research, synthetic dsRNA introduced into cells can induce suppression of specific genes of interest both *in vitro* and *in vivo*, thus enabling scientists to understand gene function. It can also be applied to large-scale screenings that systematically shut down each gene in the cell and helps in identifying components necessary for a particular cellular process or event. In biotechnology and medicine, turning off maladies, such as Huntington's or certain cancers, can conceivably use exploitation of the RNAi pathway to treat genetic diseases.

RNA interference is a gene silencing mechanism that strays away from the central dogma of molecular biology, which dictates how genetic information encoded in double-stranded DNA unzips, transcribes into RNA and subsequently translates into protein. The history of RNAi is a story of how several seemingly unconnected and unexpected phenomena observed in various organisms across kingdoms were finally linked together after discoveries to the trigger and underlying mechanism were made. As it turns out, RNAi is a fundamental mechanism that dates back millions of years where single-celled organisms cleverly employed it to defend themselves against the invasion of foreign viruses. Its modern day discovery started in the late 1980s and early 1990s in plants. At that time plant biologists were attempting to transgenically alter color in petunias by introducing an enzyme that encodes pigmentation in flowers. When the experiment was initially designed the expectation was to see gene overexpression manifested through darker colors (Krol, Leon, Beld, Mol, & Stuitje, 1990; Napoli, Lemieux, & Jorgensen, 1990). Instead to everyone's surprise, the petunias became less pigmented than their natural form producing fully or partially white flowers. This indicated that as opposed to the intended gene overexpression, activity of the enzyme had significantly decreased expression to the point of deactivating the gene responsible for regulating color pigmentation. However, both the underlying mechanism and trigger were unknown.

The story then moves to the fungal community where independently a similar phenomenon of transient inactivation of gene expression was also observed by scientists studying neurospora crassa fungi (Romano & Macino, 1992) and was separately named quelling. History repeats itself again a few years later in the c. elegan worm community where an analogous abnormal phenomenon was also documented. When scientists were attempting to understand the purpose of a particular gene that controls asymmetry in the polarity of embryo cells in the worm, they found much like co-suppression in plants that not only did the single-stranded RNA antisense silence the gene under study so did the corresponding sense RNA strand that was designed as negative control (Guo & Kemphues, 1995). Not long after, plant virologists also found a similar unexpected phenomenon when attempting to improve plant resistance from viral infections that they labeled virus-induced gene silencing (Ratcliff, Harrison, & Baulcombe, 1997).

Although all of these odd observations were not immediately recognized as related to one another, each community, plant and animal scientists, were all independently aware of the phenomenon prior to discovery of its trigger in 1998 by Andrew Fire and Craig Mello (Fire et al., 1998). In fact according to one respondent, the European plant community at the beginning of the 1990s had already started their own network of laboratories working on different plant systems with the aim of joining together and applying for funding to study the phenomenon. In the animal community, more specifically the c. elegan worm community, many had come across the phenomenon in their own experiments while being unaware of the intricacies of the underlying mechanism. Others although not always getting consistent, reproducible and potent results used the precursor technology to RNAi, antisense oligonucleotides, as a tool to inhibit and study the function of specific genes. It was also the topic of discussion at several conferences around that time as illustrated by a respondent who attended a seminar session where Craig Mello was the speaker at a Pew Scholar workshop.

> *"Craig [Mello] shared with us in that workshop in 97, RNAi. I had never heard of it until then and it wasn't about dsRNA it was just a phenomenology that people in c. elegans used as a tool for. It was discovered by Ken Kemphues at Cornell and you make this RNA in vitro, and you inject either the sense strand or the antisense strand into the worm and it would silence gene expression so you could use to basically knock out genes without having to mutate." (respondent 6)*

Fire and Mello's breakthrough insight, notable for bringing the first identification of the causal agent for the phenomenon was recognized by the Nobel Prize in Physiology and Medicine in 2006. It found that double-stranded RNA (Fire et al., 1998), which resulted from contaminated preparations of single-stranded sense and antisense RNA in test tubes as elucidated in the Nobel lecture (Fire, 2007), was the potent trigger to specific genetic interference mechanisms documented several times throughout prior literature. RNA interference was also coined as a consequence of this work.

The story of RNA interference complicates even further with what seemed for a long time to be a completely tangential finding with the discovery of the first small RNA, later labeled as microRNA (Lee, Feinbaum, & Ambros, 1993). Fire and Mello's discovery stipulated that long strands of duplex RNA would trigger the RNA interference gene silencing pathway in worms. However, in more complex organisms, notably in humans, the introduction of foreign dsRNA would instead trigger the self-defense immune mechanism of interferon. Initially when Ambros found the first small RNA, those who should have picked up on the finding did not because they thought of it as just a cute discovery since the specific sequence of microRNA was unique and idiosyncratic to worms and not present in other organisms. But when a second small RNA was discovered and also found to be present in humans (Reinhart et al., 2000) along with the discovery in plants that short sequences of antisense RNA 25-nucleotide in length would silence genes post-transcriptionally (Hamilton & Baulcombe, 1999) the link between microRNA and RNAi was finally made. These findings along with subsequent research in drosophila (Zamore, Tuschl, Sharp, & Bartel, 2000) and eventually in mammals (Elbashir et al., 2001) quickly helped solidify the belief that indeed the gene silencing phenomenon scientists were witnessing was not just a strange occurrence in worms but rather a fundamentally conserved mechanism in many organisms across kingdoms. Moreover, the feared interferon response could be evaded through the use of shorter dsRNA.

Data Collection

The interview process consisted of two stages. I first interviewed two individuals, a board member of a leading RNAi technology based company and a scientist familiar and knowledgeable about RNAi and its history without necessarily doing research in that area. These two interviews better informed me on the community of researchers beyond what was available from archival records and allowed formulation of questions in preparation for the main round with the actual actors partaking in its discovery. The interviews in this first stage lasted 30 minutes on average. They were semi-structured and discussions centered on how to define the community of scientists focusing on RNAi as well as the trajectory of discovery. For instance, one interviewee pointed out that rapid development in the field of molecular biology and genetic engineering, such as DNA sequencing, recombinant DNA, the human genome project, and the hypothesis that life originated from RNA (Gilbert, 1986) were precursors to the discovery of RNAi. Furthermore, they both brought my attention to the fact that most historical expositions of RNAi tend to include observations in plants and fungi that happened beforehand. However, it is not clear whether researchers working with animal models at the time were aware of or even associated their work to these prior anomalous results found in the plant systems. These conversations hence fine-tuned existing interview questions and triggered new ones for the subsequent set.

The second stage consisted of 18 interviews targeting scientists at the heart of the RNAi breakthrough. Respondents spanned model organisms from plants, worms, fruit flies, all the way to humans and included geneticists, molecular biologists and biochemists that contributed to RNAi research conceptually and technologically. They also included one Nobel Prize winner and three Lasker award winners. Each interview lasted between 60 to 120 minutes, averaging about 75 minutes.

The interview questions were semi-structured such that open-ended questions were asked first, followed by more specific and probing ones. I started by inquiring about the line of research each informant was undertaking during the period shortly before the breakthrough was made and covered several other topics from understanding circumstances around and factors leading to breakthrough discovery, to defining and characterizing the community of scientists prior to emergence of the RNAi field, to discussing diffusion of scientific output. The interview question guide is shown in the Appendix.

The community of RNAi scientists was defined functionality by incorporating content search of titles, abstracts and Medical Subject Headings (MeSH) keywords. Because of my need to identify the set of scientists with discovery potential before a well delineated community of RNAi researchers had yet to emerge, it was not surprising that MeSH keywords such as "RNA, Interference", and the same phenomenon of "co-suppression" in plants and "quelling" in fungi did not enter the MeSH lexicon until 2002. To circumvent this issue, I found from extensive archival searches that scientists were attempting to study gene expression regulation or gene silencing by experimenting with both dsRNA and antisense RNA as causal agents. The usage of RNA rests on the assumption that these molecules play a central role in gene silencing mechanisms rather than being restricted to the passive role of carrying genetic information. Consequently, the majority of the community of researchers with RNAi discovery potential was found by searching MeSH keyword terms[1] "RNA, Double-Stranded", "RNA, Antisense", "RNA, Catalytic", "Gene Silencing" and "Gene Expression Regulation" from published peer-review articles in MedLine. I augmented the MeSH search with title and abstract searches so as to include scientists who initially

[1] The exact search string used in PubMed query extracted on October 26, 2011: ((((gene silencing[MeSH Terms] OR gene expression regulation[MeSH Terms]) AND (RNA, double-stranded[MeSH Terms] OR rna, antisense[MeSH Terms] OR rna, catalytic[MeSH Terms])) AND "1980"[Publication Date] : "1999"[Publication Date]) AND English[Language]) NOT interferon[MeSH Terms]. We also found that dsRNA generated a lot of noise as it was heavily used by immunologists studying interferon responses. To minimize the noise from interferon we include in our MeSH search the "NOT interferon[MeSH Terms]" term.

observed the RNAi phenomenon in plants and fungi[2]. I included in this community papers that were published until 1999, since those who quickly published following the 1998 breakthrough paper had similar prospective of breakthrough discovery and were thus included in the sample set. By extracting unique authors from the set of papers obtained above, I isolated a community of RNAi scientists topically based on their publication focus. I obtained a total of 1,551 papers and 3,959 unique authors.

My method of selection for interviewees built on a regression model that attempted to predict the citation count of 1998 publications given prior bibliometric characteristics of each scientist in the sample from a prior quantitative paper. I ran residual analysis to identify scientists who were incorrectly predicted. Specifically, the predictive model employed quasi-maximum likelihood Poisson with cluster robust standard errors to regress citation count of 1998 publications on prior measures of publication history, eminence, brokerage, collaboration, core, specialization, lifecycle, affiliation type (academic or corporate), affiliation prestige and mobility computed from the beginning of each scientist's career until 1997. For the residual analysis, I calculated the error term by taking the difference between the predicted publication impact, *E(Y),* and the actual number of forward citations for 1998 publications, Y_i, as depicted graphically in Figure III-1. The group of interviewees consisted of both the top and bottom one percent of the error terms, thus elite scientists who the model severely failed to predict accurately. Furthermore, no interviewee was part of the Nobel winning team for RNA interference. This sampling provided me with unusual accounts by those who had the potential to make the groundbreaking discovery but ultimately missed it.

[2] The exact search string used in augmented PubMed query extracted on October 26, 2011: ((((cosuppression[title/abstract] OR co-suppression[title/abstract] OR quelling[title/abstract] OR RNAi[title/abstract] OR RNA interference[title/abstract])) NOT interferons[MeSH Terms]) AND "1980"[Publication Date] : "1999"[Publication Date]) AND English[Language].

The above method identified a total of 19 scientists with research focus in RNAi. The reason for such a low number of interviewees from the initial large sample of authors used in the regression models is three-fold. First, my technique of selection using top residuals cut out a vast majority of the initial sample. Second, because I am studying a nascent field that significantly grew in size only after the breakthrough occurred the number of scientists at the beginning was extremely limited. As a respondent explained, it took many years from the mid-1990s for the community of RNAi scientists to become significant in size.

> *"It's not that many actually, if you actually get into the number of people who jumped into the field, it's still rather small, I mean the number that worked on the mechanism of it. It took fifteen years to become a field of 500 people or so." (respondent 12)*

Third, because the period of interest is prior to the breakthrough no defined RNAi community existed. Therefore, in order to insure comprehensiveness in the sample of researchers with breakthrough potential a substantial amount of noise was picked up when building that set using existing MeSH keywords heavily employed by scientists working on peripheral fields such as interferon. As a respondent who had worked with the precursor technology to RNAi but later diverted into studying interferon shared the same frustration of filtering through noisy search results prior to RNAi entering the MeSH lexicon in 2002 described,

> *"I think you may have used […] a search term that's been hijacked by the [RNAi] community. Because exactly like you, sometime in around 2000 and 2002 every time I tried to see what was new in my field, I would type in dsRNA and would get all these RNAi stuff. And you have to sit there going, no that one, yes this one." (respondent 7)*

I augmented this set of interviewees with scientists who attended RNAi related conferences – Keystone Symposia and Gordon Conferences – at the beginning of the field, which increased the number of interviewees to 27. Of these 27 scientists I reached out to with interview requests 18 responded positively. During the interviews, I inquired about other potential individuals within the RNAi community at its birth stage that my interviewee would recommend I meet to validate my selection technique. Their suggestions were all amongst the sample of 27 interviewees I identified with the selection methods described above.

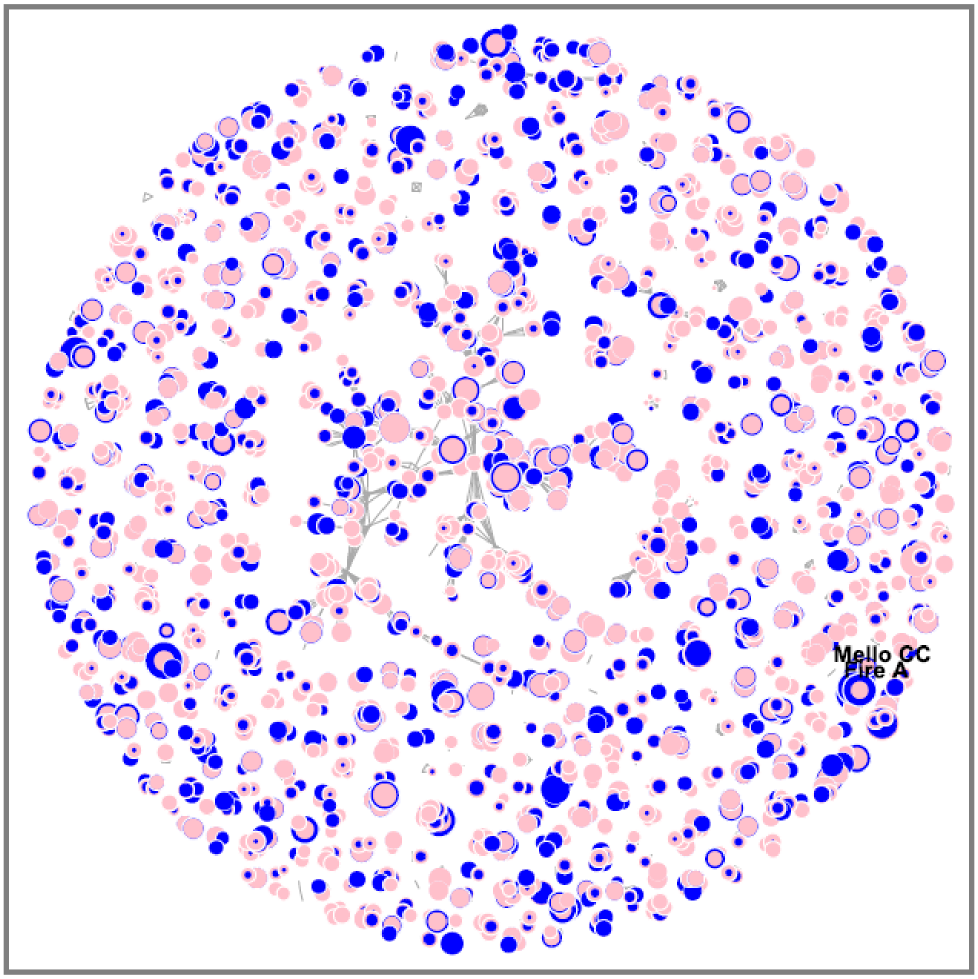


Figure III-1 – Network plot of RNAi community with nodes representing each scientist, link as co-authorship relationships. Blue represents actual impact and pink represents predicted impact.

Data Analysis

Analysis of each interview once transcribed verbatim was conducted in line with coding principles set out by qualitative researchers (Miles & Huberman, 1984). I first open coded all interviews by describing each excerpt, such as ‘attended conference’, ‘ignored mechanism’, ‘used RNAi as tool’, ‘double-checked in another organism’, ‘described antecedent to RNAi’, ‘explored at the fringe’, etc. When new data did not fit a previously identified code category I created a new category. Once I finalized the open code for all primary interviews, I proceeded to axial code the open code categories. Two salient classes emerged: barriers to breakthrough and actions scientists took to circumvent barriers. A third category included all other breakthrough related narratives such as the historical context. The two salient classes were further divided into three barriers (as well as instances where the barriers interacted with each other) that correspond to the three themes that finally emerged: being blinded by conventional science from framing barriers, being constrained by current dogma from paradigmatic pressures, and being unable to connect the dots due to boundary barriers. Also for each theme, I obtained a collection of practices that scientists put in place to circumvent barriers to breakthrough.

iv. Findings

Although my interview questions were mainly probing on circumstances that led to breakthrough discovery, the salient themes that emerged centered on how a number of scientists were on the verge of breakthrough several times but missed the seminal discovery. In other words, these results center on uncovering explanations behind the counterfactual of missing breakthroughs. My novel finding is that this delay in discovery was not only due to struggles in solving a particular problem but also because of difficulties in identifying the problem, in assessing the potential impact of the problem as well as proposing a drastically different theory than stipulated by current paradigm. Thus, throughout the discovery process that I divided into problem identification and problem-solving stages, those on the verge of discovery suffered from failures to identify and propose breakthrough opportunities. The results suggest that at the basis of this failure underlies a cognitive mechanism hinged on institutional underpinnings stemming from three barriers and their interactions. During problem identification, scientists who missed the opportunity suffered from being blinded by the pursuit of normal science by framing anomalous observations along established technologies. During problem solving, they were held back by paradigmatic pressures of being constrained by current dogma by interpreting abnormal results according to established paradigms. Established boundary barriers between communities of scientists compounded both effects as they prevented similar anomalous patterns in various fields from being connected together, thus, respectively, hindering pattern recognition and pattern labeling. Table III-1 contains quotations illustrating each barrier from all 18 respondents that I interviewed.

