Organization Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Exploring Uncharted Territory: Knowledge Search Processes in the Origination of Outlier Innovation

Madeline K. Kneeland, Melissa A. Schilling, Barak S. Aharonson

Published in Organization Science on April 01, 2020 as DOI: 10.1287/orsc.2019.1328.
This article has not been copyedited or formatted. The final version may differ from this version.

informs
http://pubsonline.informs.org/journal/orsc

ORGANIZATION SCIENCE
Vol. 31, No. 3, May–June 2020, pp. 535–557
ISSN 1047-7039 (print), ISSN 1526-5455 (online)

Exploring Uncharted Territory: Knowledge Search Processes in the Origination of Outlier Innovation

Madeline K. Kneeland,[a] Melissa A. Schilling,[b] Barak S. Aharonson[c,d]

[a] SC Johnson College of Business, Cornell University, Ithaca, New York 14853; [b] Stern School of Business, New York University, New York, New York 10012; [c] Strategic Management and Entrepreneurship, Xiamen School of Management, Xiamen University, 361005 Fujian, China; [d] Coller School of Management, Tel Aviv University, Tel Aviv 6997801, Israel

Contact: kneeland@cornell.edu, https://orcid.org/0000-0002-9097-8998 (MKK); mschilli@stern.nyu.edu, https://orcid.org/0000-0002-3573-7614 (MAS); aharonson@tau.ac.il, https://orcid.org/0000-0002-7031-2216 (BSA)

Received: April 24, 2017
Revised: March 18, 2018; May 6, 2019; July 18, 2019
Accepted: July 29, 2019
Published Online in Articles in Advance: April 1, 2020

https://doi.org/10.1287/orsc.2019.1328

Abstract. Most innovation builds closely on existing knowledge and technology, delivering incremental advances on existing ideas, products, and processes. Sometimes, however, inventors make discoveries that seem very distant from what is known and well understood. How do individuals and firms explore such uncharted technological terrain? This paper extends research on knowledge networks and innovation to propose three main processes of knowledge creation that are more likely to result in discoveries that are distant from existing inventions: *long search paths*, *scientific reasoning*, and *distant recombination*. We explore these processes with a combination of a large and unique data set on outlier patents filed at the U.S. Patent and Trademark Office and interviews with inventors of outlier patents. Our exploratory analysis suggests that there are significant differences in the inventor teams, assignees, and search processes that result in outlier patents. These results have important implications for managers who wish to encourage a more exploratory search for breakthrough innovation.

Keywords: innovation • patents • breakthrough • outliers • recombination • knowledge networks

Introduction

The universe of existing and potential technology is often conceived of as a landscape within which innovators search (Levinthal 1997; Kauffman et al. 2000; Fleming and Sorenson 2001, 2004; Rosenkopf and Almeida 2003; Ahuja and Katila 2004).[1] This landscape has peaks of exceptional opportunity, and valleys where opportunities are small or sparse. Innovators seek out the peaks of opportunity, drawing from their own experience and following the clues left by others.

Most innovators will search in the "neighborhood" of prior successful innovations—practicing what is termed "local search" (Cyert and March 1963, Nelson and Winter 1982, Dosi 1988, Cohen and Levinthal 1990). As knowledge and inventions are accumulated in a domain, they reveal the relationships between different technological elements and provide insight into new directions or permutations likely to be valuable. Well-trodden areas of the technology landscape are therefore more efficient to search. Even if the innovator does not have direct experience in the area, he or she reaps some advantages of the experience of others by searching in the vicinity of prior successful innovations. In essence, any inventor's progress up an opportunity peak provides valuable guideposts for other inventors to follow. Furthermore, the amount of innovation activity in an area is likely to be an indicator of the perceived quality of the underlying opportunities. This leads most innovators to build on established lines of inquiry, incrementally expanding or refining a technological domain (March and Simon 1958, Nelson and Winter 1982, Helfat 1994, Rosenkopf and Nerkar 2001). Consistent with this, if one examines the topology and evolution of the patent landscape, one finds that most patents are formed in close technological proximity to other patents (Aharonson and Schilling 2015).

It is also possible, however, for inventors to pioneer uncharted territory in the technology landscape and to even make discoveries that are very distant from what is known and well understood—that is, they make "long jumps" (Kauffman 1993, Levinthal 1997). Inventors might stumble upon a new peak through serendipitous discovery, such as in 1928, when Alexander Fleming noticed how a spot of mold on a laboratory culture of bacteria was surrounded by an area in which the bacteria did not grow, leading to the development of antibiotics (Koestler 1973). This serendipitous discovery is closest to Kauffman's original concept of a long jump being a random draw on dimensions of a fitness landscape (Kauffman 1993, p. 212). Other times, inventors might more purposefully explore uncharted technological space.

Understanding the processes by which inventors can explore uncharted technological space is important. Unknown domains are likely to have more untried technological combinations and thus offer more opportunity for breakthrough discovery precisely because they reveal opportunity in a new domain (Fleming 2001). Though it is typically assumed that exploring unknown spaces ("exploratory search") will result in a higher failure rate because there is less knowledge upon which to build, exploring unknown spaces is also assumed to yield more radical innovations, and some proportion of those will turn out to be exceptionally important, opening up new fields of scientific opportunity and creating new directions for technological advance. This suggests that the ability to make long jumps is important for breakthrough discovery. But other than random serendipity, how do inventors identify opportunities in what appears to be unknown space? What are the nonrandom search processes that enable inventors to make long jumps? This leads to our research question here: Can we identify systematic differences in the search processes underlying outlier innovation from nonoutlier innovation?

To explore these questions, we draw from a cognitive network perspective. In several areas of cognitive science, the cognitive process is conceptualized as a process of building a network of associations that form mental representations of what the individual knows or believes (Fahlman 1989, Martindale 1995, Simonton 1999, Steyvers and Tenenbaum 2002). Individuals form their own mental "maps" of the world and are continually adding, deleting, revising, or reinforcing elements of this map.

There are important similarities and differences between this cognitive network map and the technological landscape construct typically used in the innovation search research. Although both are representations of what is believed to be reality, a technology landscape perspective assumes that there can be some objective representation of technological opportunity that is common to all (or most) and attempts to represent what has already been discovered by anyone, whereas cognitive networks are representations that are specific to the knowledge in the minds of individuals. Cognitive networks vary across individuals and may differ substantively from what is assumed to be known about the technology landscape. Individuals, for example, do not know the whole landscape—their mental representations of the landscape are more complete or accurate in areas in which they have experience and expertise, and are sparse or inaccurate in others. Furthermore, an individual may know or suspect more than what has been publicly demonstrated and have hypotheses about relationships between dimensions on the landscape that are different from what others believe. Both the cognitive network and the technology landscape perspective may be incomplete or biased. The former, however, is in the mind and is personally constructed and idiosyncratic to the individual, whereas the latter is intended to be impersonal and to represent a collective agreement about what is known and where technology opportunities are likely to be.

Much of the previous research on innovation search has tended to focus on how inventors or firms balance exploration versus exploitation, such as through temporal sequencing or ambidexterity (e.g., Brown and Eisenhardt 1998, Nickerson and Zenger 2002, Siggelkow and Levinthal 2003, O'Reilly and Tushman 2016), creating subgroups within the firm (e.g., Fang et al. 2010), the way decisions are delegated or aggregated in organizations (e.g., Rivkin and Siggelkow 2003, Siggelkow and Rivkin 2006, Csaszar 2012), or choices of partners (e.g., Rosenkopf and Almeida 2003, Rothaermel and Deeds 2004). Our work here has a different focus. We are explicitly interested in the processes that *underlie* exploration. Work in this domain is sparser but includes important studies, such as work on deduction (Fleming and Sorenson 2004), work on how structure can influence opportunity detection (Csaszar and Eggers 2013), work on imperfect imitation or information distortion (e.g., Rivkin 2000, Csaszar and Siggelkow 2010, Schilling and Fang 2016), and work on mental representations (e.g., Gavetti and Levinthal 2000, Csaszar and Levinthal 2016). We extend this latter body of work by using a cognitive perspective to show how individuals can build a pattern of associations that enables them to arrive at a mental representation of reality that is very different from what they and others believed before. These associations may be influenced by others, thus team structure and expertise are decidedly relevant, but our focus is explicitly on the cognitive mechanism: the long jump in the mind of an individual.

In the sections that follow, we first use a technology landscape method to identify "outlier" inventions. We then utilize a cognitive network approach to derive ways in which inventors might make such outlier discoveries, and explore these ideas with qualitative comments from interviews with inventors and quantitative data on patents. In the final section, we discuss our conclusions from this exploratory analysis and their implications for future research and practice.

Table 1. Mean Comparisons of Outlier to Nonoutlier Patents, 1990–2000

	Nonoutlier	Outlier	Difference	t
Forward citations	16.969	20.802	−3.833**	−39.517
Individual inventor assignee	0.144	0.186	−0.043**	−40.246
Time to grant (years and months)	2.246	2.300	−0.054**	−13.374
Allowed claims	15.613	16.528	−0.915**	−19.880
Number of inventors	2.298	2.238	0.060**	12.078
Number of prior patents (team) in previous 10 years	21.731	17.285	4.446**	31.519
N	1,414,921	120,957		

**$p < 0.001$.

Outlier Patents

Consistent with the definition of "outliers" as "that which lies, or is away from, the main body" (Merriam-Webster 2019), we define "outlier patents" as patents that occupy technological positions that are away from the body of existing patents at their time of filing—that is, they are unusual in terms of their technological combinations. Specifically, we first identified all patents applied for (and subsequently granted) between 1990 and 2000, inclusive, from the U.S. Patent and Trademark Office (USPTO) database. We then utilized a novel technology position measure created by Aharonson and Schilling (2015) that produces individual vectors for each patent with binary indicators for each of the 9,864 mainline subclasses. The binary indicators represent a possible $2^{9,864}$ potential technology positions. Patents are considered "adjacent" to other patents that differ by only one mainline subclass. In other words, adjacent patents can be thought of as one technological step away from one another. Outlier patents are those with no adjacencies at the time of granting—that is, their binary vectors have at least two different mainline subclasses from any other existing patent in the entire USPTO database, indicating that they represent a novel combination of technological components that is at least two steps away from all preexisting patents at the time of filing (Aharonson and Schilling 2015) (for a more detailed explanation, please see Appendix A). This is, we believe, the first use of this fine-grained measure of identifying outlying patents. Of all the utility patents applied for (and subsequently granted) between 1990 and 2000 (1,535,878 patents), 120,957 (7.88%) of all patents in that time period were identified as outliers at their time of filing.[2]

Using the aforementioned binary vector measure, we are also able to identify the distance of outlying patents from the nearest occupied position. As noted earlier, our minimum criteria for being an outlier patent is two technological steps away, but some outliers were as distant as 18 steps away. We supplemented the USPTO data with data from the National Bureau of Economic Research (NBER), the Patent Name-Matching Project (Hall 2008),[3] the Disambiguated Inventor Database (Li et al. 2014), and the Patent Citation to Science Project (Marx 2019, Marx and Fuegi 2019). Table 1 compares descriptive means between the outlier patent population and nonoutlier patents on key dimensions, including the size of the inventing team and the number of allowed claims on the patent. Although our focus is on the processes involved in the creation of outliers, it worth noting that outlier patents end up garnering more forward citations, on average, as compared with nonoutlier patents ($t = 39.517$, $p < 0.001$).[4] Table 2 compares the technological subcategories (based on category definitions provided by Hall et al. 2001) in which outlier and nonoutlier patents are most common. Although the categories are more similar than we might have expected, outliers are more concentrated within these technological categories (65% across the top 10 subcategories) as compared with nonoutlier patents (54.5% across the top 10 subcategories).

To gain a richer contextual understanding of the invention of outlier patents, we attempted to track down the inventors of the 30 patents that were most technologically distant from any other existing patent at their time of filing for interviews. These patents ranged in distance from 11 to 18 steps away from any other existing patent. Of the 93 inventors represented on these 30 patents, we succeeded in finding contact information for 20 of these inventors and conducted 10 in-depth interviews with one or more inventors from nine of the most technologically distant outlying patents (see Appendix B for additional details). The interviews were unstructured and began with our explaining that we wanted to understand how inventors made unusual discoveries, our definition of outlier patents, and identification of their outlier patent that prompted our contact. We then asked them, "Can you please tell us the story of how you came to make the discovery represented in this patent?" From that point on, we asked only clarifying questions such as an explanation of a term or an approximate date of an event. These interviews lasted between 20 minutes and an hour and a half and (in most cases) resulted in detailed inventor narratives.[5] We did not ask the inventors about the dimensions

Table 2. Top 10 NBER Technological Subcategories for Outlier and Nonoutlier Patents, 1990–2000

	Nonoutlier	%	Outlier	%
1.	Miscellaneous chemical	7.83	Miscellaneous chemical	12.98
2.	Communications	7.62	Miscellaneous others	12.64
3.	Miscellaneous others	7.28	Materials processing and handling	7.16
4.	Drugs	6.48	Computer hardware and software	6.63
5.	Computer hardware and software	6.21	Drugs	5.90
6.	Surgery and medical instruments	4.05	Communications	5.30
7.	Semiconductor devices	4.03	Semiconductor devices	4.83
8.	Miscellaneous mechanical	3.90	Miscellaneous electrical	3.49
9.	Materials processing and handling	3.58	Miscellaneous mechanical	3.22
10	Information storage	3.52	Information storage	2.92

we thought would be theoretically relevant based on a cognitive network perspective, though several inventors made unsolicited comments that clearly evoke these dimensions, as detailed further in the next sections.

Knowledge Creation and Technological Search

Technological invention is the creation and application of new knowledge in the form of a useful new idea, device, or method. To really understand the dynamics of technological invention, then, we must start by understanding how knowledge is created. Knowledge creation at both the individual and supraindividual (e.g., team, organization) levels can be usefully represented as a cognitive network process.[6] New information (e.g., individual facts, data, or signals) only becomes knowledge when it is integrated into a pattern of associations that give it meaning (Bartlett 1932, Mayer and Greeno 1972). When individuals are confronted with new information, they search for connections between the new information and their existing knowledge network; the network of patterns within which the information is embedded structures how that information is understood, and how that information relates to what is believed to be true (Nonaka 1994). Knowledge creation is thus the process of integrating new information into a network of associations, or recombining existing information in new ways. In network terms, knowledge creation is the addition or change of "nodes" (e.g., ideas, facts) and/or "links" (i.e., the relationships between the nodes). Both the amount and diversity of new information integrated into the network, or the degree of change in the way existing knowledge is combined, will influence how novel the individual perceives the new knowledge to be (Schilling 2005).

Most knowledge creation will be incremental additions and modifications of the existing knowledge network. New information can be more readily assimilated when it has many obvious connections to the existing network (Ellis 1965, Cohen and Levinthal 1990, Schilling 2005). Incrementally extending or revising an existing knowledge network is thus much more efficient than attempting to search for and understand information that has fewer or less obvious connections to an individual's existing knowledge base. Analogously, and in part a result of the preceding, most technological searches will be in the vicinity of existing inventions (March 1991, Fleming 2001, Katila and Ahuja 2002, Aharonson et al. 2007, Aharonson and Schilling 2015). Inventors tend to build on the existing knowledge base that is manifest in the form of existing inventions because the opportunities may be more obvious and the search process more efficient.

Patents help to make the knowledge base underlying an invention public and, thus, contribute to the knowledge networks of other inventors. Existing patented inventions can provide inventors knowledge about both the inputs to the invention (e.g., design, materials, processes) and outcomes (e.g., what worked, what the potential uses of the technology are), in essence transferring some of the learning benefits accrued by the inventor of the existing invention to many other potential inventors. Existing patented inventions are thus repositories of knowledge upon which other inventors can build; they provide a scaffolding that inventors can incorporate into their own knowledge networks. As a result of searching this grafted knowledge network, inventors are more likely to make discoveries that are within or adjacent to known technological terrain rather than to explore uncharted territory (Aharonson and Schilling 2015).