Problem Identification Failures

Framing Barriers

Because most scientists came in contact with the phenomenon of RNAi as a technique to silence genes in their pursuit of hypothesis driven science prior to understanding the actual biological mechanism that underlies the concept, their views of the phenomenon were biased toward a useful technology rather than a topic of inquiry worthy of scientific merit. Path dependence from prior technologies reinforced the belief that the phenomenon of gene silencing is a technique, which cognitively biased and ultimately delayed discovery of its trigger. Underlying institutional logics in science where researchers were blinded by the pursuit of normal science triggered this cognitive bias. Being able to use the technique to accomplish the end goal of inhibiting specific genes mattered more than understanding why the technique worked. Thus, before even attaining the problem-solving stage, researchers were unable to identify the interesting and potentially groundbreaking problem to be solved, and subsequently passed on the breakthrough opportunity. As described by a respondent below,

> *"My sense from [others] was that they just looked at this like a bizarre tool, they couldn't explain it but it was fabulous for what they wanted to do. They could silence genes. […] They were focused on the thing at hand and kind of ignoring this elephant in the room, which was far more important and interesting." (respondent 6)*

Most researchers valued the phenomenon's ability to inhibit specific genes without having to rely on mutations. It was a means to an end rather than the end itself. This behavior is in line with Kuhn's (1962) prediction that scientific research is extremely productive at expanding the central paradigm but also self-reinforcing during periods of normal science. Case in point, the two scientists, Guo and Kemphues, who first observed the phenomenon in 1995 in worms explicitly chose not to study why it worked. One of the two scientists explained following their observation of the anomalous gene silencing phenomenon, "once we knew it was a gene specific effect we didn't really care how it worked. All we cared about was that we could use it." (respondent 10) They reported the strangeness that the control in the experimental design, sense RNA, had a similar potent effect as the treatment, antisense RNA, in silencing a gene they were studying, but decided that it was not worth following up.

This cognitive bias of focusing on the tool application of the phenomenon rather than understanding conceptually how it worked was path dependent and stemmed from the historical context of precursor technologies. In the late 1980s, large groups cornered the market in being able to produce knockout mice. They controlled the technology of making mutated knockouts in genes, all subsequent downstream phenotypes as well as the distribution of mice. This made it very hard and expensive for small laboratories to obtain such knockout samples for research purposes. Therefore, there was a large culture of people that were praying for antisense oligonucleotide technology to be the answer because it meant they could do things much faster and much more quickly than by mutation. The demand for such a gene silencing technology in the research community was very high because researchers would be liberated, as they would "no longer [be] restricted to mice and would not have to collaborate or beg for the mice to do things." (respondent 7)

This bias also explains why, despite many observations of the bizarre phenomenon by various groups of scientists, surprisingly little racing was present in the community to solve the puzzling mechanism. Because the nature of knowledge RNAi embodied was perceived as a technique of how rather than a demonstration of why, scientists did not consider solving the intrigue around the RNAi phenomenon as a priority-based incentive (Merton, 1957) and were therefore preoccupied with other scientific endeavors that met this criteria more explicitly. Besides Fire and Mello's groups working with c. elegan worms and actively attempting to solve this puzzling gene silencing phenomenon, only plant scientists were working on explaining the same mechanism (Waterhouse, Graham, & Wang, 1998). Competition, instead, intensified *after* the pathway's trigger was found as described by the two respondents below, the first working on animal models and the second working on plant models.

> *"For the actual initial discovery that you can introduce duplex RNA into cells to specifically inactivate genes, Fire and Mello were ahead of the game in that case. But once that discovery was made and the transition made to studying the mechanism and the factors involved, that's when the real competition came in." (respondent 5)*

> *"At the end of the 90s and beginning of 2000 it was really difficult, because all the things that could be found simply were found at the same time, in a range of a few months." (respondent 14)*

Following Fire and Mello's discovery of its trigger, RNAi was now established to be an open and interesting scientific question to research, as assessed by a respondent, which is in line with the norms of the scientific institution rather than a mere bizarre phenomenon used as a tool (Dasgupta & David, 1994; Merton, 1957).

> *"What Fire and Mello did is that they discovered that RNAi was real biology. Because, first of all, most people thought that the silencing phenomenon back then reported in plants and in worms, were weird things that would probably turn out to be artifacts later and they have the feeling of homeopathy." (respondent 16)*

When unexpected results appear in tangential elements not affecting core hypotheses of the research project, whether manifested in the tool or the experimental results, the decision of whether to follow and inquire deeper into a weird but interesting observation or to stay with the experiment at hand is very difficult. In particular, time and resource constraints together with the low probability that the oddity will eventually turn out to be something influential make it an especially hard decision, as often times they turn out to be mere artifacts. Consequently, blinded by the institutional barrier of pursuing normal science most ignored the weird observations and carried on. However, whenever such abnormal observations occur it is often precisely under these circumstances where breakthroughs are most likely to be discovered. As the Nobel laureate I interviewed described,

> *"When you have a well-defined system and it's telling you something you don't understand, it isn't consistent with the way you've designed the system then something is new in the system. It's paying attention to that [bizarre phenomenon] and not pushing it out of the way as you went towards your more conventional hypothesis driven science. There is a new science there. To ignore that, to do conventional science is what most people will do. […] That meant the difference between the genius and good science" (respondent 13)*

Boundary Barriers

Also present in the problem identification stage is the boundary barrier between disparate scientific communities. The history of RNAi's discovery is punctuated by several documented observations of the bizarre phenomenon first in plants (Napoli et al., 1990), then in fungi (Romano & Macino, 1992), worms (Guo & Kemphues, 1995) and plant viruses (Ratcliff et al., 1997), and perhaps even more instances of undocumented observations before the underlying trigger agent was finally found. Tracing through citations that the latter three papers refer to, I found a clear dichotomy between the plant/fungus scientists and the worm scientists. Both the 1992 fungus and the 1997 plant virus papers cited the initial 1990 plant paper, whereas the 1995 worm paper only cited works in the worm community and did not cite neither the 1990 plant nor 1992 fungus papers. Similarly, the 1997 plant virus article did not cite the 1995 worm paper (see Figure III-2 for a graphical depiction).

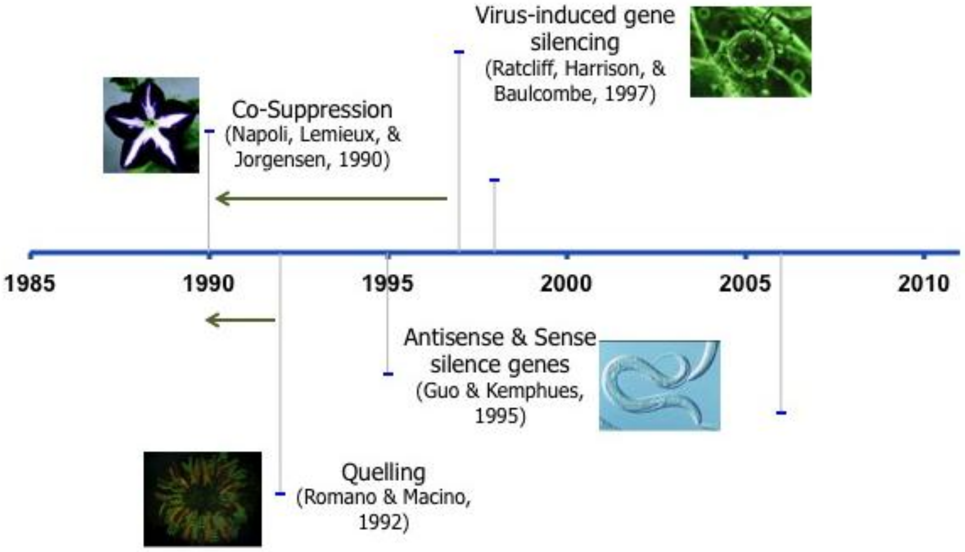

Virus-induced gene silencing
(Ratcliff, Harrison, & Baulcombe, 1997)
Co-Suppression
(Napoli, Lemieux, & Jorgensen, 1990)
1985
1990
1995
2000
2005
2010
Antisense & Sense silence genes
(Guo & Kemphues, 1995)
Quelling
(Romano & Macino, 1992)

Figure III-2 – Citation pattern of major gene silencing papers prior to Fire and Mello discovery of the trigger mechanism to RNAi in 1998.

These citation patterns and the independent results stemming from the plant and animal communities suggest that within the boundary of each community information flowed easily, but between communities diffusion was sticky. Although several observations of a similar anomaly were made in various organisms and fields, they were brushed away as a weird phenomenon that happened in the particular model organism employed. Thus, these boundary barriers lead to two levels of discontinuity, either scientists just did not know about the prior works from different communities, or for the few that did know they did not see the connection between anomalies from others beforehand. The following quote illustrates this latter discontinuity.

> *"Rich Jorgenson and Carolyn Napoli, they were telling me stories about silencing they put in. They had these flower color things trying to get purple it would turn white, it was all screwed up. But I missed it entirely. I did not see the connection." (respondent 12)*

As a consequence, scientists were unable to connect the dots and identify a repeated pattern of weird results that would provoke crisis and revolutionary changes to the established scientific paradigm (Kuhn, 1962). Had scientists made the link between similar observations in different organisms the likelihood of dismissing their one odd observation would have been lower. Two boundary barriers – disciplinary and/or organismic boundaries between scientific communities act as natural barriers to information flow, and partial opacity in academic publications – were at the basis of this inability to connect the dots.

Just like membranes in biology, organizational boundaries form natural barriers to the diffusion of information (Kogut & Zander, 1992). Similarly, for the open community of science, the flow of knowledge is deterred by the boundaries of various scientific communities, which in turn hindered scientists' ability to connect the dots. For instance, aside from citation evidence presented above, most informants I interviewed working with animal model organisms were unaware of the research done by plant and fungal scientists, and vice versa. Prior to Fire and Mello's discovery in 1998, collaboration and communication between plant and animals scientists working on gene expression and inhibition were little to none. Most researchers from the disparate model organism communities did not meet until after the links between their works became obvious. Two animal scientists describe how they perceived the plant community.

> *"The plant world tends to have its own group of people, and they don't tend to intermix too much with the non-plant people." (respondent 11)*

> *"I am on the virus division for the society of microbiology so our job is to organize annual meetings in virology for Europe, and we have a plant virus section. But they might as well be lectured in Latin; they don't really integrate with other people. […] RNAi is an absolute classic, they had stuff going on they probably thought it was very interesting but they probably didn't think others would want to know. They didn't think animals did the same thing." (respondent 7)*

Analogously from the perspective of a plant scientist, the divide between the two communities was described in a similar fashion. In fact, the plant community working on solving gene silencing was surprised that Fire and Mello's paper was published in Nature, illustrating how disparate the two communities studying the same gene silencing phenomenon were prior to the 1998 RNAi breakthrough.

> *"But for the animal people, I understand that it was really a breakthrough to consider that long dsRNA could trigger something because they were not anticipating this. In plants it was not really a breakthrough it was completely expected. That's why the discovery in the same year, in 1998, by two groups [studying plants] that dsRNA could induce very efficiently silencing and actually much more efficiently than sense and antisense directly was just the next step of something that was going on for ten years. […] In plants it was more continuous." (respondent 14)*

Furthermore, scientific literature is often blamed to some degree with a lack of transparency in published work as negative results are usually omitted. Instead of codifying knowledge, some know-how especially negative results remain tacit and significantly slow the pace of diffusion. This also hinders scientists' ability to connect the dots. Due to the priority-based reward system, scientists are afraid to get scooped or lose a race to first discovery because negative results are often sources of crucial information that may aid competitors by providing a map to success (Fleming & Sorenson, 2004). It prevents others from wasting time and resources going down unsuccessful paths and increases their research effectiveness and discovery potential. An informant scientist using an imagery of science solving a giant puzzle piece vividly describes this problem:

> *"It's almost like each of us has a little piece of the puzzle but by the time we are ready to show the puzzle piece to the audience we've filed off some of the pieces we don't like about it and now of course it doesn't fit. The other guy has got the other piece of the*

> *puzzle but of course it doesn't fit cause we have changed the shape of it." (respondent 16)*

Interplay between Framing and Boundary Barriers

During problem identification, not only were both barriers present but the interaction between the two also exaggerated the effect of misidentifying the problem. Being blinded by normal science prevented scientists from pursuing the underlying science behind gene silencing. Prior established antisense technologies, which was used as a tool to knockout genes, influenced most scientists to frame the phenomenon as a technique and led them to ignore it as an interesting subject of scientific inquiry. Instead, they were more interested in it as a means for other research topics and thereby failed to identify the interesting problem to solve. This failure in detecting the right problem to study was further compounded by boundary barriers between scientific communities that prevented anomalous observations from various fields to be linked together. Scientists were unable to recognize a repeated pattern of anomalies. Without such a critical mass and skeptical that one single anomalous instance is unique to their specific research setting contributed to inaccurate assessments of the scale of the potential breakthrough problem.

The top portion of Figure III-3 graphically summarizes the pattern recognition failure in the problem identification stage. As one respondent from the plant community described an encounter where she discussed the gene silencing mechanism with an animal scientist in the mid-1990s,

> *"I remember talking to a guy in Vienna who was one of the first big people making transgenic mice and he just scoffed at the whole idea that you'd see something like that. But even at the time [...] there still wasn't a lot of people coming together and thinking that might all lead to a single mechanism." (respondent 18)*

In fact, most of the animal community up until the Fire and Mello discovery was stuck at this stage,

> *"When people tried it and it worked, it was like ok let's work with it. Very few people thought it was worth studying, but everybody wanted to use it. So then you'd go to the worm meetings and everybody was using it." (respondent 12)*

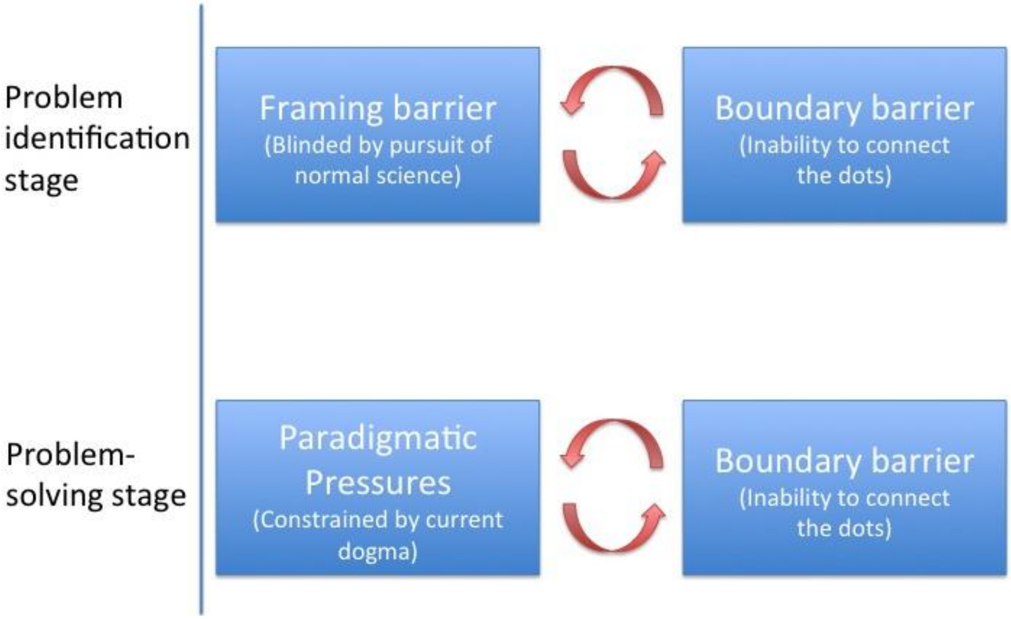

Problem identification stage
Framing barrier
(Blinded by pursuit of normal science)
Boundary barrier
(Inability to connect the dots)
Problem-solving stage
Paradigmatic Pressures
(Constrained by current dogma)
Boundary barrier
(Inability to connect the dots)

Figure III-3 – Framework of Missed Breakthrough

Problem-Solving Failures

Paradigmatic Pressures

For those who saw the phenomenon of gene silencing as a scientific endeavor worthy of pursuit, another barrier to breakthrough discovery from pursuing normal science is that scientists were constrained by current dogma when called upon to interpret unexpected results that often did not fit within the confines of current theories. Since science is preoccupied with truth seeking, scientists take great pains in ensuring that the results they present are correct (at least within the state of current knowledge and experimental techniques available at the time), instead of risking the publication of artifacts that would eventually be disproven. To avoid being wrong when faced with weird results and lacking psychological safety (Edmondson, 1999), they often chose to ignore anomalies and dismissed them as artifacts so as to avoid challenging established dogma, and, in turn, throwing away valuable opportunities for breakthrough discovery.

Scientists were confined by social and institutional factors of science that underpinned and triggered cognitive barriers. For RNAi, the difficulty in explaining the observed silencing phenomenon and identifying the causal agent stemmed from a disparity in causal pathways between the RNAi mechanism and the central dogma of molecular biology. In the central dogma, both double-stranded DNA and single stranded RNA had salient roles for long and short term information storage respectively, while most biologists at that time were brought up to believe that dsRNA was inert. There was no place left for double-stranded RNA as Fire stated in his Nobel lecture (2007). No one believed that dsRNA should work better than antisense because if you had injected an antisense provided it did not degrade, it would have found its target and taken it out. The conventional thought when inserting pre-annealed dsRNA was that it had to unzip, which was a weakness because nobody realized that there was actually machinery that accomplished that. However, it turned out that dsRNA was indeed the trigger agent in the RNAi mechanism. Thus, scientists had to get over a socio-cognitive barrier from being encultured (Simonton, 1989) in the molecular biology community with a dogma that contradicted the ability for both sense and antisense RNA strands as well as dsRNA to perform equally well in silencing gene expression. This belief reinforced the established paradigm from the central dogma, which in turn constrained scientists from interpreting their results using a revolutionary framework even when a weird and interesting problem had been identified and pursued as a path of inquiry. Given the state of knowledge at the time, two informants illustrated how implausible dsRNA was seen as a trigger to the RNAi mechanism.