But how do inventors come up with inventions that are not only new to the world but are also distant from the existing knowledge base? We first discuss the role of serendipity in discovery. We then explore three search processes that might be used alone or in combination: *long search paths*, *scientific reasoning*, and *distant recombination*. We use interview data to further develop our ideas about the roles of the three search processes before using patent data to refine empirically the prevalence and magnitude of scientific reasoning and distant recombination. For reasons we explain in the next section, we do not attempt to empirically examine long search paths with patent data and instead rely on theory and the qualitative statements that arose from the inventor interviews. There have been previous studies that have considered the role of scientific reasoning in unusual discoveries (e.g., Fleming and Sorenson 2004) and distant recombination (e.g., Barirani et al. 2015, Keijl et al. 2016). Our paper, however, is the first we know of to articulate and explore the role of long search paths in outlier patents, to compare multiple search processes that may result in outlier discoveries, and to disentangle the roles of individual breadth of experience and team breadth of experience in distant recombination.

Serendipity

Occasionally, an unusual invention is the result of serendipitous discovery—or "luck." A prime example is the development of cyanoacrylate compounds into superglue. During World War II, Harry Coover was working in the Kodak Research Laboratories searching for an optically clear plastic for casting precision gunsight lenses. Coover first worked with cyanoacrylate compounds, but because they tended to stick to everything they touched, he abandoned them. Then in 1951, when one of Coover's colleagues was apologizing for ruining a $700 refractometer by inadvertently gluing two prisms together with cyanoacrylate, Coover suddenly realized the potential of the compound as an adhesive, noting,

> Serendipity had given me a second chance but this time the mental process led to inspiration. Immediately I asked Fred for a sample of his monomer and began gluing everything I could lay my hands on [...]. Everything stuck to everything, almost instantly, and with bonds I could not break apart. In that one afternoon, cyanoacrylate adhesives were conceived, purely as the result of serendipity. (Coover 1983, p. 59)

As noted previously, serendipitous discovery is the process that is closest to Kauffman's (1993, p. 212) original concept of a long jump being a random draw on dimensions of a fitness landscape.

Our interviews with inventors revealed instances of serendipity even in situations where other processes were also at play. For example, in describing the process that led to a method for detecting glucose using a surface plasma and resonance method (patent #6576430), one of the inventors, Bruce Pitner, noted,

> My colleague Helen Hsieh was using the Beacore device to characterize some protein when she noted something unusual. She said, 'This is really odd, it gives a negative signal.' We thought, 'That's not right, let's try this again.' Several experiments later we still found the same thing.

They did not initially pursue the idea but kept it on the back burner as a curiosity. Later, they gave the project to Jason Gestwicki, a student with a National Institutes of Health (NIH) fellowship that required an industry internship. Bruce felt that the student ought to work on something he could publish so that he would be rewarded from the internship. The negative signal result was perfect for this, and as Bruce noted, "It wasn't a path to a commercial project so it made sense to bring in a student." Bruce highlighted the role of serendipity in the discovery: "If it hadn't been a negative signal, we probably wouldn't have pursued it, but it was interesting because it goes against the convention of how surface plasma resonance is used to have a negative signal."

Surachai Supattapone (patent #6322802) similarly described an unusual finding in an experiment as the catalyst for pursuing the line of inquiry into how dendrimers might be used in both therapy and in the sterilization of prions:

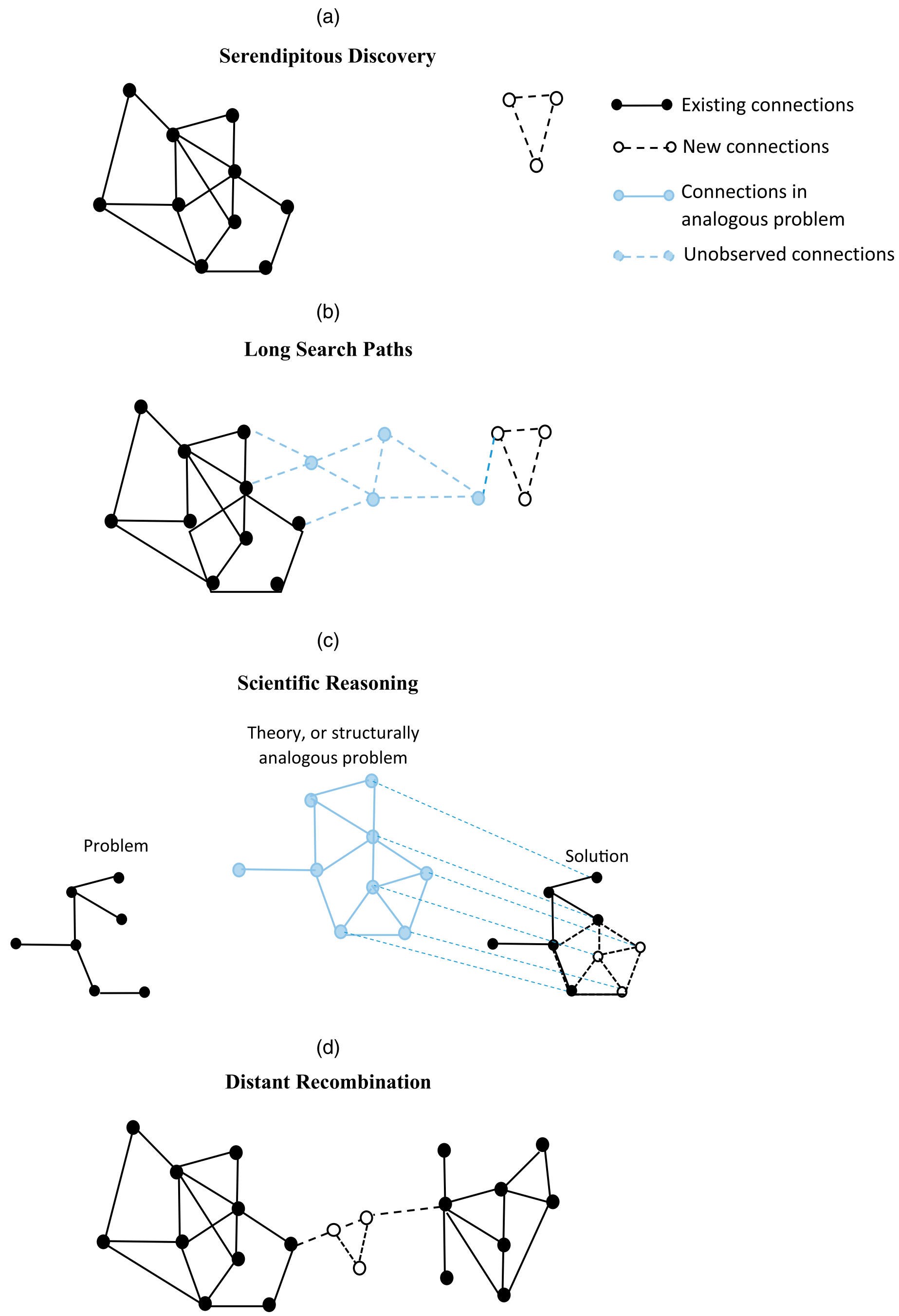
> I noted when we did the experiment that the amount of protein that we could see—it appeared that all of the samples that we transfected failed to convert into the prion form. Okay, we noticed that first. And, interestingly this was true even for the control experiment where we just used the normal form of the prion protein. We did not change it in anyway, we did not mutate it. So, that was weird, right? We expected that one to convert normally because it hadn't been changed.

However, although there is ample evidence that many discoveries have been, at least in part, unintentional, luck is rarely the entire answer. There has been a long historical debate on the degree to which any major discovery can truly be considered serendipitous (Merton and Barber 2004). First, a person may require training and expertise in an area to recognize a discovery (captured eloquently in the famous quote by Louis Pasteur in 1854, "chance only favours the mind which is prepared"), suggesting that a discovery is in part a result of a previous search. Second, particular discoveries may be ripe for the making—almost inevitable in fact—because the state of related science has brought us to their brink. This is the argument often made in studies of simultaneous discoveries (Merton and Barber 2004, Bikard 2017).

Long Search Paths

We propose here a second process that may lead to identification of a technology position that is not adjacent to any other known technology position: long search paths (see Figure 1, panel (b) and Table 3). Sometimes an outlier invention that appears to be a long jump is not really the result of a long jump from the perspective of the inventor (Adner and Levinthal 2008). Instead, it may be the result of a very *long or atypical search path* through a knowledge

Figure 1. (Color online) Knowledge Creation Processes for Outlying Technological Positions



network that would not occur to most other people (Schilling 2005). Rather than the discovery appearing as if out of thin air as in serendipity (Figure 1, panel (a)), it is many small steps on a technological trajectory that is understood by the inventor but unobserved by the outsider. The result appears to be disconnected from previous work because collectively it enables a very large leap forward, incorporating multiple improvements that set it well apart from its neighboring technologies (see Figure 1, panel (a)).

A remarkable example is illustrated in the development of BioSteel. In 1993, a McGill University biotechnology professor named Jeffrey Turner isolated the cells in mammary glands responsible for milk production. Based on his understanding of these cells, Turner quickly realized that dairy animals could be genetically manipulated to make different kinds of proteins. Turner left his tenured position to start a business using goats to produce better, cheaper enzymes for making lactose-free milk. Soon, however, he realized that there was a potentially higher-value

Published in Organization Science on April 01, 2020 as DOI: 10.1287/orsc.2019.1328.
This article has not been copyedited or formatted. The final version may differ from this version.

Table 3. Search Processes, Definitions, Indicators, and Quotes

Mechanism	Definition	Indicators from patent data	Representative quotes from interviews with inventors of outlier patents
Long search paths	A process of following a very *long or atypical search path* through a knowledge network. This process may be on a known technological trajectory but appear to be disconnected from previous work because it is a very large leap forward, incorporating multiple improvements.	• NA	"It was a long research path." (Soini 2017) "[There were] lots of small incremental steps along the way. You go from a piece of DNA all the way to something soluble, but that's not the end of it. All of the other companies would stop there. Here, you have the goat, but you're way way far from the final product." (Karatzas 2017)
Scientific reasoning	A process of reasoning that uses a generalized theory or set of assumptions from which a more specific hypothesis is derived.	• Scientific articles cited in patents	"I approached the problem solving from a fundamental chemical perspective without worrying about what was already known." (Liebeskind 2017)
		• University assignees	"Due to the biomimicry, you start with something from nature, and you try to do it with biotechnology. You try to build a bridge from learning with nature—like a mirror. With biomimicry you use models from nature and try to recreate it but improve upon it." (Karatzas 2017) "We found that and it made me start to think that, you know, under slightly acidic conditions—I'm not talking about acid like all the way like hydrochloric acid or something, it's just mildly acidic, under those conditions different thing like dendrimers could enhance the disinfection of prions." (Supattapone 2019)
Distant recombination	A process of combining two or more different trajectories—resulting in an unusual fusion of disparate knowledge domains.	• Collective team breadth	"I have a very different training than most people in the field, so I approached the problem with a distinct point of view." (Gestwicki 2018)
		• Maximum individual inventor breadth	"So, I have my scientific experience but I also have all my life stuff and the things I'm interested in. It's very unusual to find somebody who can put that together, and that's what I've always tried to do…" (Murto 2019)
		• Additional breadth of team over max individual	[About the research paths]: "It's not one of them; it's several of them at the same time…" (Soini 2017)

Note. NA, not applicable.

opportunity: spider silk proteins. Since at least the 1700s, it was well-known that spider silk has special mechanical properties. Spiders can produce several kinds of silk, and one of these—dragline silk—is one of the toughest materials known (Gosline et al. 1999). Spiders, however, are aggressive and very difficult to farm. DuPont, which claimed that a spider silk line the size of a pencil could stop a Boeing 747 in flight, had a project underway to produce spider silk proteins with genetically modified bacteria (Austen 2000). Its process, however, was slow and produced less-than-perfect clones of spider silk proteins. Turner believed that spider silk proteins could be made much more easily, and in greater quantity, by genetically modified goats. Turner and his team, which included fellow scientists Costas Karatzas and Anthoula Lazaris-Karatzis, began developing a strategy for modifying goats to produce spider silk protein. Turner licensed a spider silk cloning technology from Dr. Randy Lewis of the University of Wyoming, and the world's first "spider goats" were born in 1999.

The path to spider goats was long and circuitous, and the path to eventual applications was not immediately clear. As described by Costas Karatzas in our interview with him,

> Our target had always been the mammary gland of the animal. . . . What we had done at the University those years back was develop mammary cells, immortalized in vitro that would mimic lactation in a petri dish. We would then go from in vitro to a mouse, and then to a goat. The idea was to fail cheaply fairly quickly so that we could move on to the next stage of our evolution. We started by creating growth hormone and TPA but there were some technical complications. At that time Jeff Turner walks into my office and says "what if we tried to make something difficult? How about spider silk?" I said "Of course spider silk, that's interesting." That's part of being in a small young team—you never say "No." When God gives you lemons you don't just make lemonade; you go out and find 999 other things to do with lemons.

After the team succeeded in demonstrating that spider silk proteins could be made in milk, they had to develop bioreactors that could serve as factories of spider silk monomers. Then these monomers would be processed into stronger fibers that could be spun. Though each step of this path was novel and potentially patentable, the team was focused on progressing until they had a commercializable product:

> Natick [a U.S. Army Research and Development Center in Massachusetts] was already trying to produce fibers, and they helped us produce the spinning method. It was revolutionary. We could make it in gram quantities in these holofibers and then they could be spun, and these spun proteins gave us enough confidence to move to the next stages. That was the seminal work. That was about the time we went public; investors were very intrigued. . . . The next step was to go to mice to reduce the risk. 99% of the time we fail, so you need to reduce the risk. We used mice, targeting the specific elements in their milk (we never spun anything out of their milk because the quantity is too small).

Having shown that spider silk proteins could successfully be made in the milk of mice, the team next applied their technique to the breed-early-lactate-early (BELE) pygmy goats they had already created:

> We stumbled here because the more the mammary gland was making these proteins, the more their lactation was compromised. At high concentrations the silk has the ability to crystallize. The spider, she has all these things in her belly and when she pulls there is an automatic cord. It crystallizes upon sheer force. We were smart enough to have separated the two components of the dragline silk, but even these individual components had this property. They precipitated, which gave the goat a feedback mechanism to shut off the milk. If the lactation is compromised, you need more goats." [The team subsequently scaled up to a much larger herd of BELE goats.]

Their product, BioSteel, encountered both eager interest (notably from the U.S. military, which wanted spider silk bulletproof vests) and sharp criticism (from animal rights groups, among others). Turner's company, Nexia Biotechnologies, went bankrupt in 2009 before reaching commercial scale production, but Dr. Randy Lewis, who had supplied the initial spider protein cloning technology, continued the research on spider goats. In 2013, he received $1.15 million in grants from the U.S. Naval Research Office to develop spider silk production methods, followed by a $1 million grant in 2015 from the U.S. Army Research Office to scale up manufacturing. As Karatzas notes, "From saying 'yes' to Jeff to where we brought it with investment and research was quite a ride," adding,

> [There were] *lots of small incremental steps* along the way. You go from a piece of DNA all the way to something soluble, but that's not the end of it. All of the other companies would stop there. Here, you have the goat, but *you're way way far from the final product*. Because the goat is now your starting material. You need to take it and modify it to become a fiber, and that fiber has to become a product. . . . When you are on these innovative paths, you never know what's going to happen. . . . Many ideas die on the vine. You have all of these ideas that never see the light of day.

The creation of a spider silk–lactating goat at first sounds like fantastical fiction from a movie plot, but if one follows the long search path from the isolation of mammary cells, to genetically modified goats, to producing highly sought-after spider silk proteins from goat milk, the reasoning sequence does not appear at all random.