> *"Nobody would ever inject the sense strand cause psychologically you could imagine how the antisense strand could work with the base pairing but the sense didn't make sense even though they showed they both worked equally well. No one ever did the*

> *sense strand cause they just thought that just can't be right. They just kind of ignored it and thought it's antisense." (respondent 1)*

> *"It's weird and not expected because basically we all knew that we make dsRNA and that's a dead end, it's an inhibition of the other RNA, you can't use that to make something." (respondent 15)*

Boundary Barriers and Interplay with Paradigmatic Pressures

Similar to the problem identification stage, boundary barriers were present during the problem-solving phase. This boundary barrier is driven by scientists' belief of how fundamental the phenomenon of gene silencing traces back to a common ancestor between animals and plants. As an informant explained,

> *"We know that plants evolved as a multicellular life forms independently from animals, so the last common ancestor of plants and animals was a single cell organism. And so when you're talking about how the cells are organized and develop, that happened independently. […] So when you're talking about very fundamental processes that were there in the last ancestor, last single cell ancestor, those operate across kingdom. So in general it just depends on whether you think it's an ancestral process or whether you think it's more derived." (respondent 3)*

For those who saw the scientific merit of pursuing research on gene silencing, boundary barriers also aggravated the institutional pressures from the current paradigm as illustrated in the bottom half of Figure III-3. Paradigmatic pressures coupled with boundary barriers reinforced each other in contributing to the failure of identifying and proposing a breakthrough opportunity. If one is faced with unexpected results in one single research setting and is unable to gain more confidence from similar results in other settings due to boundary barriers, her ability to think in a revolutionary manner is compromised and she stays locked within the same mindset. Analogously, if one is constrained by current dogma and does not consider the possibility of a groundbreaking perspective, substantiation in other organisms and fields will not be sought out. In both cases, crisis will be missed and breakthroughs overlooked or delayed. An informant describes this reticence that if one is alone in finding a contrarian result it is very hard to muster the courage to submit it for publication without having substantiated it somewhere else.

> *"Cause if you think about it if you were sitting in a lab in the middle of nowhere injecting dsRNA into c. elegans, and seeing it having an effect, a really good effect, a really strong effect on gene expression and it doesn't work with single-stranded RNA, and no one has ever seen this before, you can't write this up. You must have put out a few fingers to see, whether anyone have heard of anything before." (respondent 7)*

Unlike the animal community that was caught in the problem identification phase, most of the plant community was trapped in this problem-solving stage,

> *"Why didn't the plant people get to where Fire & Mello did? My main insight is that we were so focused on transgenes to manipulate DNA expression. We never got to introducing RNA, it was regarded as unstable, that was never going to work. […] So there are these different mindsets that are so ingrained that you don't even appreciate that there is another way to look at this. And I think that's why we were really locked into that. It traced back to the discovery of DNA and the genetic material and the structure of DNA." (respondent 17)*

Table III-1 (following nine pages) – Informant quotations respectively illustrating framing barriers, paradigmatic pressures and boundary barriers that contributed to the delay of the RNAi breakthrough discovery.

Resp #	Model Organism	Framing Barriers	Paradigmatic Pressures	Boundary Barriers
1	c. elegans	It's really puzzling, people had just filed this away and just thought this doesn't make sense but didn't think about it. They just kind of ignored it and just thought it's antisense, people used to call it antisense even though it wasn't.	It was funny because the sense and the antisense strand both worked. Nobody would ever inject the sense strand cause psychologically you could imagine how the antisense strand could work with the base-pairing but the sense didn't make sense even though they showed they both worked equally well. No one ever did the sense strand cause they just thought that just cant be right and dsRNA hadn't been shown to have any effect. People talk themselves out of doing experiments all the time. They'll say I won't try that because it will probably look like this. Maybe. Maybe not! If you don't do it you never will know.	No one at that time, no one I had talked to was even thinking that it was related to the plant things. No one before 98 I had ever heard anyone mention anything to do with plants.
2	c. elegans	We were obviously intrigued by it, but we could use to probe some biology that we were interested in it. And you want to do in science, it's almost like you see something and you want to harvest it. So we could harvest RNAi in a way by using it as a novel method, it allows you to leverage some biology. You didn't have to get mutations and you could get some information and learn something about it. The community started to adopt it as a method, because they knew it was specific.	It's hard to do these breakthroughs where you really have to step beyond your comfort zone.	We were trying to penetrate what we thought what we thought was a novel phenomenon. We didn't believe that it really represented anything general. If you have, if your suspicion let's say the weight of your suspicions is that it's probably kind of worm specific, what's the point of devoting a lot of resources to it. Because we are trying to figure out things that are general and broad right. So the fact that the worm could do this and that other things couldn't do it. I mean flies don't do it, and it's

				inconceivable that mammals would do it. So you’re thinking it’s a worm thing.
3	drosophila	Scientifically you know that this is working and these people were just using this as a tool. Then you have to decide ok. On the one hand, this is just a tool and the reason you’re using this tool is because you want to study the biology of these genes and you’re really focused on that biology and so you’re convinced that using this antisense method is teaching about the function of those genes and you go on and you focus on the function of those genes. And, you don’t get distracted by this oddity that the sense is also working.		I didn’t know that the plant phenomenon would be related to the worm phenomenon. Obviously the people working in plants, were in fact trying to explain the same phenomenon but we didn’t know any of those details. So when you’re talking about very fundamental processes that were there in the last ancestor, last single cell ancestor, those operate across kingdom. So in general it just depends on whether you think it’s an ancestral process or whether you think it’s more derived.
4	plants	I mean a lot of people have been doing the sorts of experiments, putting genes in viruses into plants. A lot of people had been seeing exactly the same thing, they had seen that the transgene that conferred resistance was not expressed that it was silenced. And they just ignored it. So people sometimes ignore data when it stares them in the face.	[In plants] it was largely phenomenology, there wasn’t a lot do to with the mechanism. However we had done some experiments that implicated dsRNA. We had done 2 experiments one set of which never got published. We did submit them for publication and then they came back and the editor said they were of insufficient general interest. And so the reviewers were not convinced that, they thought that it’s an interesting illustration of how a field can get preconceptions.	We were aware of the worm story to some very small extent, partly because what we knew was Ken Kemphues’ original micro injection experiments. So we knew about those and those looked like some sort of co-suppression phenomenon and also I knew about Ruvkun and Ambros’ work. We missed the link between that work and our work, so the small temporal RNA, the Ruvkun and Ambros work so that looked as if it were a translational regulation thing. I don’t think there were really conferences that brought the animal and plant fields together until probably as late as 2001.

5	c. elegans	People who were using antisense were using it to inhibit genes so they could show in vivo or address inactivity in a gene in vivo.		I think the effects of antisense were not that satisfactory cause they were not that potent and hard to control for. So again it gets back to this question, yeah if you have a control that doesn't make a lot of sense, you are not going to report it. Cause there are probably a lot of observations that were not, experiments that were not actually included in papers. You know one of the problems with science is that negative results often do not go reported. And they are left un-described.
6	drosophila	They just looked at this like a bizarre tool, they couldn't explain it but it was fabulous for what they wanted to do. They could silence genes, so it was kind of like this. They were focused on the thing at hand and kind of ignoring this elephant in the room, which was far more important and interesting.	Craig got up to share with us that workshop in 97 RNAi. We thought this was really bizarre. I remember being there; everyone in the room was bedazzled, because I was so bizarre. I ran counterintuitive to everything we've been taught. We knew [the experiment] had worked. It's like holy shit, although you're really scared that you're over interpreting it or something.	The quelling people were kind of off on their own; they didn't interact very much with us. There is this really bizarre phenomenon but it never occurred to me at that point that it would be applied to other organisms. So it never occurred to me at that time that I should try RNAi in drosophila.

7	c. elegans	Everybody was wanting a way to knock a gene down. So it was a receptive community, when people saw that they thought that's very interesting. And I don't think it was long before everyone was trying it out. Everyone could see the application of it, everyone was already primed to apply it.	And you know, a lot of people had described similar things or talked about it certain things. But nobody took it terribly seriously. Why should dsRNA work better than antisense. Because if you had sticked in an antisense providing it doesn't get degraded, it should find it's target and take it out. When you put in pre-annealed dsRNA we thought it had to unzip, it's actually a weakness, because nobody realized that there is a machinery that does that.	[The plant section] might as well be lectured in Latin, they don't really integrate with other people. They didn't think animals did the same thing. So there you will find that there is a subculture of plant people that are just doing stuff. If you were sitting in a lab in the middle of nowhere injecting dsRNA into C. Elegans, and seeing it having a really strong effect on gene expression, it doesn't work with single-stranded RNA and no one has ever seen this before, I can't write this up. You must have put out a few fingers to see, whether anyone have heard of anything before.
8	c. elegans	So I was also able to also use the technique to inhibit that gene activity and see.		We didn't even know that, it would become such a general phenomenon. No one knows because we thought that it is something peculiar with worm. Actually I was going to discuss some of these [plant results] in my original paper but my advisor felt it was a little too premature to make that kind of link.

9	mammals	So we used gene inhibition technologies in order to understand the pathways. But at the same time we also worked on developing this type of technologies, so I think I was more or less among the first, the ten first, to use antisense oligonucleotides which were very popular, but much less efficient than RNAi.	Who is going to think let’s put a double-stranded short RNA, if you don’t know the system, who is going to say let’s put a short double-stranded RNA in a cell by chance and it could be something. It’s impossible. It was so anti-dogmatic, because there was DNA, RNA and protein. I guess it took time also and very bright and inventive people to really go against the dogma. And say ok, maybe something is wrong. It is always very difficult.	At this time, I had never followed what was going on in plants. Preconception that whatever perhaps happens in plants is different in animals even in mammals, that there wasn’t much attention paid to them, even within the community. The same thing happened with c. elegans in a way. Because at the beginning everybody thought, ok this thing is interesting but most of them thought just in c. elegans.
10	c. elegans	And so we never asked the question in a serious way other than talking out of this work. So that we then continued to use the technique because it was clear that one of the key control experiment was to show that it wasn’t any old RNA that did this effect so it was a very gene specific effect and so once we knew it was a gene specific effect we didn’t really care how it worked. All we cared about was that we could use it. Everybody was very excited about it because of the potential for its use to target specific genes without going through the trouble of making mutations. People were intrigued but that’s different from going after it.	Because I didn’t have that information, that knowledge. I wouldn’t have made that connection it never occurred to me that there was both strands in our reaction.	As far we knew at the time, it was a very specific phenomenon for c. elegans. It was pretty much just thought of as a c. elegans phenomenon at the time.

11	mammals	Fire and Mello were trying to do an antisense. So people were using antisense oligonucleotides for a long time to try to do what RNAi does.	dsRNA molecules were not typically viewed as being naturally occurring molecules. They were typically viewed as being part of virus or whatever.	The paper by Jorgenson on petunias in plant cell, the name of the journal was Plant Sciences, Plant Cell or something. That really nobody followed, nobody. It's interesting the plant world tends to have its own group of people. And they don't tend to intermix too much with the non-plant people.
12	c. elegans	Cause it's a tool that everybody wants to use like recombinant DNA, many people who wanted to use it don't care how it works it just becomes a tool that they use. When people tried it and it worked it was like ok, let's work with it. very few people thought it was worth studying. But everybody wanted to use it. So then you'd go to the worm meetings and everybody was using it.	First reaction was: it can't be right, it's too weird.	Rich Jorgenson and Carolyn Napoli, they were telling me stories about silencing they put in. They had these flower color things, trying to get purple it would turn white, it was all screwed up. But I missed it entirely. I did not see the connection.
13	drosophila	But they were so focused upon the objective they were studying – was this gene required for this mutant phenotype. It confirmed the issue they designed the experiment for that was their objective. So they went and published the results, saying this is the function of this gene. But what they didn't do it that they didn't say that this control that didn't work is likely to be more important than this paper. And we should put aside the results of this paper and pursue that control.	Yes, and there was DNA, RNA and protein. And then we started to get information from RNA to DNA. It started to be a big change that RNA also could be considered as an information and not just as an intermediary.	

14	plants	You could silence genes, which was a tool that could be valuable.	We are not really discovering a new thing, in fact what we are discovering are things that already existed but that we are simply ignoring or that we underestimated. Also you discover it was something really different than the dogma that dsRNA played a role in animals.	Here we are too isolated; we don't have enough interaction with people because we are only working on plants. We are experts on plants but we are only working on plants. There was not that many meetings that mixed organisms. So we extended this from plant to fungus, but still in 1996 there was nothing published on animals, at all.
15	e. coli	Even though nobody knew really how it worked, the success rate was enormous. You could use it. And there were all these patents in the early. Because you could try sense or antisense, or you use a sense gene or antisense gene generally speaking it worked.	Almost none of us thought it would be the discovery that dsRNA is the trigger, that is something we did not expect. It's weird, not expected because basically all we knew that we make dsRNA and that's a dead end, it's an inhibition of the other RNA, you can't use that to make something. They didn't really know what to do with it cause dsRNA doesn't do anything.	The people who should have picked up on this and Victor [Ambros]' discovery in the beginning they didn't because it was a worm thing, worms are doing strange things.
16	c. elegans	Somebody took me aside and said whatever you do don't just work on RNAi, because it's not biology it's just a technique.	The hypothesis that were successful for you in explaining phenomenon A, kind of get recycled as the first choice in explaining phenomenon B. Because it's what you're most comfortable with and you know how to test it and if it's wrong you now have this set of experiments that helps point you in the right direction. So I was just fascinated with the idea that dsRNA could do anything, I had been brought up to believe it was inert.	Because I have always thought that the real barrier to productivity in science was people not communicating immediately when they see the common thread. Once you wait until your discovery is polished and presented then you've already filtered out some of the things you don't understand that the right person can explain to you because they have the other piece of the puzzle.

17	plants	You have a particular objective you want to understand X you want to solve that you’re using hypothesis testing, I think that turns out to be kind of a trap.	So there are these different mindsets that are so ingrained that you don’t even appreciate that there is another way to look at this. And I think that’s why we were really locked into that. It traced back to the discovery of DNA and the genetic material and the structure of DNA. We were introducing constructs that it turns out in retrospect did make dsRNA. We were thinking in terms of RNA being produced and then what happens to it well it get degraded, and I always thought well it gets turned over so who cares what the degradation products look like. We were manipulation DNA not RNA. That was the one missing piece, had we gone into introducing RNA directly we could have done things like Fire & Mello did and we could have done it years before them.	We were publishing in different journals then a lot of the animal folks that yeast folks wouldn’t see if they were at the wrong kind of institution, and that created an artificial barrier that doesn’t exist now but was an important one then.

18	plants	So I would say that in that case we were trying to do something different, this gene replacement, but in the process of doing those experiments we stumbled upon this gene silencing and at that point it was so interesting, it seemed so new and not explainable by anything that we had known before that we had started focusing on that phenomenon. Everybody wanted to use this technology first as a technology for research for knocking down a gene.	We were subconsciously ignoring a lot of science. We were also testing with what kind of thing do you need to trigger this gene silencing and we had already setup this experiment to test this that there would be some kind of RNA signal involved, and results at the time also suggested that it was likely to be dsRNA.	There wasn't a lot of dialogue then between the plant and animal community. […] And at the time I don't think we were thinking too much about necessarily the animal work. But during the initial years when we were working on it I think we weren't talking with animal people very much, it was more just a small group of plant scientists who were first trying to figure out what was going on. Plant scientists find that a lot of animal scientists don't take you very seriously. But there are so many fundamental biological findings made in plant systems beginning with Mendel and his peas and the genetics. But we sort of felt like we were on the side, the animal people would always listen to animal people.

Success in the Discovery of the First RNAi Causal Trigger

The above analysis begs the question of why did Fire and Mello discover the revolutionary mechanism to RNA interference and why not someone else? In short, Fire and Mello were able to surmount all barriers that incite failures to identify the opportunity for potential breakthroughs.

First, although Fire and Mello also first came in contact with the phenomenon from a tools development perspective while trying to inactivate genes using antisense oligonucleotide technology, they quickly realized that the phenomenon itself was interesting, important and worth studying. Instead of dismissing it as just a useful tool or a mere worm oddity they believed that it was a fundamental process conserved in other organisms. They were not blinded by the pursuit of conventional science, and were able to explore the phenomenon. The following quotes exemplify both Andy Fire and Craig Mello's motivation to study the phenomenon from the point of view of their colleagues.

> *"[Andy Fire] has always […] said look I think I can figure this out, and sometimes it's boring stuff, but he just latches on and keeps going." (respondent 12)*

> *"Craig [Mello] was very excited about it and he just wanted to figure it out. He thought it was fundamental, and he was right. He believed […] that if he figured it out, he would have done something good." (respondent 10)*

> *"Craig [Mello] believed that it's something. When you think about it, you say why would somebody entertain the possibility that what they are seeing is something broadly conserved. You have to have a deep, a serious amount of faith that there are uncovered phenomena in the life sciences." (respondent 2)*

Second, the fact that Fire and Mello were able to see passed the inertness of dsRNA they were taught to believe throughout their academic careers up to that point was an indication that they circumvented the constraints established by current dogma to propose their theory of RNA interference, and were less encultured in the current thinking of their fields.

And finally throughout their discovery process, Fire and Mello were well aware of the work done by plant scientists and were able to connect the dots between these works, those from the c. elegans community and the results that they observed from their own experiments. From the Nobel paper citations which made reference to several related articles in plants, Fire and Mello were not only aware of phenomenon in plants they also believed that it was similar to what they had discovered in worms.

Fire and Mello's success which hinged on surmounting all three barriers as necessary conditions to breakthrough discovery provides further evidence that these barriers cannot be viewed independently, but are rather interconnected and interact with one another.

v. Implications to Bibliometric Literature from Remedial Practices

In this section, I move away from the above inductive framework on mechanisms of breakdown throughout the breakthrough process towards a set of deductive propositions from remedial practices that scientists employed to circumvent the barriers. Although these propositions are extracted directly from my data and have theoretical foundations from prior research, they require more empirical validation as they have yet been operationalized as sources of breakthrough in traditional measures of bibliometrics. Further, it is important to stress that all scientists did not perform these practices. Instead they reflect probabilistic central tendencies that are necessary but not sufficient for breakthrough discovery.

Circumventing Framing Barriers

While some scientists pursued their initial path of research and maintained narrow research agendas when something intriguing and new manifested from experiments, others would encourage side project as the primary method to avoid being blinded by conventional science and potentially missing precious opportunities where new revolutionary science may be hidden. Those who took side projects explored at the fringe while carrying their regular research program, in order to balance the two. They maintained their principal lines of research so as to keep a steady stream of publications, to hedge away the heightened risk of side projects and satisfy grant evaluations. At the same time, they increased the number of radical attempts (Fleming, 2002) by actively seeking out the more unconventional results at the periphery so as to increase the likelihood of discovering a breakthrough. A respondent describes having taken on various high-risk projects with potential of yielding high impact results,

> *"I think there is quite a core of people who are prepared to do a few risky things on the side that may influence. I have done a lot of stuff like that." (respondent 7)*

This behavior is reminiscent of the extensively studied topic of exploration and exploitation that firms undertake in their innovative quest (March, 1991) and provides empirical evidence at the individual or laboratory level for ambidexterity despite the inherent discordancy shown in the literature of simultaneously exploring and exploiting (Burgelman, 1983; March, 1991)

Circumventing Boundary Barriers

Scientists hinted to the fact that those with broader exposure to and awareness of work produced within disparate but related scientific communities were less likely to be isolated and had higher chances of linking disparate sources of knowledge. This diversification of exposure to various sources of information acts as a way to break down boundary barriers and enhance the likelihood of breakthrough. For instance, cross-disciplinary and cross-organismic conference attendance affords exposure to research done outside the immediate area of focus or model organism as illustrated by this informal interaction that happened while one respondent was on the way to a conference with a colleague.

> *"The most important thing at conferences is what you hear in the halls and in the coffee breaks. For example, I have heard about microRNAs way in advance before there were publications in a train station on my way to a conference." (respondent 9)*

Similarly, changing and mixing conferences that one attends also enables diversification in awareness as illustrated by another respondent.

> *"Changing the conferences that you go to. You have a new discovery in a field it's not part of your field then you have to go to conferences to tell people what you've learned and also to learn what's there in the field." (respondent 1)*

Conferences not only provide opportunities for scientists to hear about odd negative results that may have been shelved, they also enable attendees to sample current research topics at the forefront of related fields as they expose scientists to a variety of opinions fostering divergent thinking (Nemeth, 1986) and spurring creativity. This diversity in contexts cultivate brokering of ideas and finding analogies between seemingly disconnected and unrelated fields (Hargadon & Sutton, 1997), thereby enabling intriguing new avenues to be pursued.

Additionally, teaching cross-disciplinary courses can also force scientists to go beyond the comfort zone of their immediate research area and become familiarized with tangential topics and organisms. Aside from expanding scope of knowledge through thorough literature search, teaching also leads scientists to reach out to colleagues they would otherwise not connect with outside their fields, as discussed by two respondents below.

> *"Because when you teach you need to read about things which you are not directly involved in [...] For example, I have one paper which has been cited more than six hundred times, and this paper actually came from the fact that I was teaching in a university." (respondent 9)*

> *"Our most highly cited paper was a consequence of [teaching]. Although that wasn't why we did the research but it was clear that we could make half the problem go away*

> *overnight. So that was really very much influenced by my having to teach that course." (respondent 16)*

Circumventing Paradigmatic Pressures

In many instances bizarre experimental observations were shelved because scientists were reticent to challenge established truth in existing theory that constrained them to think within the confines of current dogma. They failed to identify and propose solutions with breakthrough potential outside of socio-cognitively delimited borders. To avoid being locked in, many informants stressed the importance to be open-minded and not be bogged down by the confines of current dogma or existing models (Simonton, 1989), which forces one to think about the implications of a set of experiments more comprehensively and heightens creativity. As illustrated in the following quotes,

> *"We as scientists want to be doing something that's different. You want to be following things that aren't the same as what had been looked at before." (respondent 4)*

> *"For further advancements to be made and more breakthroughs to be discovered one cannot believe that theories that are proven are there forever. Moreover, one needs to be open-minded and be ready to admit being wrong sometimes." (respondent 14)*

> *"So I think that there may not be a lot of mystery to why people find these breakthroughs it's a matter of considering the possibility of something in that biological system that we have no clue about and if you're setting aside this allusion that we have all the pieces […] Breakthroughs also get leveraged by the culture shifting towards acceptance of the idea that there are things that we don't understand." (respondent 2)*

Another way to decrease the amount of outliers from being discarded and improve research effectiveness is by substantiating results against other organisms instead of quickly ruling them out as artifacts. Viewing these practices from a social network perspective, scientists are building informal ties or inter and intra-laboratory collaborative ties to validate results. Thus, social ties not only act as conduits of information (Granovetter, 1973) in the production of knowledge but also as a mechanism of substantiation. This creates social validation (Cialdini & Trost, 1998). When scientists want to propose a new theory that would nullify existing ones it is easier to be contrary if many people are on board.

Multiple levels of validation are used in the substantiation of work: first, ensuring that results are internally consistent using well-controlled experiments, then evoking evolutionary conservatism as a mechanism of validating results by seeking parallels. These mechanisms enable scientists to develop enough confidence in obtained results to break away from the socio-cognitive confines established by the field and realize that theories are not always valid forever. Evolutionary conservatism is a double-checking mechanism, where observing the same artifact in multiple organisms or settings hints that the anomalous result is not an artifact but rather something real and substantial. To achieve evolutionary conservatism several practices are available: attending conferences, setting a laboratory in close proximity to other labs and facilitating collaboration, and finally running a laboratory that studies multiple organisms instead of one single organism.

Conferences are used as one of the mechanisms to confirm the soundness of abnormal results through informal social ties, as illustrated by two informants,

> *"And it only requires you going along to one seminar. We've been clearly influenced. We had a theory, we didn't have any confidence in it, and this guy from Harvard shows up and talked about something utterly different, and you think that's worth doing a few experiments." (respondent 7)*
>
> *"Both of you will hear a talk you can discuss what you think are the reasons, what's really happening there, to what extent you think it's going to be reproducible, to what extent is this really going to change the way people think, are there other explanations. All these things you can do between sessions, and also talk to people about some surprising thing that you're finding and get input and be able to test ideas with." (respondent 3)*

Collaborating with other proximate labs is another way to confirm evolutionary conservatism. Instead of relying on other laboratories to report findings in the literature or share results at conferences to ensure evolutionary conservatism, researchers can perform the necessary experiments with collaborators. Geographically proximate colleagues ease the collaboration even further as proximity to other labs that focus on other model organisms bolsters spillover (Jaffe, Trajtenberg, & Henderson, 1993; Thompson & Fox-Kean, 2005). It increases the chances of running into someone in an unplanned manner and the development of non-professional relationships that may lead to conversations that advance the exchange of ideas or validate results that one may be doubtful of. For instance,

> *"If you're at a place like MIT where there are labs that have the expertise in each of these systems, usually in the same building or across the street, it's very easy for students and post-docs to start a project in these systems and get help from their friends in those labs." (respondent 3)*

> *"If I think of a question, for example, if we had made some discoveries in fish that we would like to know if it's conserved, I would approach it by collaboration rather than having to re-invent the wheel." (respondent 8)*

Finally established scientists who have large enough research groups can also opt to study multiple model organisms in their own lab and avoid coordination costs associated with collaboration. A respondent described his experience working as a research fellow in a multi-organism laboratory of another respondent,

> *"What I really liked is that even in a single lab we were working on ten different organisms. We were a few people working on plants but there were people working on mice, on c. elegans, on drosophila, on zebra fish, on chicken… and this diversity of material that we were studying was really providing an exciting discussion. We could not only go and find the details of silencing we were studying in each organism, we could also make the parallel and trying to find what was common between these different mechanisms, how did it start, how did it evolve." (respondent 14)*

Bibliometric Operationalization

Although operationalizing cognitive constructs is still undeniably difficult, several new bibliometric measures can be derived from these practices, and as future work tested and generalized empirically using large datasets. For instance, exploring at the fringe to circumvent framing barriers can be proxied from the frequency distribution of scientist's MeSH keywords, where individuals who have a tendency to try high-risk explorative projects on the fringe are characterized by having a set of high frequency MeSH keywords representing the core of their research while at the same time having many one-off MeSH keywords mimicking the explorative nature of side projects. Similarly, exploiting intra and inter laboratory collaborative ties as a substantiation mechanism to avoid being confined paradigmatically can also be captured by animal model specific MeSH keyword of papers written by a single author. If a single scientist is associated with multiple animal models from their published works' MeSH keywords, they have either worked with other labs that use different model organisms or run a lab that supports research in multiple organisms.

To broaden one's awareness in related research, scientists also resort to attending conferences and teaching besides turning to the literature. The role of conferences has been understudied in the literature but findings in this paper suggest that the number of conferences and the breadth of conferences, whether interdisciplinary or cross-organism an individual attends is important to take into account as a source of breakthrough. Conferences are important not only as a perturbation to boundaries between communities of science to gain diversity of opinions and knowledge, but also act as a mechanism for result validation and provide a glimpse of informal scientific networks not captured through purely co-author collaborations. The number of cross-disciplinary courses a scientist teaches can also serve as a proxy for a source of breakthrough. Although these two measures are not readily available in archival data, they can be obtained by running surveys, and for the former by combing through conference attendance lists.

Furthermore, another measure that proxies the scope of awareness of related research communities is the breadth of backward citations that scientists reference in their own publications. This measure is also one of few bibliometric measures that can capture cognitive processes, as scientists only cite papers that they are aware of. I implement this measure and test it on the predictive regression models in the previous chapter. I construct a distribution of MeSH keywords for all publications that scientists cite prior to the 1998

RNAi breakthrough discovery. And exactly like the publication depth measure introduced in the previous chapter, I calculate a citation depth measure using all MeSH keywords that characterizes the backward citations each scientist in the RNAi community referenced. Adding this new measure to the extensive list of explanatory variables yield significant results. As this current chapter specifically focuses on better understanding through counterfactuals the process of breakthrough, I expect this new measure of citation depth to be negatively significant for strict operationalizations of breakthrough especially for the case of elites situated in the top 10 percent of citations. Surprisingly for the sample that includes all papers whether first, middle or last authored, results are not significant for citation depth, however the results are significant for the sample of first and last authored papers. These results can be interpreted such that citation breadth is only significantly associated to breakthroughs when scientists are the first or last author on a paper – author positions in a publication where most of the writing and thus the referencing decisions are made. The effect size is calculated by increasing the citation depth measure by one standard deviation, and yields a decrease of 20.1% to the dependent variable. Thus, broader citations are correlated with increased probability of being in the elite of the top 10% of citations. Table III-2 shows the regression result similar to those in Table II-4 with the added explanatory variable for citation depth. The sample in Table III-2 is smaller because of missing values for the newly introduced variable.

Table III-2 (following page) – Predictive models of the top 10% of citations with logit, number of forward citations of 98 papers and number of 98 papers both with quasi-maximum likelihood Poisson where the dependent variables are derived from papers published as first or last author in 1998.

	Logit Top10c	QML impact	QML prod
DV	top10cite	ncite98	npub98
	b/se	b/se	b/se
lnpub97_fl	0.359**	0.202	1.052**
	(0.13)	(0.15)	(0.06)
lncite97_fl	0.990**	0.748**	-0.027
	(0.09)	(0.06)	(0.03)
constraint	-0.823*	-0.570*	-0.290*
	(0.32)	(0.29)	(0.11)
lncoauthor	0.169*	-0.036	0.014
	(0.08)	(0.07)	(0.03)
pubdepth	0.401	0.293	0.779
	(1.74)	(1.44)	(0.65)
lexp	-1.718**	-1.053**	-0.931**
	(0.24)	(0.19)	(0.08)
prestige	0.006	0.007	0.003
	(0.01)	0.00	0.00
collabcore	0.506	0.205	-0.015
	(0.32)	(0.18)	(0.12)
techcore	-0.202	0.058	0.093+
	(0.15)	(0.13)	(0.05)
academic	2.843	1.780*	1.238+
	(2.35)	(0.87)	(0.75)
prestiged	-0.033	-0.109	-0.144*
	(0.18)	(0.12)	(0.07)
citdepth	-4.859**	-0.424	-0.511
	(1.77)	(1.50)	(0.56)
affil1p	0.175	0.082	0.158
	(0.25)	(0.22)	(0.11)
constant	-5.495*	-0.278	-1.251
	(2.42)	(1.00)	(0.79)
N.Obs	2504	2504	2504
Log-Likelihood	-697.377	-42836.565	-3868.446

+ p<0.10, * p<0.05, ** p<0.01

vi. Discussion

The unifying theme that emanates from the above discussion is that at different stages of the discovery process scientists on the verge of discovery failed to identify or propose the breakthrough opportunity. This failure is based on a cognitive process triggered by institutional factors stemming from various interacting barriers at both the problem identification and solving stages of the creative process. Besides contributing to the literature by showing and proposing a framework by which the seminal discovery was missed several times, this work also provides a collection of practices that scientists use to remedy the barriers discussed above and increase the likelihood of breakthrough. These practices can be operationalized as testable sources of breakthrough, although it is important to note, however, that many of them are necessary but not sufficient.

Additionally, this work adds to the micro-foundations of innovation. The literature in innovation has thus far mostly assumed constant input to innovation. My results suggest, instead, that individual inputs are quite heterogeneous and should be accounted for. Indeed, scientists behave differently with regard to conference attendance, teaching, taste for exploitation versus exploration, collaborative preference and willingness to take closer or further leaps.

My findings also shed light on how the institutional differences and divergent nature of knowledge produced between science and technology are manifested in the discovery of scientific breakthroughs. The understanding of the RNA interference phenomenon is puzzling in that several documented observations were witnessed before discovery was made. Contrary to technological innovations where a breakthrough invention happens at first successful occurrence, a number of scientists were on the verge of breakthrough but missed it. In these cases, the novelty component is not in the observation but rather in the explanation of why a particular abnormal result occurs. In other words, the definition of success in science is different from that in technology. Hence what would have been a success in the technology realm is not considered as one in science because mere observations or descriptions of a phenomenon are insufficient, they also need to be explained. Scientists are preoccupied by understanding real mechanisms in nature and are, consequently, worried about the validity and correctness of their findings. Moreover, truth seeking cultivates a requirement of being right and a fear of being wrong, which translates into being constrained by the limits of current theory. Technologists, on the other hand, mainly care about whether their inventions function as intended without necessarily needing to comprehend why it works.

The other conundrum around RNAi centers on the fact that its initial use and perception as a tool did not facilitate discovery of its trigger mechanism but rather delayed it. RNAi is a perfect illustration of the tension between concepts and tools because it effectively embodies both. Historians of science have extensively explored the two, and described how scientific revolutions arise from each. Thomas Kuhn (1962) perceived science from the point of view of a theoretical physicist, thereby emphasizing the great leaps of theoretical and conceptual insight that give rise to scientific revolutions for understanding nature while taking for granted experimental data. Whereas thirty years later, Peter Galison's (1997) argument that new tools drive the process of scientific discovery stems from an experimental physics viewpoint where he described great leaps of practical ingenuity for observing nature enabled by the acquisition of new data.

RNAi, however, is a hybrid that does not fit squarely in one camp or another. Instead, it is a tool based on an underlying biological concept. The case of RNAi suggests that the nature of the underlying knowledge should be a continuum rather than simply having two distinct categories – concepts and tools. RNAi debuted as a tool that arose from observations in plants, fungi and worms, but not understanding the causal mechanism to the phenomenon impeded its stability as a technique and consequently its initial widespread use and diffusion. It was not until the trigger agent was identified that the community of scientists started to study the intricacies of its mechanism in all organisms including complex ones and RNAi became truly revolutionary. However, although in the beginning its perception as a tool delayed understanding of the concept, it promoted diffusion once the trigger mechanism was understood and the technique was stabilized. Familiarity brought about by its use eased acceptance of the underlying concepts.

One weakness of this work is its sole focus on cognitive barriers driven by institutional factors to breakthrough. Without doubt, the above discussions on constrained resources and taking on side projects allude to the role that incentives play in discovering breakthroughs, which have been extensively studied in the innovation literature. The institution of science is based on the priority-reward system where one is recognized for being first to discovery thus pushing scientists to take on high risk and high rewards projects. But scientists also face the realistic pressure of producing a steady stream of papers for funding purposes unless they benefit from sources that tolerate early failure, reward long-term success, and give its appointees great freedom to experiment (Azoulay, Graff Zivin, & Manso, 2011). Thus, intermixed with the barrier of framing the puzzling phenomenon as a tool is an incentive pressure of consistently producing publications.

The funding of research grants and the evaluation process within academia all play significant roles in determining the research path that scientists take. For instance, strict funding schemes with frequent short-term deadlines and deliverables will most likely force scientists to stay closely on track with the proposed grant project, whereas more flexible grant structures would afford the scientist to experiment more. The tight timeframes of an academic scientist’s tenure evaluation can be another incentive barrier which may lead young scholars to pursue more incremental and less uncertain projects to guarantee publishing. Compared to cognitive barriers these economic barriers, however, are more deliberate.

The main limitation of this work is the generalizability of findings from one single case study especially when studying idiosyncratic rare events like breakthroughs. Therefore, these above findings should be interpreted with caution. However, if breakthroughs are conceptualized as a process of multiple attempts mainly characterized by failure with eventual success, then that process can be studied and characterized. Moreover, the goal of this work is to generate theory and extend current understanding of sources that enhance breakthrough potential, which can then be operationalized and more broadly tested quantitatively so as to show generalizability. These include expanding research scope through exploration at the fringe to avoid being blinded by conventional science, exploiting social ties as a mechanism of substantiation to overcome being constrained by current dogma, and broadening exposure and awareness of work across multiple scientific communities to mitigate the inability of connecting the dots.

Furthermore, RNAi being a discovery that occurred fifteen years ago, one must also be mindful of how findings in this work should be applied given the changing processes in which science is done and published today. New processes for innovation such as open innovation, as well as alternate venues of publication outside of the conventional peer-reviewed articles such as web blogs, chat groups, preprints have emerged. In the case of the former, given that problems to be solved are open to anyone who can plausibly provide a solution, framing the problem in the correct way is crucial. With changes in the latter where steps to publishing have seemingly been simplified, anomalies should surface earlier, more easily and be readily accessible to everyone. However, although these structural changes may have simplified the publication process, it is not obvious whether scientists, especially untenured scientists, will freely post their expected or unexpected results on websites without changes to the institutional priority-based reward system of science.

vii. Conclusion

Moving beyond bibliometric measures by using qualitative interviews, I proposed a cognitive framework driven by institutional factors on the emergence of breakthroughs through failures and found that the seminal discovery was missed several times because of failures to identify and propose the breakthrough opportunity. At the basis of this failure are three barriers. In the problem identification stage, path dependence from established technologies and the quest toward normal science blinded scientists from recognizing a prospective breakthrough. Instead, they framed RNAi as a tool while ignoring it as a scientific concept worthy of study. Existing boundary barriers between communities of scientists aggravated this difficulty in identifying the breakthrough opportunity by misrepresenting the magnitude of the problem as it prevented recognition of links between several prior instances of odd observations. In the problem-solving stage, scientists suffered from the socio-cognitive barrier of being constrained by current dogma. Due to fear of being wrong, they hesitated to propose solutions that significantly strayed away from the confines of established theory. Coupled with boundary barriers, similar anti-dogmatic results stayed isolated and diminished confidence to propose a new revolutionary paradigm.