The use of long search paths was also amply evident in our interview with Juhani Soini, one of the inventors on patent #6361956, "Biospecific, two photon excitation, fluorescence detection and device." As Soini noted,

> *This is a long story* [emphasis added], it is part of my family history. My father has been working in the high-tech industry since the '60s. . . . At that time there was a breakthrough in the diagnostics industry using radio assay [a method of precisely measuring concentrations of substances like hormones or glucose in blood or on cell surfaces]. My father started developing a similar technology here [in Finland], for detecting minute amounts of radioactivity. . . . There was a desire to reduce the amount of radioactive material used in hospitals because there was too much gamma waste. Everyone wanted to find a way to make more sensitive immune assays with less material; to make it smaller, cheaper, and with less waste.

Juhani Soini's father, Erkki Soini, took a sabbatical year at the end of the 1980s in Heidelberg, where he was exposed to emerging technologies in cell counting and microscopy. He began working on a way to combine image processing, microscopy, and computerized cell counting. He also began collaborating with other scientists:

> There was heavy competition from other university groups. Russia was breaking down and research teams had been sent to Siberia where they had no resources, so we invited a Russian team to join us and we supported them. We also invited from Heidelberg a young researcher, Stefan Hell.... We were hunting for solutions.

The team finally landed on "functional flow cytometry," but in the course of this work, they also improved the microscopy significantly, and Stefan Hell was ultimately awarded a Nobel Prize for the development of superresolved fluorescent microscopy:

> We took the methodology and combined them into in-vitro diagnostics.... Some of the steps on the way weren't patentable.... We tried to protect only the immunoassays.... We have a long-term research tradition, from the '60s. *It was a long research path* [emphasis added]. My father, my brother, myself.... My father is so stubborn. The intuition is so strong, and the will to do is was just so strong. It's a way of thinking.

Soini's story not only illustrates a very long search path but also shows the possibility that inventors will choose not to patent some of the claims they encounter, or that claims will end up in other patents that do not appear proximate to the outlying patent.

When inventors are traveling down a long search path, it can result in a patent that appears disconnected from existing technologies because we do not observe the intermediate steps in the discovery process—that is, they were not patented. These intermediate steps may have been disclosed by other means, such as through academic publications, or not disclosed at all. The nondisclosure of these intermediate steps may be strategic (e.g., the inventor may believe that disclosing the intermediate steps would lead to earlier competition) or an inadvertent consequence (e.g., inventors in pursuit of a solution to a particular problem might not be interested in investing the time or effort in disclosing intermediate steps on their search path). If we could observe the knowledge network of the inventor rather than the technology landscape manifest in granted patents, we might realize an invention was, from the perspective of the inventor, not an outlier at all. Although occasionally it might be possible to document a long search path by, for example, tracing published articles or other archival evidence of the search process, often there may be no archival evidence of the path that is accessible. We thus do not attempt to empirically examine the long search path mechanism in our patent data, though developing a way to trace long search paths would be a valuable area for future research.

Scientific Reasoning

Inventors might use *scientific reasoning*—induction, deduction, or both—to predict that a fruitful combination exists, despite the fact that there has been no work proximate to that combination. A scientific theory or model may give us insight into solutions that ought to work, even if nothing proximate to that solution has ever been tried before, thus yielding solutions that are not adjacent to prior experiments. Einstein's revolutionary general theory of relativity, for example, made predictions about the travel of light that were not derivative extensions of prior work in the measurement of light; rather, they were consequences of his more general reasoning that light was actually tiny particles ("photons") and would be subject to gravity. This was quite a radical departure from the existing theory that light was a continuous electromagnetic wave. Einstein's hypotheses about light were subsequently verified by Sir Arthur Eddington during a complete solar eclipse, much to the amazement of the physics community. Fleming and Sorenson (2004) also provide an example of researchers predicting that single-wall carbon nanotubes would either conduct or semiconduct, depending on the angles of the carbon bonds. This helped inventors target a specific position on the technology landscape that was previously unknown, and that gave rise to a new generation of chip technology.

The preceding illustrates that scientific reasoning can point to where a solution ought to be, even if no one has yet explored that space. Theory can provide a structure of cause-and-effect relationships that helps us identify patterns in incomplete information (see Figure 1, panel (c) and Table 3). For example, by using arguments about when and why a relationship exists (or should exist) between two events or things, scientists can generate predictions about the results of untried experiments (Fleming and Sorenson 2004). For example, when it was discovered that Clustered Regularly Interspersed Short Palindromic Repeats – protein nuclease Cas9 (CRISPR-Cas9) was an adaptive immune system of bacteria that defended the single-celled prokaryote by modifying its genetic code, Feng Zhang (who had previously worked with other more complicated gene-editing tools) correctly deduced that CRISPR-Cas9 could be adapted for use in eukaryotes (including humans). Zhang set about proving his hypothesis and, in 2014, was granted the first patent for a CRISPR-Cas9 engineered for use in living mammalian tissues. This breakthrough technology offered the potential for correcting genetic disorders, even in adult humans. As alluded to, there were other genetic-editing tools in development, but they were expensive, difficult to produce, and posed major challenges for delivering them to the target cells. CRISPR-Cas9, on the other hand, was a simple, inexpensive, and elegant solution that was rapidly adopted by researchers around the world and named *Science Magazine's* 2015 Breakthrough of the Year.

The preceding does not imply that the use of scientific theory is more likely to generate outlier inventions than nonoutliers; outlier patents are uncommon, and the vast majority of patent applications, whether they are based on scientific reasoning or not, are for inventions that are not outliers (i.e., scientific reasoning does not always lead to a giant leap—it often leads to incremental advance). However, because scientific reasoning can sometimes enable an inventor to identify an invention opportunity that is very distant from existing occupied technology positions, we might expect to see it figure more prominently in the origination of outlying patents than nonoutlying patents (i.e., of the small number of patents that are outliers, we might expect a larger percentage of them to exhibit scientific reasoning relative to the percentage of nonoutlier patents that exhibit scientific reasoning).

Scientific reasoning can also be part of a long path search; it is therefore useful to note the conceptual difference between these two processes. Long path search means that the individual might have made no long jumps at all, but rather had many incremental steps that were unobserved by others. The outcome could look like a discovery that emerged out of thin air, even though it was actually the accumulation of many small steps. Some of these steps may have been based on scientific reasoning, some may have arisen through serendipity, some may be trial and error, etc. In some cases, scientific reasoning and long path search might overlap completely, as when the scientific reasoning process can be broken into many steps. However, there can also be long path searches that do not use scientific reasoning, and there can be scientific reasoning that leads to leaps of understanding that are not easily divisible into many small steps. Scientific reasoning can also be the impetus for distant recombination, and distant recombination can be an element of long path search. These processes are different, but they are not mutually exclusive; they can occur in tandem, and they can interact with each other.

It is relatively easy to identify the use of scientific reasoning in a search process if one talks to the inventor directly. For example, one of the far outlying patents, #6632805, "Methods for using water-stabilized organosilanes," is a method for giving objects an antimicrobial surface that is stable in water. This invention yielded a way of creating an antimicrobial surface on many kinds of products, including food containers; medical devices; latex gloves; heating, ventilation, and air conditioning (HVAC) systems; and more. The technological distance of this patent was 13 steps away from the nearest existing patent when it was filed. According to Lanny Liebeskind, a chemistry professor at Emory University and one of the two inventors on the patent, prior to making this discovery, he had been working with simple molecules that are antimicrobial—they disrupt cell membranes of microbes. Liebeskind noted,

> People asked me, 'Is there a way to stabilize it so you could use it in water? And make it more environmentally friendly?' So I took it on as a side project. It was outside of what I normally do in the lab. . . . *I approached the problem solving from a fundamental chemical perspective without worrying about what was already known.* [emphasis added] We conceived of a way to stabilize it in water—once attached to the surface it polymerizes. . . . We came up with this unusual solution to polymerize in water, and it worked. It was low tech enough that anybody out there could reproduce it, which is what happened. It wasn't worth prosecuting the patent cases.

Another inventor, Jonathan Stamler, described his use of scientific reasoning in the discovery of a method of coating medical devices with nitric oxide, thereby inhibiting infection and tissue damage (patent #6255277). As he noted,

> [Nitric oxide] was thought to have a lifetime in the body of less than one second. . . . The atoms of that molecule were very famous—they were not thought to be controllable. It was heresy that nitric oxide could last. . . . I argued that you can control it. . . . *Its chemistry was such that it should be amenable to binding to an amino acid called cysteine, and I demonstrated that to be the case. I deduced that with some chemistry it would happen—I turned out to be right*. [emphasis added] It was a magical molecule. I could attach it to any protein. If a protein was causing injury, I could put this on it and change its function. It worked. You can block re-stenosis with it.

Surachai Suppattapone (#6322802) discussed a similar process of reasoning that took his team from a biological understanding of the ways in which a cell interacted with a dendrimer to a set of testable hypotheses (which his team subsequently verified with experiments) about ways in which dendrimers might be useful in both therapy and sterilization processes to prevent prion infection:

> And what was interesting is that when you gave the dendrimers to the cells, the dendrimers were taken up into part of the cell called the lysosome, which is an acidic organelle. It's the part of the cell that contains acid, and it also contains prions and enzymes that can degrade proteins. So, we think what was happening is

Published in Organization Science on April 01, 2020 as DOI: 10.1287/orsc.2019.1328.
This article has not been copyedited or formatted. The final version may differ from this version.

> that the dendrimers get into the lysosomes and under the acidic pH that helps to unfold the infectious form of the prion, and maybe the dendrimers help that unfolding process and then maybe the enzymes that are there degrade it.

It might also be possible to identify use of scientific reasoning using archival evidence from patents. One rough proxy that we explore is the use of science in the form of citations to academic articles. Because academic science is a source of theories that inventors may use to deduce inventive solutions in uncharted spaces, we might see more references to academic articles cited in outlying patents than in nonoutliers. Academic articles are not, of course, the only source of theories, so this is a rough measure, but it is a place to start our exploration. The use of scientific theory in innovation also suggests that we might expect more university assignees on outlier patents. Drawing from Brooks (1994), we anticipate that universities may be disproportionately represented as assignees on outlier patents because of their focus on basic science. Whereas firms generate a significant portion of their revenue on commercial applications of existing technologies and derivative enhancements of existing products, university researchers are typically rewarded based on their scientific contribution and impact, and are far more likely to work on basic science and to explore newly emergent technological areas, or areas of unknown commercial opportunity (Jiang et al. 2010, Arora et al. 2015).

In the empirical evidence section of this paper, we explore the independent and joint effects of relying on scientific reasoning using academic articles and having university assignees. We find support for these arguments, with both citations to academic articles and university assignees independently predicting higher likelihood of outlier patent creation.

Distant Recombination

One of the ways that inventors might create inventions that appear distant from known inventions is through *distant recombination*—that is, a combination that fuses two or more disparate knowledge domains (see Figure 1, panel (d) and Table 3). Other studies have similarly posited that distant recombination might lead to inventions that are more basic or fundamental in their nature, or that have greater impact (e.g., Barirani et al. 2015, Keijl et al. 2016). An excellent example is the development of the camera pill. Gavriel Iddan was an electrooptical engineer developing the "eye" of a guided missile for the Israeli military when he befriended Eitan Scapa, a gastroenterologist (Iddan and Swain 2004). Scapa educated Iddan about the devices used to image the inside of the gastrointestinal system and their limitations. With Scapa's help and encouragement, Iddan developed a very small guided missile–like device that could travel through the intestinal system with a tiny camera eye that broadcast the images to a video pack worn by the patient. The unlikely fusion of gastroenterology knowledge and guided missile knowledge resulted in a breakthrough product called the PillCam, which received Food and Drug Administration (FDA) approval in 2001.

The role of distant recombination was evoked in several of our interviews. For example, Jason Gestwicki (one of the inventors on patent #6576430, a surface plasma and resonance method for detecting glucose) noted, "I have a very different training than most people in the field, so I approached the problem with a distinct point of view." Another inventor, Juhani, when asked about the research path that had led to the invention of the method for detecting microparticles in biospecific fluorometric assays, replied, "It's not one of them; it's several of them at the same time," and then explained that the invention had been due to the team's ability to merge research streams on image processing, fluorescent microscopy, and computerized cell counting.

James Murto (#6689615) explained how he and his coinventor, Mike Salvati, approached solving the problem of how to extract platelets from a blood sample without an expensive centrifuge:

> He and I, we didn't limit ourselves to the industry norm. We look at things from a little bit of a different mindset. I work with a lot of different scientists, and it's very unusual to find scientists that, um, will take things they've learned from working on their car or on their plumbing at home. . . . It's very unusual to go and find people who will go and look outside of their—I'm probably saying this terribly—their scientific experience. So, I have my scientific experience but I also have all my life stuff and the things I'm interested in. It's very unusual to find somebody who can put that together, and that's what I've always tried to do.

Murto and Salvati ultimately invented and patented a process through which coating superparamagnetic microparticles with a lectin (a type of protein that can bind to carbohydrates) allows a simple, inexpensive magnet to separate platelets from the other elements in a blood sample.

If an outlying patent is formed because of an unusual combination between knowledge domains, we might expect to see more diversity of expertise in the individual inventor or team of inventors that created it. As with a long search path, this could be strategic or inadvertent: an inventor or inventor team that intends to create an invention that fuses disparate bodies of knowledge is more likely to seek members with the diverse expertise needed (strategic choice), and an inventor or inventor team that happens to have diverse expertise is more likely to identify inventions that fuse disparate bodies of knowledge (inadvertent consequence).

Though there is a growing body of management research examining whether the combination of diverse knowledge domains leads to highly impactful innovations (e.g., high citation rates to articles or patents, or high economic value) (e.g., Ahuja and Lampert 2001, Fleming 2001, Rosenkopf and Nerkar 2001, Hargadon 2003, Singh and Fleming 2010, Schilling and Green 2011), there is far less work examining whether combining diverse knowledge domains results in innovations that are highly unusual in their form or function. An exception is Kaplan and Vakili (2015), who found (contrary to our expectation) that a broader search in nanotechnology does not directly result in innovations of greater novelty. Their context, however, is a single industry and utilizes very different measures, so we believe there is still much to be gained by examining this question.

There is a second, related issue here. If it turns out that breadth of experience matters, does it matter whether that breadth inheres at the individual or team level? Individuals might have particularly broad expertise because they had a varied career history that brought them into contact with multiple technological domains. Alternatively, they might have worked in an area of basic science that could be applied to multiple technological domains, or worked in a field that sits at the intersection of multiple domains, either of which could have made them a "technology broker" (Hargadon 2003). However, patents are also created by teams of inventors who integrate their knowledge through a transactive memory system (Wegner 1987, Moreland and Argote 2003) that does not require each inventor to possess the knowledge of the other. Individuals could thus each have relatively narrow prior technological experience, yet a team composed of multiple specialists could collectively exhibit quite diverse experience if the individuals worked in different technological domains.

At the heart of this contextually specific question is a more general question: To what degree is invention a process driven by individual ideation versus integration across a group? Many studies have shown that individuals working alone generate more varied or novel outcomes than those in a group (Dahlin et al. 2004, Fleming 2007, Stroebe et al. 2010). Groups can inhibit creative ideation due to fear of judgment, norms of convergence, and production blocking (i.e., when one person is speaking, that person prevents others from contributing) (Stroebe et al. 2010). This implies that teams diminish both number and variance in ideas. On the other hand, Singh and Fleming (2010) argue that teams might influence variance in ideas asymmetrically by selecting out the really bad novel ideas without diminishing the really good novel ideas. They found that, compared with lone inventors, teams of inventors generated more extremely good patent outcomes and fewer extremely bad ones, though it is important to note that their dependent variable is impact, not novelty.