This work has implications in the design of organizations and institutions that partake in scientific discovery. Understanding the barriers to scientific knowledge creation is vital not only for academic administrators but also from both managerial and policy standpoints. It illustrates the fundamental differences inherent in the production of scientific and technological knowledge, and directly speaks to the organizational design of science-based firms (where the literature has mainly focused on technological innovation and remains thin) by providing structural characteristics and policies that foster the production of groundbreaking discoveries. These include facilitating interdisciplinary research teams, encouraging cross-organism and cross-field conference attendance, and providing incentives that enable the flexibility to take on side projects on the fringe. From a policy vantage point, this work characterizes which scientists have the highest potential of breakthrough. This is a first step in eventually moving up levels of analysis to locating communities of scientists more likely to discover breakthroughs and, thus, enabling more targeted governmental subsidies and private investments into them (Lane, 2009).

A natural extension to the current work is to further quantitative operationalizations and test sources that enhance breakthroughs uncovered herein. Keeping in mind tradeoffs in the practices scientists employ to circumvent barriers to breakthrough opportunity identification and dynamics between each theme, quantitatively testing these new sources of breakthroughs can shed light on equilibrium points as well as interaction effects. Another extension follows from the conventional wisdom of 'having smart people at the right place at the right time' when eliciting about breakthroughs. Therefore, in future work the question of when is the right time for breakthroughs to emerge can be studied. For instance, at what point in the maturity of a field are breakthroughs most likely to be made, what role do complementary discoveries – for instance microRNA and genome projects in the case of RNAi – play in spurring or stunting revolutionary discoveries, how much of an installed base is required within a community for breakthroughs to emerge are all intriguing questions to further explore.

IV. Fostering Translational Research

i. Introduction

A frequent question in the innovation and entrepreneurial finance literature is the impact of different funding schemes on firm performance and innovative output. The interest in understanding how ideas are produced and the means by which idea production is enhanced have been driven by the belief that knowledge from scientific research and its subsequent translation into technological inventions is a driver for wealth creation and stimulates economic growth. In scientific research, studies have investigated the effect of various research grant designs (Azoulay, Graff Zivin, & Manso, 2011). Similarly in the technology sector, the effect of angel investments (Kerr, Lerner, & Schoar, 2011), venture capital (Kortum & Lerner, 2000; Samila & Sorenson, 2011), banks (Black & Strahan, 2002) and initial public offerings (Bernstein, 2012) on innovation and entrepreneurship is another area of great interest. These studies mainly investigate the impact of various funding schemes on innovation within the well-defined boundaries of the scientific and technological institutions. Therefore, implicit in these works is a dichotomy between the knowledge created in science and that produced for technological and commercial purposes, as well as the assumption that knowledge created in the scientific and technological realms are produced independently.

Alongside these prevalent funding structures that separately focus on science and technology, many countries have invested in academic-industry partnership grant schemes that target translational research at the intersection of science and technology. In the United States, National Science Foundation (NSF) shared resources centers often require some form of partnership with private firms to accelerate product development, while the

National Institutes of Health (NIH) academic-industry partnership program seeks to identify the most compelling opportunities for cross-boundary research that would link biomedical research to commercial opportunities. In Germany, the Fraunhofer-Gessellschaft is a partially state-supported application-oriented research organization that undertakes applied research of direct utility to private and public enterprises. The Technology Strategy Board in the United Kingdom supports a range of research collaborations and runs programs such as its Knowledge Transfer Partnerships, which support UK businesses wanting to improve their competitiveness and performance by accessing the knowledge and expertise available within UK universities and colleges. Though there are many such programs globally, little research has been performed to assess the impact of these approaches on the quantity, impact and collaborative nature of knowledge produced especially from the perspective of the firms that receive them.

We examine academic-industry partnerships sponsored by the Danish National Advanced Technology Foundation (Højteknologifonden), an agency of the Danish government. In its unique mediated funding model, DNATF awards grants for projects that encompass cooperation between at least one academic institution and one firm. DNATF kindly provided us with a novel dataset for this study that enabled us to determine the efficacy of their academic-industry funding model in terms of the quantity, impact and collaborative nature of innovative outputs of the firms.

This work bridges the literature between innovation funding and the coevolution of science and technology by lending empirical evidence on the impact of academic-industry partnership grants on knowledge creation. This study differs from other works in the innovation funding literature in that instead of focusing on traditional sources of funding such as venture capital, debt, initial public offerings, or basic research grants it investigates a setting that blurs institutional boundaries of science and technology. It also takes a distinctive perspective from the literature stream that investigates the effect of academic scientists crossing scientific and technological boundaries on knowledge production, by centering on the firm as the level of analysis and investigating the impact of cross academic-industry projects on firm behavior and performance. Specifically, we assess how this novel combination of funding and mediation with public research institutions is effective in helping firms survive, and partake in riskier and wider explorative activities that spur innovation. We contrast a sample of funded firms with those that applied for DNATF funding but did not ultimately receive a grant. Since all proposal applications to DNATF are ranked, we develop several sample specifications to ensure that we do not suffer from selection bias by including qualitatively similar funded and unfunded firms.

Our results show that with subsamples of qualitatively similar small and medium enterprises and younger firms, the receipt of funds helps alleviate capital constraints by decreasing the likelihood of going bankrupt for funded firms. Moreover, it also has consistent positive effects on both filed and granted patents for funded firms. With regard to peer-reviewed publications, we only observe that forward citations to papers published by funded firms are significantly higher than those of unfunded firms, but surprisingly find no significant result for the quantity of publications nor for the cross-institutional collaborative nature of publications for funded firms.

The structure of this work is as follows. We begin by presenting the theoretical framework from the literature and develop testable hypotheses. We then elaborate on the setting from which we compiled our data, detail the estimation methodology employed to run our analyses, and interpret our results. Finally we discuss the contributions this work brings to the extant literatures and consider the implications for policymakers and managers.

ii. Theoretical Framework and Hypotheses

Literature on the financing of innovation has extensively explored the effect of funding on organizational performance and innovative output in the form of grants for academic research (Azoulay et al., 2011), of early-stage funding such as angel investments (Kerr et al., 2011) and venture capital (Kortum & Lerner, 2000), and of more mature financing outlets such as initial public offerings (Bernstein, 2012). Scholars in entrepreneurial finance have theoretically and empirically studied consequences of early-stage funding. Theoretical works suggested that the role of entrepreneurial financiers is not only to provide funding that relieves capital constraints but also alleviates agency problems between entrepreneurs and investors through monitoring and improved governance (Admati & Pfleiderer, 1994; Hellmann, 1998). Empirical researchers have also causally tested these theoretical propositions. For instance, Kerr, Lerner and Schoar (2011) showed that angel funding benefits ventures in improving subsequent survival, exit, employment, patenting and financing using regression discontinuity estimation. Exploiting exogenous shocks, venture capital funding has been shown to causally lead to higher patenting rates (Kortum & Lerner, 2000) and positive impacts on employment and aggregate income (Samila & Sorenson, 2011).

Although firms in our setting are not necessarily in early entrepreneurial stages, they still suffer from the same capital constraints that prevent them from undertaking risky innovative projects. Therefore, we posit that firms successful in obtaining academic-industry partnerships funding are less likely to file for bankruptcy, and more likely to take on R&D projects with resulting inventions encoded in patents.

Hypothesis 1: *Firms that receive funded, mediated academic-industry partnerships are less likely to go bankrupt compared to non-funded firms*

Hypothesis 2: Firms that receive funded, mediated academic-industry partnerships produce more patents relative to non-funded firms

Before continuing further, it is important to first elaborate the institutional differences between science and technology assumed throughout this study, where science is seen as a distinctive incentive system compared to technology. The scientific institution is primarily embodied in research universities based on a priority-based reward system where outputs are mainly in the form of peer-reviewed publications. The technology institution, in contrast, encodes ideas in protected modes, using for example patents, trademarks or copyrights, to facilitate commercialization and appropriation of economic rewards (Dasgupta & David, 1994). Moreover, the two institutions differ in the nature of the goals accepted as legitimate and the norms of behavior, especially with regard to the disclosure of knowledge. Science is concerned with additions to the stock of public knowledge, whereas technology is concerned with additions to the stream of rents that may be derived from possession of private knowledge. Given this distinction between science and technology, the prior two hypotheses were derived for the technology institution that firms are part of. However, given the cross institutional nature of academic-industry partnerships, we posit that obtaining such funding grants also alters firms' behavior in partaking in basic research activities more deeply rooted within science.

Firms have little incentive to undertake basic research because of the difficulty in protecting and patenting resulting knowledge since natural laws and facts are not patentable. Very few firms are broad and diverse enough to directly benefit from all the new technological possibilities opened up by successful basic research. Moreover, they are also confronted with the free rider problem that enhances use by others (Nelson, 1959). Thus, the high uncertainties and risks associated with basic research combined with the difficult appropriability problem diminish incentives for firms to pursue basic research and may prompt those with limited funding to completely avoid it.

With the help of governmental funding for academic-industry partnerships, we postulate that it provides firms with the motivation and the risk mitigation mechanism to assume more basic research, as encoded in peer-reviewed publications, that they otherwise would not have undertaken. Thus, as suggested by Rosenberg (1990), firms with basic research capabilities can make more effective decisions about applied activities, build the capability to monitor and evaluate research being conducted elsewhere (such as in universities), and evaluate the outcome of applied research to recognize possible implications. Moreover, since academic-industry partnership projects condense interactions between scientists and technologists and blur their institutional boundaries, spillover effects stimulate firms to take on more basic research.

Hypothesis 3: *Firms that receive funded, mediated academic-industry partnerships produce more peer-reviewed publications compared to non-funded firms*

One stream of literature examining the interplay of science and technology has investigated the effect of patenting on scientists' efforts to engage in subsequent scientific research. These works take the perspective of academic researchers originating in the scientific institution who also take on patenting activities. Thus the question of interest axes on understanding the effect of intellectual property rights on scientific research. Findings show that both the flow and the stock of scientists' university patenting are positively related to subsequent publication rates (Azoulay, Ding, & Stuart, 2009), and even though patent volume does not predict publication volume, it positively affects paper citations, providing insight on the research impact of patents (Agrawal & Henderson, 2002). These results imply that patenting is a complementary activity to fundamental research rather than a substitute. Even though our setting differs from these studies in that we are studying firms that originated in the technological institution but partake in basic research activities in science, we can still postulate from these findings that firms granted academic-industry funding are more effective at applying basic science results and therefore their peer-reviewed publications will receive more forward citations.

Hypothesis 4: *Firms that receive in funded, mediated academic-industry partnerships produce more frequently cited peer-reviewed publications relative to non-funded firms*

Only a few articles have empirically assessed the extent of overlap between science and technology (Murray, 2002, 2004). Cockburn and Henderson (1998) provided empirical evidence that spillover effects between science and technology were not a simple waterfall model in which the public sector produced knowledge that spilled over costlessly to downstream researchers. They showed that in order to take advantage of public sector research, firms must do more than simply hire the best scientists and invest in in-house basic research with appropriate pro-publication incentive systems. Industry researchers must also actively collaborate with their academic colleagues, which improves access to public sector research and quality of research conducted within the firm. Thus we postulate that given the close interactions between scientists and technologists when working on academic-industry partnership projects, collaboration and co-authoring across institutions increases.

Hypothesis 5: *Firms that receive in funded, mediated academic-industry partnerships produce more cross-institutional collaborative outputs relative to non-funded firms*

iii. Methodology

Setting

Our setting is the Danish National Advanced Technology Foundation (DNATF) founded in 2005 by the Danish government, whose broad objective is to enhance growth and strengthen employment by supporting strategic and advanced technological priorities. It was created with the aim of making Denmark one of the world's leading advanced-technological societies. DNATF provides governmental funding for academic-industry partnership collaborations, facilitating bridge building between Danish public research institutions and Danish companies in order to generate growth and technologies that benefit Danish society as a whole.

DNATF is the only Danish governmental funding source that exclusively supports academic-industry research collaborations. Funding for such collaborations, however, can also be obtained from other Danish governmental sources.[1] DNATF uses a bottom-up approach in the application process. It seeks to fund the best ideas within the broad realm of advanced technology. The investment portfolio covers sectors ranging from robotics, agriculture, livestock, biotechnology and medicine, all the way to telecommunications. Based on all funded projects since DNATF's inception in 2005 to 2011, the largest sector in DNATF's portfolio is biomedical sciences, making up 30% of all investments, while 26% are in energy and environment, 20% in IT and communication, 14% in production, 5% in agricultural produce and food, and 5% in the construction sector. Applications must include at least one academic scientist and one firm. In choosing the best ideas, DNATF screens on three criteria: obvious business potential, internationally recognized high

[1] The largest alternative state funding sources in Denmark are the Energy Technology Development and Demonstration Programme (EUDP), Green Development and Demonstration Programme (GDDP), The Danish Counsil for Strategic Research, the Business Innovation Fund, The Danish Counsil for Technology and Innovation, and finally, The Danish Public Welfare Technology Fund.

quality research and innovation, and entreneurship. Applications are screened in two stages by the board of DNATF, which consists of nine leaders from Danish industry and science who have extensive and unique knowledge in their respective fields.

The first application stage is the submission of a short expression of interest which identifies the core idea of the proposed project. Each expression of interest is read and scored A, B, or C by each board member before a board meeting. Individual board members form their own opinion *a priori*. At the meeting, the aggregate scores by all board members are tallied at the starting point of the discussion on deciding whether to approve the particular expression of interest for the second round. About 30% of the first round applications are approved and move into the second round, in which applicants prepare a more comprehensive proposal that explains the project idea in detail. These applications are then subjected to a peer review process by two independent reviewers, and armed with these peer reviews DNATF's board members again score each application with an A, B or C. Based on the aggregate scores and discussion, the board reaches a consensus on whether to fund each application. From the applications that proceed to the second stage, about 40% ultimately receive funding. During the final board meeting every year, a fixed pool of funding is awarded until fully exhausted, thus eliminating the potential endogeneity issue of reverse causality where innovation drives funding.

DNATF's mediated facilitation model entails active follow-up on each investment throughout the project period. A Single Point of Contact (SPOC), an individual who is part of the small DNATF staff, is assigned to each investment to act as a gatekeeper and link between the project and DNATF for the project duration. The SPOC practices active follow-up by participating as an observer in steering-group meetings, engaging in day-to-day dialogue with project participants, reporting quarterly to the board, and challenging the project participants on progress and issues throughout the project period. The SPOC focuses on facilitating effective collaboration between projects participants, maximizing the collaborative gains in the project.

By the end of 2012, DNATF had made 238 investments with a total project budget of DKK 5,320[2] million of which DNATF invested half in accordance with its 1-2-3 investment model where public research institution(s) fund 1/6 of the total budget, private firm(s) 2/6 while DNATF funds 3/6. Neither participating firms nor academic institutions are required to pay back the awarded funding, therefore using the self-financing scheme ensures that all parties have something at stake. Full requested amounts are committed at the time of award, but progress payments are contingent on performance. A project has a typical duration of 4 years and on average receives DKK 12 million from DNATF. Figure IV-1 shows the distribution of funded amounts DNATF has awarded by project.

DNATF project awards typically go to a team of one or two public research institutions teamed with an average of two companies. In 2012, 84% of all investments had one or more universities as the participating public research institution. The remaining 16% were either hospitals or universities and hospitals in cooperation. Foreign companies are allowed to participate but cannot receive funding. Of the unique companies in DNATF's portfolio (duplicates not included), 59% have 49 or fewer employees, 17% have 50-249 employees, 12% have 250-999 employees, and 12% have more than 1000 employees.[3] The age distribution for DNATF funded firms is skewed towards younger firms with 38% of firms aged 5-years and younger, 22% between 6 and 10 years old, 8% from 11 to 15 years old, and 36% being 15 years and older.

[2] DKK5,320 million is the equivalent of USD925 million at the October 2012 exchange rate of 5.75DKK/USD

[3] Additional numbers are provided by DNATF's yearbook.

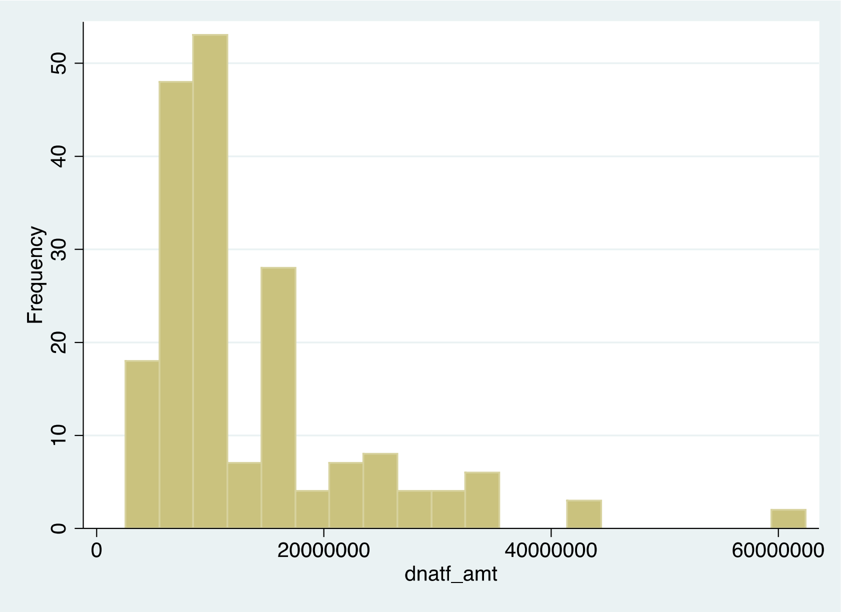


Figure IV-1 – Frequency distribution of amount funded by DNATF (in DKK)

Empirical Approach and Identification Strategy

Full Sample from Second Stage of Selection Process

The two-stage application process that projects undergo enables us to eliminate projects that failed to advance to the second stage of selection and concentrate only on those that did. These projects are more similar in quality and partially resolve our problem of unobserved heterogeneity stemming from selection bias where the funded projects are more promising and have higher potential of success. Thus, our first specification is the entire sample of firms that proceed to the second round of the evaluation process.

At the end of 2011, a total of 49 investments had been finalized. These finalized investments were all funded between 2005 and 2008. Out of the projects that DNATF chose to invest in, 47 were finalized as usual and two were stopped before nominal project completion by DNATF. Since there was no upper limit on the number of firms per project, the 49 invested projects corresponded to 102 participating companies. Among these 102 companies, 16 were duplicates, i.e. companies which participated in more than one of the 49 investments. Thus, in total there were 86 unique companies which have been part of finalized DNATF investments and these make up our funded group. For the matched control group, which consists of 105 companies we used firms that applied for DNATF funding from 2005 to 2008 and selected into the second round of review, but did not ultimately receive funding. All firms in the control group were part of applications that would have been finalized by the end of 2011 or before. Among the 105 companies 8 were duplicates, which amounted to a total of 97 unique companies in the control group.

Qualitatively Similar Small and Medium Enterprises Sample

A more detailed look at the sample of firms that received funding shows that it encompasses an extremely heterogeneous set along the dimension of firm size. While most of the firms that received funding are small and medium size enterprises (SME) defined as companies with 250 employees or less, some funding recipients boasted headcounts into the thousands of employees. Given the limited range (DKK 2,550,000 to DKK 62,400,000) in the amount of funding provided by DNATF, its impact would be more substantially felt in small and medium enterprises where the amount of funding comprises a sizable portion of the firm's R&D budget. Although larger companies still benefit from the influx of capital brought by funding, its impact would be likely less evident, as the funded amount only represents a small fraction of the firm's R&D budget.

Despite dropping firms whose projects did not advance to the second round of the application process as well as those with more than 250 employees, the reader may still argue that the difference between the best firms in the funded sample and the worst firms among the unfunded ones is still significant and that the sample specification still suffers from selection bias and unobserved heterogeneity. To address this issue, our second sample comprises of qualitatively similar *ex ante* projects except in their probability of funding. We exploit scores given by DNATF board members in their assessment for each application proposal as a quasi-ranking system, and drop from the sample the best funded firms and the worst unfunded firms. Interviews with DNATF staff revealed that an assessment of A for a project indicates that a board member believes that the project is highly worthy of support, B indicates that the project is worthy of support, whereas a C indicates not worthy of support. We translate this evaluation into a normalized score as dictated by Equation IV-1 for firm i, where A, B and C are binary variables equal to 1 based on the assessment of board member k. Moreover an A assessment is assigned a score of 10, B a score of 0 and C a score of -10.

Equation IV-1 $$score_i = \frac{10 \cdot (\sum_k A - \sum_k C)}{\sum_k (A+B+C)}$$

For each tranche of the normalized score, we identify the fraction of firms that are funded. In column 2 of Table IV-1, we observe that the fraction of funded firms increases monotonically as the normalized score increases. We see that at the lower end no applications with a normalized score of less than -2.5 were funded, and are therefore dropped from the sample. We also drop the top 5% of firms with normalized scores of above 8.5. Consequently, we define our narrow band of qualitatively firms to be those with normalized score in the range [-2.5, 8.5], effectively creating a matched sample of funded and unfunded firms.

Several characteristics of the data led us to believe that observable heterogeneity from sample selection can be eliminated. First, since DNATF does not have explicit funding rules that lead to systematic funding decisions as the selection process hinges on board member assessment and votes, the cutoff score for funding is not known in advance to applicants and therefore cannot be gamed or manipulated.

Normalized score	Funded (%)	Number of applications	Applications (%)	Cumulative applications (%)
[-7.5,-5)	0.0%	9	7.3%	7.3%
[-5, -2.5)	0.0%	17	13.8%	21.1%
[-2.5, 0)	15.4%	13	10.6%	31.7%
[0,2.5)	37.1%	35	28.5%	60.2%
[2.5, 5)	42.1%	19	15.4%	75.6%
[5, 7.5]	86.7%	15	12.2%	87.8%
[7.5, 10]	100.0%	15	12.2%	100.0%

Table IV-1 – DNATF funding selection by normalized score

Second, if we were to use unfunded firms as a matched sample to the funded ones, there should be no significant difference in the observables for unfunded and funded firms within of narrow range of normalized scores. We test this criterion using two-sided t-tests. Table IV-2 shows that firms situated within this narrow bandwidth were not significantly different on all observable dimensions at the time of application. These results are critical in order to draw causal inferences on the effect of the funding on firm performance and innovative performance. Moreover, a predictive logit regression model of the probability of funding – regressing a dummy *funded* variable on all observable explanatory variables listed in Table IV-2 – yields no significant result on any variable.

Consequently, our second sample specification consists of the region in which firms are most comparable – those with normalized scores in the range of [-2.5, 8.5] – dropping from the sample firms at the lowest and highest ends of the normalized score distribution, which amounts to 39 funded and 43 unfunded firms.

Characteristic	Unfunded	Funded	Two tailed t-test
age of firm	8.16	7.79	0.84
proposed duration	3.00	3.00	1.00
funding amount	12700000	12400000	0.85
number of parties	5.12	5.56	0.54
patents filed	3.29	1.16	0.29
patents granted	2.66	0.84	0.35
publications	4.02	6.44	0.46
forward citations	119.19	136.79	0.85
cross-institutions	2.28	3.28	0.62
n	43	39	

Table IV-2 – Comparison of funded and unfunded firm observables for SMEs

Qualitatively Similar Younger Firm Sample

Characteristic	Unfunded	Funded	Two tailed t-test
age of firm	5.16	5.26	0.89
proposed duration	3.09	3.00	0.63
funding amount	14800000	13200000	0.46
number of parties	5.13	5.37	0.73
patents filed	3.70	1.85	0.40
patents granted	2.98	1.34	0.43
publications	7.84	6.71	0.83
forward citations	147.13	141.80	0.96
cross-institutions	4.09	3.97	0.97
n	45	38	

Table IV-3 – Comparison of funded and unfunded firm observables for young firms

Instead of small and medium enterprises, we take another cut at the data using age. From the skewed age distribution of firms, we first define a subsample of firms that are 15 years and younger, which yields 55 funded and 74 unfunded firms. Then, following the same method described above, we determine qualitatively similar younger firms. We find the same normalized score range of [-2.5, 8.5], which amounts to 38 funded and 45 unfunded firms. Similar to Table IV-2, Table IV-3 shows that observable measures of younger firms are not significantly different for the funded and unfunded samples at the time of project application.

Outcome Variables

Hypothesis 1 explores whether receiving DNATF mediated funding decreases the likelihood of bankruptcy for firms. We obtain data on whether a firm in our sample is bankrupt up to four years subsequent to funding application. The outcome variable is an indicator (*bankrupt*) that takes on the value of 1 if the firm is bankrupt four years after funding application and 0 if it is still operating. We stopped tracking this outcome variable and all other outcome variables four years after application to maintain consistency throughout our dataset, since our sample includes firms that applied for funding in 2008 where only four years have elapsed at time of data gathering in October 2012.

Hypothesis 2 investigates the relationship between academic-industry funding and the quantity of knowledge produced measured by the number of patents. We use the number of granted patents (*patents granted*) assigned to the firm as filed for each year up to four years after the year of application as well as the number of unissued patents filed (*patents filed*) for each year up to four years after the year of application. All outcome variables for hypothesis 2 onward are in long panel form by firm-year.

Similarly in hypothesis 3 for peer-reviewed academic papers, we count the number of peer-review papers (*publications*) researchers of the firm have published for each year up to four years after the year of application.

Hypothesis 4 focuses on the impact of knowledge produced from mediated academic-industry partnership projects. We operationalize the impact of publications using the commonly employed measure of forward citations. Consequently, we count the number of citations (*forward citations*) garnered in all peer-reviewed publications for each year up to four years after the year of application.

Finally, hypothesis 5 explores the co-evolutionary nature of science and technology in mediated academic-industry partnership projects. To see the effect of the academic-industry partnership funding, we count the number of instances (*cross-institutions*) where peer-review publications by a firm are published in collaboration with at least one co-author affiliated with an academic institution for each year up to four years after the year of application. We planned to develop a similar measure for patents, but affiliation data for inventors do not include the institution for which they work and therefore we could not make any rigorous inferences as to their professional affiliation.

Datasets

As described earlier, our dependent variables fell within three categories – bankruptcy data, patent variables and publication variables – each of which required a different data source.

Bankruptcy data was collected using BiQ Erhvervsinformation (BiQ), a database that includes all registered Danish firms and provides yearly information on each firm from 20-30 years ago onwards. The company data we used from BiQ is updated daily from the Danish Business Authority, a governmental database that keeps comprehensive information on all Danish firms.[4]

Data for patent variables was collected at the firm level using Google patents. Firm name was used to match for patent assignees, with some minor adjustments due to Danish letters not found in the English alphabet. The dataset for both filed and granted patents is

[4] Other Danish alternatives to BiQ include Statistics Denmark (Danmarks Statistik), which is the main data source for census type data. Statistics Denmark also keeps information on firms, however this information is only available on an aggregated level in contrast to the more nuanced yearly firm level data from BiQ.

in long panel form from time t_{-4} to time t_4, for fours years before and after the application year amounting to a total of nine years of data (four years prior to funding, four years after funding and t_0).

Publication variables were collected from the Web of Science. Again, we used firm name to search for publications with relevant organizational affiliation, where we extracted the quantity and impact of publications per firm using respectively the number of publications and the number of citations garnered by these publications as proxies. One additional variable on cross-institutional co-authorship, papers published in cooperation between firm(s) and universities, was also constructed. Similar to patents all publication variables were collected annually for four years before and after the year of funding application as well as the year of funding itself.

Finally, a number of basic variables were obtained from DNATF's database and integrated into the dataset. These consisted mainly of information on the specific project or application each firm has been part of, such as the year of application used to derive the *post* indicator as well as whether a project was *funded* or not. Variables such as industry sector, project duration and amount of funding were all included as comparable *ex ante* observables in the analyses.

Regression Model Estimation

In hypothesis 1 because the *bankrupt* outcome variable is not in panel form but rather an indicator of whether a firm is bankrupt 4 years after funding application, we used probit models with cluster robust standard error on all three sample specifications described above.

To test for hypotheses 2 to 5 on each sample specification described above, we employed for our estimation a diffence-in differences (DiD) model, specified as follows:

Equation IV-2 $Y_{i,s,t} = \alpha + \gamma funded_s + \lambda post_t + \beta_1(funded_s \cdot post_t \cdot t_1) + \beta_2(funded_s \cdot post_t \cdot t_2) + \beta_3(funded_s \cdot post_t \cdot t_3) + \beta_4(funded_s \cdot post_t \cdot t_4) + \beta X_{i,t_0} + \varepsilon_{i,s,t}$

The outcome variable is $Y_{i,s,t}$ for firm *i* at time *t* for funded state *s*. Since we are assessing the effect of academic-industry partnership funding, the first difference is that between funded and unfunded firms, and the second difference is that between the pre and post funding periods. Thus *funded* is an indicator of whether a firm *i* has received funding at time t_0, while *post* is an indicator of being after the funding event. The difference-in-differences is captured by the interaction effects of $funded_s$ and $post_t$, and since we are interested in effect trends, we also interact the DiD with a time indicator of t_1 to t_4 for each year after funding. Thus coefficients β_1 to β_4 are our coefficients of interest. For each firm *i* in the vector X_{i,t_0} of length *j*, we also control for observables by including application year fixed effects and industry fixed effects.

Since all variables for patents and papers (number of patents and papers, number of citations and number of cross-institutional papers) are non-negative and over-dispersed counts, we used quasi-maximum likelihood Poisson models with cluster-robust standard errors to circumvent the assumption of equal mean and variance distribution for Poisson models and minimize estimation bias.

iv. Results

This section shows results for the hypotheses we proposed earlier in an effort to empirically bring evidence to the research questions of how does academic-industry partnership funding affect firm innovative performance. Table IV-4 shows the summary statistics including the mean, standard deviation, minimum and maximum for each dependent variable as well as the *funded* and *post* indicator variables.

Variable	Observation number	Mean	Std. Dev.	Min	Max
normalized score	1629	2.0196	4.564143	-7.5	10
proposed duration	1845	3.063415	0.7333031	1	5
amount funded by DNATF	1728	1.35E+07	9756608	2550000	6.24E+07
number of parties	1845	5.663415	3.674086	2	19
funded	1845	0.4926829	0.500082	0	1
post	1845	0.4444444	0.4970387	0	1
SME	1737	0.6839378	0.4650714	0	1
young firm	1845	0.6292683	0.4831317	0	1
bankrupt	202	0.0445545	0.206836	0	1
patents filed	1827	3.217843	11.19599	0	178
patents granted	1827	1.383142	4.961285	0	49
publications	1782	2.070707	18.09965	0	337
forward citations	1773	15.40835	80.91633	0	1553
cross-institutions	1773	0.6739989	2.536591	0	41

Table IV-4 – Summary statistics

Placebo Test on Period Prior to Funding Event

Before showing our main results for the period after the funding event, we ran placebo tests to ensure that trends in the outcome variables prior to the funding event were not significantly different between the would-be funded and unfunded firms. Figure IV-2 graphically depicts one such trend for granted patents in the qualitatively similar younger firm sample with pre and post funding periods. Thus for all outcome variables except *bankrupt* (since all firms at time of application t_0 were all operating), we ran DiD regressions using the same estimation model as in **Error! Reference source not found.** with data for time t_{-4} to time t_0 as if the funding event occurred at time t_{-4} and include four subsequent years of data after funding from time t_{-3} to t_0. For all outcome variables of innovative quantity, impact and collaborative nature, we found no significance in the DiD coefficients, which implies that no significant difference in our outcome variables of interest existed between funded and unfunded firms prior to the actual funding event at t_0. The placebo test regression tables are not included herein, but can be obtained from the author upon request.

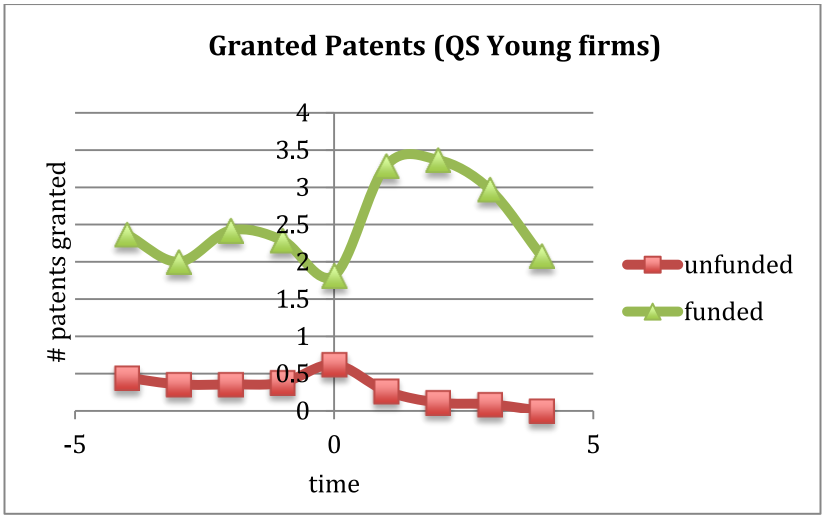


Figure IV-2 – Graphical depiction of the number of granted patents for both funded and unfunded firms for periods before and after funding at t_0 using the qualitatively similar sample of younger firms.

Effects on Firm Survival

Table IV-5 reports results for the firm's likelihood of going bankrupt for all three sample specifications, controlling for industry and application year fixed effects. We find for all three sample specifications that funded firms are significantly less likely to go bankrupt four years after applying for funding. The first model in Table IV-5 shows that firms successful in obtaining funding are 1.34 times more likely to survive up to 4 years after receiving funding compared to non-funded firms that went through the same application process while employing the full sample of firms in the second round of selection. Similarly restricting the sample to qualitatively similar SMEs, the second model shows that funded firms are 2.48 times more likely to survive. And finally, for the sample of qualitatively similar younger firms, we find analogous effects where funded firms are 2.31 times more likely to survive than non-funded ones. Specifically, hypothesis 1 is confirmed. Therefore our results provide empirical evidence on the hypothesis that mediated funding for academic-industry partnerships alleviates capital constraints for firms and increases the likelihood for a firm to remain in business.

Probit Models	Full b/se Model 1	QS SME b/se Model 2	QS Young b/se Model 3
funded	-1.343**	-2.480**	-2.306**
	(0.39)	(0.79)	(0.81)
constant	-4.682**	-4.094**	-4.087**
	(0.54)	(1.01)	(1.03)
N.Obs	179	63	67
Log-Likelihood	-29.849	-12.269	-12.808

+ p<0.10, * p<0.05, ** p<0.01

Table IV-5 – Bankruptcy data. Probit regression models with bankruptcy as indicator outcome variable four years after receiving funding, run on all three sample specifications: full sample in second round selection, qualitatively similar SMEs, and qualitatively similar young firms.

Effects on Quantity of Firm Innovations

We now shift to explore the effect of receiving funding on firm innovative performance, including productivity in patents and publications. We measure the effect of receiving funding on a firm's innovative productivity by counting the number of filed and issued patents after application, as well as the number of peer-reviewed publications. Table IV-6 shows our findings for the number of filed and granted patents, while Table IV-7 displays results for publication count.

In Table IV-6, we find that especially for the two qualitatively similar sample specifications the number of filed patents after funding application is significantly higher for funded firms than for non-funded ones. Specifically, in models 2 and 5 for the narrowly defined qualitatively similar SMEs, we find that funded firms file between 3.6 times ($e^{1.282}$) and 4.7 times ($e^{1.557}$) more patents than unfunded firms in the four years after funding application, and funded firms receive between 4.0 times ($e^{1.376}$) and 6.9 times ($e^{1.928}$) more granted patents when filed up to three years after funding application. Comparable strong significant results are also observed for qualitatively similar younger firms. Models 3 and 6 respectively show that funded firms file between 2.1 times ($e^{0.723}$) and 2.3 times ($e^{0.812}$) more patents up to four years after funding application, and receive between 3.4 times ($e^{1.218}$) and 5.5 times ($e^{1.701}$) more granted patents when filed up to four years after funding application. Thus, we find strong empirical evidence that confirms hypothesis 2.