Melero and Palomeras (2015) took a different approach and studied the effect of having a generalist (an individual with particularly broad experience) on an inventing team. They found that generalist inventors are very valuable (i.e., create economically valuable patents that are jointly filed in the European, Japanese, and U.S. patent offices) when inventing in areas of high uncertainty. Generalists may have a better "big picture" of the technological landscape that helps them make better choices (consistent with Fleming and Sorenson 2001 and Gruber et al. 2013) and play a bridging role that helps integrate the knowledge of specialists (similar to work by Rulke and Galaskiewicz 2000).

We do not expect there to be one answer that applies to all teams; every inventive team has a different structure, composition, and process. It should be immediately apparent that the type of diversity of experience, power structure of the team, length of time working together, domain of the invention, and more can all influence the type of outcomes obtained. In the following section, our objective, then, is merely to explore whether individual-level breadth of experience and group-level breadth of experience each has a significant independent effect. We find suggestive empirical evidence that both processes play a role in predicting outlier creation, although the individual-level breadth has a comparatively larger effect.

Empirical Evidence

In this section, we aim to identify the presence of the focal search processes and their role in generating outlying innovation. We empirically explore how two of the search processes, *scientific reasoning* and *distant recombination*, are used in the creation of outlier patents. We do not attempt to examine empirically the long search path mechanism in our patent data because there may be no reliable or consistent archival evidence of the path. From the population of utility patents applied for (and subsequently granted) between 1990 and 2000 (1,535,878 patents), we use the previously described method to identify 120,957 outlier patents (7.88%), which represent distant technological invention. We first examine scientific reasoning and distant recombination as independent processes before exploring their combined effects and relative magnitudes.

Scientific Reasoning

To explore the possibility that scientific reasoning may facilitate the discovery of outlier patents, we first

Table 4. Descriptive Statistics and Correlations

Variables	Mean	SD	Min	Max	1	2	3	4	5	6	7	8
1. *Outlier*	0.08	0.27	0	1								
2. *Scientific articles cited*	0.48	1.12	0	9.66	−0.00							
3. *Patents cited*	3.04	1.20	0	10.19	0.06	0.03						
4. *University*	0.02	0.14	0	1	0.01	0.26	−0.05					
5. *Collective breadth*	3.11	2.96	1	37	0.01	0.02	0.04	−0.03				
6. *Max individual breadth*	2.90	2.68	1	37	0.01	0.02	0.04	−0.02	0.98			
7. *Difference in breadth*	0.21	0.66	0	14	0.01	0.01	0.03	−0.02	0.52	0.33		
8. *Number of inventors*	2.29	1.65	1	41	−0.01	0.14	−0.01	0.03	0.32	0.24	0.44	
9. *Prior patents (10 years)*	3.34	1.69	1	10.71	−0.04	0.15	0.03	−0.01	0.70	0.67	0.40	0.56

Notes. $N = 1{,}535{,}878$. SD, standard deviation.

examined the relationship between the likelihood of a patent being an outlier and citations to science. Following Fleming and Sorenson (2004), who studied the use of scientific deduction, we measure *scientific articles cited* as a simple count of academic articles referenced by the focal patent. We add one to this value and take the log (base 2) due to the heavily skewed distribution of values, ranging from 0 to 807 with the mean close to two before transformation.[7] This count is a sum of the unique scientific articles identified by a linking algorithm developed by Marx and Fuegi (2019), through which items listed in the "other references" section of the front page of each patent were matched to a specific article within proprietary databases, such as *Web of Science*, as well as open source databases.[8]

We also examined whether outlier patents are more likely when the assignee is an academic institution. *University* is a binary variable that takes the value of 1 for patents assigned to academic institutions. The coding of assignee types was developed by The Patent Name-Matching Project (Hall 2008), which classifies the first assignee into one of six mutually exclusive categories.[9] The intuition behind this variable is that the use of scientific theory and/or a greater interest in basic science in academia may motivate exploration of unusual technological coordinates.

We also included several controls (see Table 4 for summary statistics and correlations). First, we use data from the disambiguated inventor database (Li et al. 2014) to match the inventors on our sample patents to all of their prior patents. *Prior patents* is a logged count of granted patents of the inventor(s) over the previous 10 years at the time of the focal patent's application date to capture patenting experience at the team level.[10] We also calculated a sum of *number of inventors*, a count of the number of the inventors on the patent, ranging from 1 to 41. This variable is designed to ensure the prior patenting history of the team is not driven by the size of the team.

We further include a control for the field of the invention. *Field* is a set of dummy variables for the 38 industry classifications created by NBER (Hall et al. 2001) in which the 400 patent classes were sorted into 38 categories. The classification used was from the data set updated in 2012.[11] We include these dummy variables for field to account for patenting differences across technological areas. *Patents cited* is the logged (base 2) count of the number of patents listed as prior art upon which the patent builds. We also include application year dummy variables to capture time effects.

Because outlier is a binary dependent variable, we estimate the coefficients using the following logit model:

$$P(y_i = 1|X_i) = \frac{\exp(X'_i\beta)}{1 + \exp(X'_i\beta)},$$

where $y_i = 1$ is the likelihood that patent i is a technological outlier, X_i is a vector of variables at the patent level, and β is the vector of coefficients of interest. Table 5 shows the coefficient estimates and standard errors for logit models predicting the likelihood of an outlier. Model 1 includes only the controls, including year, field, size of the team, and the collective number of approved patents filed in the previous 10 years, as these are likely to affect search processes. Model 2 includes scientific articles cited and patents cited. We separated out the *scientific articles cited* and the *patents cited* to see how the referencing of prior art independently influences the likelihood of being an outlier patent. We find that the number of scientific articles cited is positively associated with outlier patents ($z = 14.30$, $p < 0.001$), and exponentiating the coefficient indicates that, for each doubling of articles cited, a patent's odds of being an outlier increases by about 1.05 times—a modest but significant effect.[12] The count of patents cited is also positively associated with outlier patents ($z = 65.20$, $p < 0.001$): each doubling of the number of patents cited increases the odds of a patent being an outlier by about 1.20 times. Overall, then, we find that controlling for field of invention, year of application, team size, and prior patenting experience, outlier patents cite both more articles and more patents.

We include *university* in Model 3 (Table 5) and find that universities are significantly more likely to be

Published in Organization Science on April 01, 2020 as DOI: 10.1287/orsc.2019.1328.
This article has not been copyedited or formatted. The final version may differ from this version.

Table 5. Logit Regressions Predicting Outlier Patents with Use of Scientific Reasoning

	Models			
Variables	1	2	3	4
Articles cited		0.046** [0.003]		0.038** [0.003]
Patents cited		0.179** [0.003]		0.181** [0.003]
University			0.311** [0.021]	0.270** [0.022]
Number of inventors	0.042** [0.002]	0.043** [0.002]	0.041** [0.002]	0.043** [0.002]
Prior patents (10 years)	−0.093** [0.002]	−0.107** [0.002]	−0.092** [0.002]	−0.105** [0.002]
Constant	−2.020** [0.048]	−2.436** [0.048]	−2.030** [0.048]	−2.447** [0.048]
Field dummies (37)	Yes	Yes	Yes	Yes
Application year dummies (10)	Yes	Yes	Yes	Yes

Notes. Logit coefficients shown with standard errors in brackets. N = 1,529,701.
$^{**}p < 0.001$.

the assignees on outlier patents (z = 14.75, $p < 0.001$). The odds of a university filing an outlier patent are 1.36 (EXP(0.311)) times greater than a nonuniversity assignee. Our results are consistent when including both *scientific articles cited* and *university* assignee in Model 4 (Table 5). These results are also robust to using a different data indicator of university assignees, with a more inclusive range of what counts as a university patent.[13]

Overall, then, the data suggest that one of the ways inventors may pioneer new areas of the technology landscape is through application of scientific reasoning, and future research should explore how and when this strategy will be most valuable.