	Patents filed			Patents granted		
Poisson	Full	QS SME	QS Young	Full	QS SME	QS Young
Models	b/se	b/se	b/se	b/se	b/se	b/se
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
post	-0.012	-0.820**	-0.421	-0.997*	-1.386**	-1.265**
	(0.35)	(0.16)	(0.27)	(0.43)	(0.50)	(0.24)
funded	2.504**	1.162+	1.684**	2.583**	1.464	1.602
	(0.42)	(0.65)	(0.48)	(0.70)	(1.00)	(1.08)
post*funded*t1	0.086	1.557**	0.723*	1.013*	1.928**	1.677**
	(0.37)	(0.33)	(0.29)	(0.48)	(0.65)	(0.28)
post*funded*t2	0.334	1.282**	0.792**	1.032*	1.376*	1.701**
	(0.42)	(0.32)	(0.28)	(0.45)	(0.64)	(0.31)
post*funded*t3	0.416	1.319**	0.744*	0.995*	1.427*	1.576**
	(0.36)	(0.48)	(0.30)	(0.43)	(0.63)	(0.24)
post*funded*t4	0.251	1.423**	0.812**	0.427	0.629	1.218*
	(0.38)	(0.37)	(0.31)	(0.49)	(5.92)	(0.55)
constant	-1.513+	-2.145	-0.152	-2.114+	-2.124	-0.211
	(0.79)	(11.19)	(1.15)	(1.15)	(11.18)	(1.65)
lnalpha constant	1.727**	1.250**	1.297**	1.998**	1.516**	1.524**
	(0.14)	(0.26)	(0.19)	(0.14)	(0.22)	(0.33)
N.Obs	1818	729	738	1818	729	738
Log-Likelihood	-2720.374	-826.474	-1077.889	-1387.208	-338.517	-525.375

+ $p<0.10$, * $p<0.05$, ** $p<0.01$

Table IV-6 – Patent data. DiD QML Poisson count regression models with cluster robust standard errors for filed and granted patents filed up to four years after funding, run on all three sample specifications: full sample in second round selection, qualitatively similar SMEs, and qualitatively similar young firms.

Similarly we show results for the effect of mediated academic-industry partnership funding on the count of peer-reviewed publications in Table IV-7. Surprisingly, we find no consistent significant results for the three sample specifications; although for the qualitatively similar sample of SMEs, results are weakly significant for one and three years after funding and significant four years after funding (funded firms publish 3.0 times ($e^{1.091}$) more peer-reviewed papers). Overall the results are such that even though funded firms participate in cross-institutional projects that are arguably based on more basic science, they do not publish their findings in peer-reviewed papers more than unfunded firms. Thus hypothesis 3 is only weakly supported for the qualitatively similar SME sample.

	Publications		
Poisson	Full	QS SME	QS Young
Models	b/se	b/se	b/se
	Model 1	Model 2	Model 3
post	0.343	0.441	0.41
	(0.21)	(0.36)	(0.30)
funded	1.447**	-0.507	0.214
	(0.55)	(1.12)	(1.13)
post*funded*t1	-0.003	0.771+	0.422
	(0.29)	(0.47)	(0.47)
post*funded*t2	0.08	0.653	0.141
	(0.28)	(0.52)	(0.56)
post*funded*t3	0.148	0.896+	0.457
	(0.30)	(0.49)	(0.45)
post*funded*t4	0.275	1.091*	0.656
	(0.30)	(0.45)	(0.48)
constant	-0.904	-1.404	-0.119
	(0.96)	(5.57)	(1.21)
lnalpha constant	1.957**	1.666**	1.477**
	(0.14)	(0.31)	(0.26)
N.Obs	1773	729	702
Log-Likelihood	-1190.805	-392.433	-465.262

+ p<0.10, * p<0.05, ** p<0.01

Table IV-7 – Publication data. DiD QML Poisson count regression models with cluster robust standard errors for the number of peer-reviewed papers up to four years after funding, run on all three sample specifications: full sample in second round selection, qualitatively similar SMEs, and qualitatively similar young firms.

Effects on Impact of Firm Innovations

Beyond assessing the quantity of innovative productivity, we also explore their impact. We employ the commonly used measure of citations to operationalize impact for peer-reviewed papers. These results are shown in an analogous setup in Table IV-8 for all three sample specifications. We find positive results for the effect of mediated funding on innovative impact. For qualitatively similar SMEs in model 2, we find the most consistent significant results with funded firms being cited between 3.3 times ($e^{1.196}$) and 9.6 times ($e^{2.261}$) more than unfunded firms. For qualitatively similar younger firms in model 3, even though the coefficients of interest are sometimes only weakly significant, publications from funded firms are still directionally more cited than those from unfunded ones. Thus, we find evidence for hypothesis 4.

Effects on Collaborative Nature of Firm Innovations

The last set of analyses investigates the collaborative nature of the academic-industry partnership. Table IV-8 shows whether participation in such cross-institutional projects changed the collaborative nature of the innovation produced. Our outcome variable is defined as the number of papers published up to four years after application in which co-authors are affiliated with different institutions. For a publication to count as cross-institutional at least one author has to be from academia while another one from a firm. Surprisingly again, all three sample specifications yield no significant results which implies that researchers in funded firms do not collaborate more with their peers in academic institutions than those in unfunded firms, thus hypothesis 6 is not verified.

	Forward Citations			Cross-Institutions		
Poisson	Full	QS SME	QS Young	Full	QS SME	QS Young
Models	b/se	b/se	b/se	b/se	b/se	b/se
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
post	-0.177	-0.415	-0.42	0.526*	0.673	0.598
	(0.35)	(0.31)	(0.42)	(0.23)	(2.83)	(0.56)
funded	0.861	-1.117	-0.233	1.296**	-0.46	0.364
	(0.70)	(1.75)	(2.04)	(0.41)	(3.27)	(1.48)
post*funded*t1	0.518	2.023*	1.589+	-0.182	0.618	0.182
	(0.62)	(0.82)	(0.82)	(0.27)	(4.02)	(0.74)
post*funded*t2	0.3	1.483**	0.995	-0.105	0.707	0.079
	(0.44)	(0.43)	(0.73)	(0.28)	(2.77)	(0.79)
post*funded*t3	0.633	2.261*	1.784+	0.069	0.789	0.275
	(0.65)	(0.93)	(1.06)	(0.30)	(2.81)	(0.72)
post*funded*t4	0.08	1.196+	0.783	0.336	1.159	0.643
	(0.51)	(0.69)	(0.69)	(0.33)	(2.77)	(0.74)
constant	2.422	1.772	3.129	-0.624	-1.626	-0.303
	(1.81)	(4.09)	(2.34)	(1.02)	(4.89)	(1.85)
lnalpha constant	2.713**	2.446**	2.338**	1.871**	1.604**	1.409**
	(0.14)	(0.21)	(0.19)	(0.14)	(0.39)	(0.37)
N.Obs	1764	729	702	1764	729	702
Log-Likelihood	-14605.259	-5094.846	-6299.557	-1043.267	-352.015	-422.861

+ p<0.10, * p<0.05, ** p<0.01

Table IV-8 – Publication data. DiD QML Poisson count regression models with cluster robust standard errors for the number of forward citations and cross-institutional collaborations of peer-reviewed papers filed up to four years after funding, run on all three sample specifications: full sample in second round selection, qualitatively similar SMEs, and qualitatively similar young firms.

v. Discussion and Conclusion

Implications for Literature

This work provides empirical evidence on the effect of a novel source of governmental funding using mediated academic-industry partnerships on firm innovative performance, by bridging the entrepreneurial finance literature with that studying the interplay between institutions of science and technology. To the best of our knowledge this work is the first to show the effect of such funding sources using a setup that eliminates observable selection bias at the level of the firm.

To summarize our results, we observe compelling evidence that mediated academic-industry partnership funding alleviates capital constraints and increases the financial viability of a firm, thereby decreasing the likelihood of bankruptcy four years after funding. Moreover, academic-industry partnership funding also increases the number of patents firms file and are granted. Thus when provided with more funding, firms are able to take advantage of the extra capital to increase their stock of knowledge and encode them in the familiar method of patents.

Unexpectedly, obtaining academic-industry partnership funding does not increase the number of peer-reviewed papers published by funded firms. Despite the extra capital from funding that should incentivize firms to take on more basic research and higher risk projects as well as spillover effects during the projects while working alongside academic partners, funded firms do not significantly encode more knowledge in peer-reviewed papers compared to unfunded firms. This may be explained by the lack of increased cross-institutional collaborations in peer-reviewed publications despite the cross-institutional composition of partners in the projects. Thus partners are siloed and ingrained within their initial institutional logics, while institutional norms are still prevalent despite participation in a setup conducive to enhanced spillovers. However, the significant positive result for the impact of peer-reviewed publications in funded firms is an indication that even though the amount of basic research encoded in publications is not significantly higher for funded firms, the scientific knowledge that does get published garners more applications and is more easily diffused.

Implications for Practitioners and Policymakers

From these above results, one must be careful in making policy prescriptions. As evidenced by our results when implementing such funding programs, governments are able to incentivize firms in undertaking more R&D projects translated into patents and that become more widely applied as evidenced by increased forward citations of peer-reviewed publications. As a way to help companies remain competitive, governments can view this approach as a potential policy tool for faster application and commercialization.

However, whether firms take on more basic science R&D projects that enable faster and more efficient recognition of spillover opportunities between science and technology is uncertain. Considering peer-reviewed publications as a method primarily used by the scientific institution to encode basic scientific knowledge, the absence of a significant positive result for peer-reviewed publications in funded firms reflects the lack of success in incentivizing for more basic research. The partnership structure of requiring academics to work in collaboration with researchers in private firms suggests that both science and technology develop concurrently instead of via the waterfall model proposed in earlier works (Freeman, 1992; Mansfield, 1995). However, the lack of increased collaboration between scientists from different institutions in funded firms may be an indication that institutional partners are still isolated within the project and that a division of labor between academic scientists and industry scientists is still customary.

This partnership structure creates the potential for a novel model of interaction between the realms of science and technology that strays away from the conventional belief of dedicated gatekeepers that straddle both institutions (Cockburn & Henderson, 1998; Murray, 2004). Instead of having single actors transfer knowledge back and forth between independent silos of science and technology, our setting temporally breaks down the boundaries between the two institutions and enables teams of individuals from both sides to work together alongside one another. However, it is difficult to assess the effectiveness of such a spillover setup from our results.

Limits and Weaknesses

Despite showing interesting outcomes of mediated academic-industry partnership funding on firm innovative performance, this work still suffers from several limitations and weaknesses. Thus, the interpretation of our results should be made with care. We are unable to address an important question for practitioners: how partnerships in which team members come from very different institutional roots can be effectively managed. In effect, we show the relationship between input – mediated funding and output – firm performance – without delving inside what remains a black box. Preliminary qualitative interviews (*n =12*) with project managers of these academic-industry partnership projects indicated that some big challenges they faced were getting individuals from different institutions to align their goals, understand each other and collaborate effectively.

From a policy standpoint, this work has difficulty teasing apart the effect of providing funding from the novel mediated intervention model specific to DNATF since our sample of firms does not provide us with any source of variation on this intervention dimension. As explained in the Setting section, DNATF's mediated intervention model implies active follow-up on each project throughout the project period where a DNATF staff member is assigned and acts as the single point of contact throughout the funded project's lifetime. Compared to more conventional funding schemes where funded projects are left more or less on their own to meet pre-established deliverable deadlines, DNATF stays much closer to each project, frequently mediating conflicts that arise among funded parties.

Finally, since we have studied one specific funding scheme in one specific country, the generalizability of our results may have limitations. However, as we have not concentrated on the intricacies and idiosyncrasies specific to our setting, and instead attempted to explore more largely the effect of funding, we strongly believe that the implications of our results can be interpreted more broadly.

Future Research

Despite these limitations and weaknesses, we have exposed several interesting future research topics beyond the research question explored herein of how mediated academic-industry funding affects firm innovative performance. From a management perspective, understanding the challenges of managing conflict inside partnerships that are "virtual companies" with multiple cross-institutional stakeholders is vital. Research can explore how such projects can be effectively managed and what factors make these projects more successful. For policymakers designing effective funding programs, understanding DNATF's mediated funding and intervention model can offer powerful insights into cross-discipline and cross-boundary project management. Finally, from the perspective of the literature on the micro-foundations of innovation we can lower our level of analysis to understand the effect of such partnerships on individual level productivity and subsequent impact.

V. References

i. References for Chapter II

Azoulay, P., Zivin, J. S. G., & Wang, J. 2010. Superstar Extinction. ***The Quarterly Journal of Economics***, 125(2): 549-589.

Brusoni, S., Prencipe, A., & Pavitt, K. 2001. Knowledge Specialization, Organizational Coupling, and the Boundaries of the Firm: Why Do Firms Know More than They Make? ***Administrative Science Quarterly***, 46(4): 597-621.

Burt, R. S. 2004. Structural Holes and Good Ideas. ***The American Journal of Sociology***, 110(2): 349-399.

Coleman, J. S. 1988. Social Capital in the Creation of Human Capital. ***The American Journal of Sociology***, 94: S95-S120.

Collins, R. 1998. ***The sociology of philosophies : a global theory of intellectual change***. Cambridge, Mass.: Belknap Press of Harvard University Press.

Couzin, J., Enserink, M., & Service, R. F. 2002. Breakthrough of the Year: Small RNAs Make Big Splash. ***Science***, 298(5602): 2296-2303.

Dougherty, D. 1992. Interpretive Barriers to Successful Product Innovation in Large Firms. ***Organization Science***, 3(2): 179-202.

El Ghaoui, L., Li, G.-C., Duong, V.-A., Pham, V., Srivastava, A., & Bhaduri, K. 2011. Sparse Machine Learning Methods for Understanding Large Text Corpora. ***Proc. Conference on Intelligent Data Understanding***.

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. 1998. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. ***Nature***, 391(6669): 806.

Fleming, L., Mingo, S., & Chen, D. 2007. Collaborative Brokerage, Generative Creativity, and Creative Success. ***Administrative Science Quarterly***, 52(3): 443-475.

Fortunato, S. 2010. Community detection in graphs. ***Physics Reports***, 486(3-5): 75-174.

Furman, J. L., & Stern, S. 2011. Climbing atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research. ***American Economic Review***, 101(5): 1933-1963.

Gieryn, T. F., & Hirsh, R. F. 1983. Marginality and Innovation in Science. ***Social Studies of Science***, 13(1): 87-106.

Girotra, K., Terwiesch, C., & Ulrich, K. T. 2010. Idea Generation and the Quality of the Best Idea. ***Management Science***, 56(4): 591-605.

Granovetter, M. S. 1973. The Strength of Weak Ties. ***The American Journal of Sociology***, 78(6): 1360-1380.

Janis, I. L. 1971. Groupthink. ***Psychology Today***, 5(6): 43-46, 74-76.

Jeppesen, L. B., & Lakhani, K. R. 2010. Marginality and Problem Solving Effectiveness in Broadcast Research. ***Organization Science***, Forthcoming.

Jones, B. F. 2009. The Burden of Knowledge and the "Death of the Renaissance Man": Is Innovation Getting Harder? ***Review of Economic Studies***, 76(1): 283-317.

Kaplan, S., & Vakili, K. 2012. Breakthrough innovations: Using topic modeling to distinguish the cognitive from the economic. ***Rotman School Working Paper.***

Kerr, W. R., Lerner, J., & Schoar, A. 2011. The Consequences of Entrepreneurial Finance: Evidence from Angel Financings. ***Review of Financial Studies***.

King, G., Keohane, R. O., & Verba, S. 1994. ***Designing social inquiry : scientific inference in qualitative research***. Princeton, N.J.: Princeton University Press.

King, G., & Zeng, L. 1999a. Logistic Regression in Rare Events Data. ***Department of Government, Harvard University***.

Leonard-Barton, D., & Swap, W. C. 1999. ***When sparks fly : igniting creativity in groups***. Boston, Mass.: Harvard Business School Press.

McEvily, B., & Zaheer, A. 1999. Bridging Ties: A Source of Firm Heterogeneity in Competitive Capabilities. ***Strategic Management Journal***, 20(12): 1133-1156.

McFadyen, M. A., & Cannella, A. A. J. 2004. Social Capital and Knowledge Creation: Diminishing Returns of the Number and Strength of Exchange Relationships ***Academy of Management Journal***, 47(5): 735-746.

Merton, R. K. 1949. ***Social theory and social structure*** (1968 enl. ed.). New York: Free Press.

Mowery, D. C., & Ziedonis, A. A. 2002. Academic patent quality and quantity before and after the Bayh-Dole act in the United States. ***Research Policy***, 31(3): 399-418.

Obstfeld, D. 2005. Social Networks, the Tertius Iungens Orientation, and Involvement in Innovation. ***Administrative Science Quarterly***, 50(1): 100-130.

Reagans, R., & McEvily, B. 2003. Network structure and knowledge transfer: The effects of closure and range. ***Administrative Science Quarterly***, 48: 240-267.

Romer, P. M. 1987. Growth Based on Increasing Returns Due to Specialization. ***The American Economic Review***, 77(2): 56-62.

Schumpeter, J. A. 1942. ***Capitalism, socialism, and democracy*** (1st Harper Perennial Modern Thought ed.). New York: HarperPerennial.

Simonton, D. K. 1989. Age and creative productivity: Nonlinear estimation of an information-processing model. ***International Journal of Aging and Human Development***, 29: 23-37.

Simonton, D. K. 1999. ***Origins of genius : Darwinian perspectives on creativity***. New York: Oxford University Press.

Singh, J., & Fleming, L. 2010. Lone Inventors as Sources of Breakthroughs: Myth or Reality? ***Management Science***, 56(1): 41-56.

Stern, S. 2004. Do Scientists Pay to Be Scientists? ***Management Science***, 50(6): 835-853.

Swanson, D. R., Smalheiser, N. R., & Torvik, V. I. 2006. Ranking indirect connections in literature-based discovery: The role of medical subject headings. ***Journal of the American Society for Information Science & Technology***, 57(11): 1427-1439.

Terwiesch, C., & Ulrich, K. T. 2009. ***Innovation tournaments : creating and selecting exceptional opportunities***. Boston, Mass.: Harvard Business Press.

Torvik, V. I., & Smalheiser, N. R. 2009. Author Name Disambiguation in MEDLINE. ***ACM transactions on knowledge discovery from data***, 3(3): 1-29.

Uzzi, B. 1997. Social Structure and Competition in Interfirm Networks: The Paradox of Embeddedness. ***Administrative Science Quarterly***, 42(1): 35-67.

Wuchty, S., Jones, B. F., & Uzzi, B. 2007. The Increasing Dominance of Teams in Production of Knowledge. ***Science***, 316(5827): 1036-1039.

ii. References for Chapter III

Azoulay, P., Graff Zivin, J. S., & Manso, G. 2011. Incentives and creativity: evidence from the academic life sciences. ***The RAND Journal of Economics***, 42(3): 527-554.

Berson, J. A. 1992. Discoveries missed, discoveries made: creativity, influence, and fame in chemistry. ***Tetrahedron***, 48(1): 3-17.

Bijker, W. E., Hughes, T. P., & Pinch, T. J. 1987. ***The Social construction of technological systems : new directions in the sociology and history of technology***. Cambridge, Mass.: MIT Press.

Burgelman, R. A. 1983. A Process Model of Internal Corporate Venturing in the Diversified Major Firm. ***Administrative Science Quarterly***, 28(2): 223-244.

Burt, R. S. 2004. Structural Holes and Good Ideas. ***The American Journal of Sociology***, 110(2): 349-399.

Chai, S., & Fleming, L. 2012. Bibiometric Predictions of Scientific Breakthroughs: A Cautionary Tale. ***Working paper***.

Cialdini, R. B., & Trost, M. R. 1998. Social influence: Social norms, conformity and compliance. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), ***The handbook of social psychology, Vols. 1 and 2 (4th ed.)***: 151-192. New York, NY, US: McGraw-Hill.

Cockburn, I. M., & Henderson, R. M. 1998. Absorptive Capacity, Coauthoring Behavior, and the Organization of Research in Drug Discovery. ***The Journal of Industrial Economics***, 46(2): 157-182.

Cohen, W. M., & Levinthal, D. A. 1990. Absorptive Capacity: A New Perspective on Learning and Innovation. ***Administrative Science Quarterly***, 35(1): 128-152.

Collins, R. 1998. ***The sociology of philosophies : a global theory of intellectual change***. Cambridge, Mass.: Belknap Press of Harvard University Press.

Corbin, J. M., & Strauss, A. L. 2008. ***Basics of qualitative research : techniques and procedures for developing grounded theory*** (3rd ed.). Los Angeles, Calif.: Sage Publications, Inc.

Couzin, J., Enserink, M., & Service, R. F. 2002. Breakthrough of the Year: Small RNAs Make Big Splash. ***Science***, 298(5602): 2296-2303.

Dasgupta, P., & David, P. A. 1994. Toward a new economics of science. ***Research Policy***, 23(5): 487-521.

Dosi, G. 1982. Technological paradigms and technological trajectories : A suggested interpretation of the determinants and directions of technical change. ***Research Policy***, 11(3): 147-162.

Dougherty, D. 1992. Interpretive Barriers to Successful Product Innovation in Large Firms. ***Organization Science***, 3(2): 179-202.

Dyson, F. J. 1972. Missed Opportunities. ***Bulletin (New Series) of the American Mathematical Society***, 78(5): 635-652.

Edmondson, A. 1999. Psychological Safety and Learning Behavior in Work Teams. ***Administrative Science Quarterly***, 44(2): 350-383.

Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., & Tuschl, T. 2001. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. ***Nature***, 411(6836): 494-498.

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. 1998. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. ***Nature***, 391(6669): 806.

Fire, A. Z. 2007. Gene silencing by double-stranded RNA. ***Cell Death & Differentiation***, 14(12): 1998-2012.

Fleming, L. 2001. Recombinant Uncertainty in Technological Search. ***Management Science***, 47(1): 117-132.

Fleming, L. 2002. Finding the organizational sources of technological breakthroughs: the story of Hewlett-Packard's thermal ink-jet. ***Industrial & Corporate Change***, 11(5): 1059-1084.

Fleming, L., & Sorenson, O. 2004. Science as a Map in Technological Search. ***Strategic Management Journal***, 25(8/9): 909-928.

Galison, P. 1997. ***Image and logic : a material culture of microphysics***. Chicago: University of Chicago Press.

Garud, R., & Rappa, M. A. 1994. A Socio-Cognitive Model of Technology Evolution: The Case of Cochlear Implants. ***Organization Science***, 5(3): 344-362.

Gieryn, T. F., & Hirsh, R. F. 1983. Marginality and Innovation in Science. ***Social Studies of Science***, 13(1): 87-106.

Gilbert, W. 1986. Origin of life: The RNA world. ***Nature***, 319(6055): 618-618.

Girotra, K., Terwiesch, C., & Ulrich, K. T. 2010. Idea Generation and the Quality of the Best Idea. ***Management Science***, 56(4): 591-605.

Golden, B. R. 1992. The Past Is the Past--Or Is It? The Use of Retrospective Accounts as Indicators of past Strategy. ***The Academy of Management Journal***, 35(4): 848-860.

Granovetter, M. S. 1973. The Strength of Weak Ties. ***The American Journal of Sociology***, 78(6): 1360-1380.

Guo, S., & Kemphues, K. J. 1995. par-1, a gene required for establishing polarity in C. elegans embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. ***Cell***, 81(4): 611-620.

Hamilton, A. J., & Baulcombe, D. C. 1999. A Species of Small Antisense RNA in Posttranscriptional Gene Silencing in Plants. ***Science***, 286(5441): 950-952.

Hargadon, A., & Sutton, R. I. 1997. Technology Brokering and Innovation in a Product Development Firm. ***Administrative Science Quarterly***, 42(4): 716-749.

Henderson, R. M., & Clark, K. B. 1990. Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms. ***Administrative Science Quarterly***, 35(1): 9-30.

Jaffe, A. B., Trajtenberg, M., & Henderson, R. 1993. Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. ***The Quarterly Journal of Economics***, 108(3): 577-598.

Jeppesen, L. B., & Lakhani, K. R. 2010. Marginality and Problem Solving Effectiveness in Broadcast Research. ***Organization Science***, Forthcoming.

Jones, B. F. 2009. The Burden of Knowledge and the "Death of the Renaissance Man": Is Innovation Getting Harder? ***Review of Economic Studies***, 76(1): 283-317.

Kaplan, S., & Vakili, K. 2012. Breakthrough innovations: Using topic modeling to distinguish the cognitive from the economic. ***Rotman School Working Paper.***

Kogut, B., & Zander, U. 1992. Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology. ***Organization Science***, 3(3): 383-397.

Krol, A. R. v. d., Leon, A. M., Beld, M., Mol, J. N. M., & Stuitje, A. R. 1990. Flavonoid Genes in Petunia: Addition of a Limited Number of Gene Copies May Lead to a Suppression of Gene Expression. ***The Plant Cell***, 2(4): 291-299.

Kuhn, T. S. 1962. ***The structure of scientific revolutions***. Chicago: University of Chicago Press.

Lane, J. 2009. Assessing the Impact of Science Funding. ***Science*** 324: 1273-1275.

Lee, R. C., Feinbaum, R. L., & Ambros, V. 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. ***Cell***, 75(5): 843-854.

Leonard-Barton, D., & Swap, W. C. 1999. ***When sparks fly : igniting creativity in groups***. Boston, Mass.: Harvard Business School Press.

March, J. G. 1991. Exploration and Exploitation in Organizational Learning. ***Organization Science***, 2(1): 71-87.

McEvily, B., & Zaheer, A. 1999. Bridging Ties: A Source of Firm Heterogeneity in Competitive Capabilities. ***Strategic Management Journal***, 20(12): 1133-1156.

McFadyen, M. A., & Cannella, A. A. J. 2004. Social Capital and Knowledge Creation: Diminishing Returns of the Number and Strength of Exchange Relationships ***Academy of Management Journal***, 47(5): 735-746.

McFadyen, M. A., Semadeni, M., & Cannella, J. A. A. 2009. Value of Strong Ties to Disconnected Others: Examining Knowledge Creation in Biomedicine. ***Organization Science***, 20(3): 552-564.

Meister, G., & Tuschl, T. 2004. Mechanisms of gene silencing by double-stranded RNA. ***Nature***, 431(7006): 343-349.

Merton, R. K. 1957. Priorities in Scientific Discovery: A Chapter in the Sociology of Science. ***American Sociological Review***, 22(6): 635-659.

Miles, M. B., & Huberman, A. M. 1984. ***Qualitative data analysis : a sourcebook of new methods***. Beverly Hills: Sage Publicaions.

Mueller, P. 2006. Exploring the knowledge filter: How entrepreneurship and university,Äìindustry relationships drive economic growth. ***Research Policy***, 35(10): 1499-1508.

Murray, F. 2002. Innovation as co-evolution of scientific and technological networks: exploring tissue engineering. ***Research Policy***, 31(8-9): 1389-1403.

Napoli, C., Lemieux, C., & Jorgensen, R. 1990. Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. ***The Plant Cell***, 2(4): 279-289.

Nemeth, C. J. 1986. Differential contributions of majority and minority influence. ***Psychological Review***, 93(1): 23-32.

Obstfeld, D. 2005. Social Networks, the Tertius Iungens Orientation, and Involvement in Innovation. ***Administrative Science Quarterly***, 50(1): 100-130.

Ratcliff, F., Harrison, B. D., & Baulcombe, D. C. 1997. A Similarity Between Viral Defense and Gene Silencing in Plants. ***Science***, 276(5318): 1558-1560.

Reinhart, B. J., Slack, F. J., Basson, M., Pasquinelli, A. E., Bettinger, J. C., Rougvie, A. E., Horvitz, H. R., & Ruvkun, G. 2000. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. ***Nature***, 403(6772): 901-906.

Rivkin, J. W., & Siggelkow, N. 2002. Organizational sticking points on NK Landscapes. ***Complexity***, 7(5): 31-43.

Romano, N., & Macino, G. 1992. Quelling: transient inactivation of gene expression in Neurospora crassa by transformation with homologous sequences. ***Molecular Microbiology***, 6(22): 3343-3353.

Schumpeter, J. A. 1934. ***The theory of economic development; an inquiry into profits, capital, credit, interest, and the business cycle***. Cambridge, Mass.: Harvard University Press.

Schumpeter, J. A. 1942. ***Capitalism, socialism, and democracy*** (1st Harper Perennial Modern Thought ed.). New York: HarperPerennial.

Simonton, D. K. 1989. Age and creative productivity: Nonlinear estimation of an information-processing model. ***International Journal of Aging and Human Development***, 29: 23-37.

Singh, J., & Fleming, L. 2010. Lone Inventors as Sources of Breakthroughs: Myth or Reality? ***Management Science***, 56(1): 41-56.

Teece, D. J., Pisano, G., & Shuen, A. 1997. Dynamic Capabilities and Strategic Management. ***Strategic Management Journal***, 18(7): 509-533.

Thomke, S. H. 2003. ***Experimentation matters : unlocking the potential of new technologies for innovation***. Boston Mass.: Harvard Business School Press.

Thompson, P., & Fox-Kean, M. 2005. Patent Citations and the Geography of Knowledge Spillovers: A Reassessment: Reply. ***The American Economic Review***, 95(1): 465-466.

Torvik, V. I., & Smalheiser, N. R. 2009. Author Name Disambiguation in MEDLINE. ***ACM transactions on knowledge discovery from data***, 3(3): 1-29.

Uzzi, B. 1997. Social Structure and Competition in Interfirm Networks: The Paradox of Embeddedness. ***Administrative Science Quarterly***, 42(1): 35-67.

Waterhouse, P. M., Graham, M. W., & Wang, M.-B. 1998. Virus Resistance and Gene Silencing in Plants can be Induced by Simultaneous Expression of Sense and Antisense RNA. ***Proceedings of the National Academy of Sciences of the United States of America***, 95(23): 13959-13964.

Wuchty, S., Jones, B. F., & Uzzi, B. 2007. The Increasing Dominance of Teams in Production of Knowledge. ***Science***, 316(5827): 1036-1039.

Zamore, P. D., Tuschl, T., Sharp, P. A., & Bartel, D. P. 2000. RNAi: Double-Stranded RNA Directs the ATP-Dependent Cleavage of mRNA at 21 to 23 Nucleotide Intervals. ***Cell***, 101(1): 25-33.

iii. References for Chapter IV

Admati, A. R., & Pfleiderer, P. 1994. Robust Financial Contracting and the Role of Venture Capitalists. ***The Journal of Finance***, 49(2): 371-402.

Agrawal, A., & Henderson, R. 2002. Putting Patents in Context: Exploring Knowledge Transfer from MIT. ***Management Science***, 48(1): 44-60.

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

Azoulay, P., Graff Zivin, J. S., & Manso, G. 2011. Incentives and creativity: evidence from the academic life sciences. ***The RAND Journal of Economics***, 42(3): 527-554.

Bernstein, S. 2012. Does going public affect innovation? ***Working paper***.

Black, S. E., & Strahan, P. E. 2002. Entrepreneurship and Bank Credit Availability. ***The Journal of Finance***, 57(6): 2807-2833.

Cockburn, I. M., & Henderson, R. M. 1998. Absorptive Capacity, Coauthoring Behavior, and the Organization of Research in Drug Discovery. ***The Journal of Industrial Economics***, 46(2): 157-182.

Cohen, W. M., & Levinthal, D. A. 1990. Absorptive Capacity: A New Perspective on Learning and Innovation. ***Administrative Science Quarterly***, 35(1): 128-152.

Dasgupta, P., & David, P. A. 1994. Toward a new economics of science. ***Research Policy***, 23(5): 487-521.

Fleming, L., & Sorenson, O. 2004. Science as a Map in Technological Search. ***Strategic Management Journal***, 25(8/9): 909-928.

Freeman, J. 1992. ***Formal scientific and technical institutions in the national systems of innovation.*** . London :New York: Pinter Publishers ;Distributed exclusively in the USA and Canada by St. Martin's Press.

Hellmann, T. 1998. The Allocation of Control Rights in Venture Capital Contracts. ***The RAND Journal of Economics***, 29(1): 57-76.

Kerr, W. R., Lerner, J., & Schoar, A. 2011. The Consequences of Entrepreneurial Finance: Evidence from Angel Financings. ***Review of Financial Studies***.

Kortum, S., & Lerner, J. 2000. Assessing the Contribution of Venture Capital to Innovation. ***The RAND Journal of Economics***, 31(4): 674-692.

Lai, R., D'Amour, A., & Fleming, L. 2009. The careers and co-authorship networks of U.S. patent-holders, since 1975. ***Harvard Business School Working Paper***.

Lee, D. S., & Lemieux, T. 2010. Regression Discontinuity Designs in Economics. ***Journal of Economic Literature***, 48(2): 281-355.

Mansfield, E. 1995. Academic Research Underlying Industrial Innovations: Sources, Characteristics, and Financing. ***The Review of Economics and Statistics***, 77(1): 55-65.

Merton, R. K. 1957. Priorities in Scientific Discovery: A Chapter in the Sociology of Science. ***American Sociological Review***, 22(6): 635-659.

Mueller, P. 2006. Exploring the knowledge filter: How entrepreneurship and university‚Äìindustry relationships drive economic growth. ***Research Policy***, 35(10): 1499-1508.

Murray, F. 2002. Innovation as co-evolution of scientific and technological networks: exploring tissue engineering. ***Research Policy***, 31(8-9): 1389-1403.

Murray, F. 2004. The role of academic inventors in entrepreneurial firms: sharing the laboratory life. ***Research Policy***, 33(4): 643-659.

Murray, F. 2010. The Oncomouse That Roared: Hybrid Exchange Strategies as a Source of Distinction at the Boundary of Overlapping Institutions. ***American Journal of Sociology***, 116(2): 341-388.

Murray, F., & Stern, S. 2007. Do formal intellectual property rights hinder the free flow of scientific knowledge?: An empirical test of the anti-commons hypothesis. ***Journal of Economic Behavior & Organization***, 63(4): 648-687.

Nelson, R. R. 1959. The Simple Economics of Basic Scientific Research. ***The Journal of Political Economy***, 67(3): 297-306.

Nelson, R. R. 1995. Recent Evolutionary Theorizing About Economic Change. ***Journal of Economic Literature***, 33(1): 48-90.

Rosenberg, N. 1990. Why do firms do basic research (with their own money)? ***Research Policy***, 19(2): 165-174.

Samila, S., & Sorenson, O. 2011. Venture Capital, Entrepreneurship and Economic Growth. ***Review of Economics & Statistics***, 93(1): 338-349.

Stokes, D. E., & Brookings, I. 1997. ***Pasteur's quadrant : basic science and technological innovation***. Washington, D.C.: Brookings Institution Press.

Varga, A., & Schalk, H. J. 2004. Knowledge Spillovers, Agglomeration and Macroeconomic Growth: An Empirical Approach. ***Regional Studies***, 38(8): 977-989.

VI. Appendix

i. Interview Questions for Chapter III

Open-Ended Questions

- Describe your work leading to 1997, in 1998 and after 1998.

Probing Questions

Breakthrough

- In the period of 1997-1998 were you and your peers aware that a breakthrough was about to be discovered? Was there excitement due to a potential impactful discovery?
- Were scientists trying to solve a specific puzzling mechanism or did they just happen to stumble on the RNAi mechanism by chance while looking for something else?
- Were there many teams working towards solving the same problem? Was there racing?
- Do you feel like the breakthrough could have been made earlier? Why? What was the missing link that prevented it?
- Was the discovery and its results a surprise? In terms of simplicity or complexity of the solution, in terms of who made the discovery?
- Before you chose your research direction, how do you evaluate the potential impact of your research? How?

- What papers or findings spurred your interest in RNAi research? What works had a decisive influence on your research interests?
- What experiments, field or prior breakthroughs do you believe paved the road to the discovery? What inventions (tools), environment fostered the discovery?
- Were you aware of the similar co-suppression and quelling results obtained in plants and fungi? / As a plant scientist did you think that co-suppression and quelling would be present in animals?

Community

- Was there a defined community of RNAi scientists prior to breakthrough?
- How would you define the community of RNAi scientists prior to breakthrough? Which subfields of biology came together to form such a community?
- How would you characterize this community? Social, open or collective?
- How open was the community of scientists working towards solving this discovery? Was there an informal group established that frequently communicated and shared their ideas? Or were results withheld?
- What kind of conference/research seminars did you attend at the time, was it phenomenon-based, organism-based or something else?
- How do you think about conferences? What role do conferences play in your research?
- In your opinion, did the breakthrough come from within the community or from outside?
- In your opinion, who were the big contenders in the community to discover the mechanism to RNAi? Why?

This page is intentionally left blank.