Distant Recombination

We begin exploring distant recombination by creating several breadth measures based on the number of unique fields in which the inventor(s) has successfully patented within the 10 years prior to and inclusive of the year of application of the focal patent (see Table 4 for descriptive statistics). We created this measure using the disambiguated inventor database (Li et al. 2014) and NBER's 38 industry classifications (Hall et al. 2001) as our fields. We first identified the patent portfolio of each individual inventor in the USPTO database that filed any patent between 1981 and 2000. For each inventor-patent pair, we then identified each unique field represented in the patent portfolio for the 10 years leading up to and including the focal patent. Each field is counted only once; thus, our breadth measures can theoretically range from 1 to 38. These individual patent portfolios were constructed in the same spirit as the Fleming et al. (2007) measure of focal inventor experience to capture the "breadth of experience that the inventor brings to the creative effort" (Fleming et al. 2007, p. 454). Our measure, however, is based on NBER's technology categories, which are significantly broader than the subclasses used in the Fleming et al. paper. This approach is more consistent with the concept of distant recombination representing a fusion of disparate technology domains.

We next calculate a measure of the collective breadth of the inventor team (*breadth*) by aggregating the individual portfolios up to the inventor team level, treating each of the 38 fields as a dummy variable to retain the identity of unique fields. We counted the total number of unique fields represented by the inventor team for each focal patent. In our data, we find that the maximum number of subcategories an inventing team has a history of patenting in is 37 (mean = 3.11). Our team-level measure of experience departs from the collaborative breadth measure of Fleming et al. (2007) in that they consider prior experience of all past collaborators, not just those on the focal patent. We believe our measure is appropriate for our purposes of exploring the cognitive search processes of a team.

To examine the importance of breadth at the individual level, we constructed the variable *max individual breadth*, which is a count of the technology categories in which the broadest inventor on the team has patented in the previous 10 years. For patents filed by individual inventors, this value is equal to the previously described breadth measure; however, for a team it represents the broadest patenting experience of any single individual on the team. In our data, we find that the maximum individual breadth is also 37 (mean = 2.90). To test if team breadth explains

Table 6. Logit Regressions Predicting Outlier Patents with Distant Recombination and Full Model

	Models				
	1	2	3	4	Marginal effects[a]
Articles cited				0.045** [0.003]	0.045
Patents cited				0.188** [0.003]	0.227
University				0.243** [0.022]	0.243
Collective breadth	0.098** [0.001]				
Max individual breadth		0.105** [0.002]	0.103** [0.002]	0.109** [0.002]	0.303
Difference in team/individual breadth			0.066** [0.005]	0.064** [0.005]	0.030
Number inventors	0.050** [0.002]	0.066** [0.002]	0.056** [0.002]	0.059** [0.002]	0.091
Prior patents (10 years)	−0.228** [0.003]	−0.228** [0.003]	−0.231** [0.003]	−0.251** [0.003]	−0.333
Constant	−1.858** [0.048]	−1.908** [0.048]	−1.873** [0.048]	−2.320** [0.048]	
Field dummies (37)	Yes	Yes	Yes	Yes	
Application year dummies (10)	Yes	Yes	Yes	Yes	

Notes. Logit coefficients shown with standard errors in brackets. N = 1,529,701.
[a]Marginal effects are calculated after Model 8. Each marginal effect represents a one-standard-deviation increase in the likelihood of a patent being an outlier. The exception is for the independent variable for *university*, which is a binary indicator for which the marginal effect is the increase from zero to one.
**$p < 0.001$.

additional variance above and beyond the broadest inventor on the team, we created *difference in breadth*, measuring the difference between the breadth of the broadest individual and the collective team breadth. This variable captures the additional, unique fields of experience on the team not captured by the individual with the greatest breadth of experience. In our data, the maximum value for difference in breadth is 14 (mean = 0.21).

To explore the roles of inventor and inventor team breadth in the knowledge search process, we add our breadth variables to our controls in Models 1–3 in Table 6. First, we look at collective team-level breadth, *collective breadth*, measured by the sum of unique NBER subcategories the inventors on the team have previously patented in during the prior 10 years. The result is significant and positive (z = 68.30, $p <$ 0.001), and the marginal effects indicate that at the mean breadth, an additional 10 fields of experience more than double the probability of being an outlier, from 0.067 to 0.162 (Model 1, Table 6). This suggests that an individual or team with diverse prior experience is more likely to generate an unusual combination of technologies. Notably, our breadth measure is not being driven by the number of inventors on the team, as we have controlled for this separately. It is also interesting to note that although breadth could be an indirect measure of prior patenting experience (those who have filed more patents in the past have more opportunities to file in different fields, thus accumulating both breadth and more experience generally), we find that prior patenting experience is consistently significant and negative across all models. This suggests that the inventor(s) of outlier patents has significantly *less* patenting experience than those generating patents in closer proximity to preexisting technological inventions. This is an interesting finding and lends support to arguments made by Schilling (2005) that experience can be somewhat double-edged: an individual with extensive experience in an area is more likely to become trapped in known paradigms and may be less likely to generate breakthrough ideas. Repeated experience at a task can cause a phenomenon known as "einstellung," or functional fixedness, whereby learners form a problem-solving set that mechanizes their problem solving, constraining them from developing creative solutions (Luchins 1942, Mayer 1995).

Next, we disaggregate the breadth measure to disentangle the roles of individual breadth and team breadth.

Model 2 (Table 6) looks at the maximum individual breadth of experience, measured by the greatest number of subcategories any one inventor on the team had previously patented in over the prior 10 years, and we find evidence of a significant effect of the *maximum individual breadth* ($z = 67.96$, $p < 0.001$). Marginal effects indicate that, at the mean of *maximum individual breadth*, a standard deviation for the variable (i.e., roughly three fields of experience) increases the probability of an outlying patent from 0.067 to 0.087 (a 30% increase), and an additional 10 fields (which is well within the range of the *maximum individual breadth*) increases the probability to 0.171 (a 155% increase)—a large amount given the generally low likelihood of filing an outlier patent.

To investigate whether additional breadth of experience offered by the team matters, we add the difference between team and maximum individual breadth into Model 3. We find that although individual breadth is still highly significant, additional units of breadth above and beyond the broadest inventor are a significant predictor of creating technologically outlying patents ($z = 12.95$, $p < 0.001$), with the marginal effects of the additional team breadth being large (but smaller than those for maximum individual breadth); that is, an additional 10 fields above the mean increase the probability of being an outlier to 0.122, an increase of 82%. A more modest increase of 3 additional fields yields an increase of 21%. These results suggest strong support for our arguments about distant recombination: inventors may make discoveries in uncharted technological space by making unusual fusions between disparate technology domains. Furthermore, our findings suggest that both individual-level breadth and team-level breadth contribute to the process of generating novel combinations.

As noted previously, these processes are not mutually exclusive, and many instances of knowledge creation will combine elements of multiple processes. Some of the steps in a long search path, for example, might be applications of scientific reasoning, distant recombination, or serendipity. Furthermore, scientific reasoning may enable distant recombination. We thus did not intend to identify inventions as the outcome of a single process, nor are we posing a "horse race" about which process yields the most radical inventions. Rather, these processes likely co-occur within the inventing process. Model 4 (Table 6) includes all variables associated with the *scientific reasoning* process and with the *distant recombination* variables, along with their respective marginal effects for comparison. The combined model yields results consistent with previous estimates, further substantiating the independence of these two processes while not ruling out the simultaneous or sequential use of these processes. Our empirical results are robust to using a higher threshold of four technological steps away (instead of two) to define outliers (top 13.48% of most distant outliers) as well as to using a continuous measure of technological distance as the dependent variable (Appendix C). Our sole purpose is to explore processes that are more likely to be observed in outlying inventions than nonoutlying inventions. Answering that question is a prerequisite for future research that may have more normative implications for inventors and organizations.

Discussion and Conclusions

Most technological invention builds closely on existing inventions; this is highly efficient because the knowledge and resources in a known technological domain tend to spillover, at least in part, to adjacent domains. That is, what is known about an existing technology can be leveraged in pursuit of other closely related technologies, lowering the cost of a search and improving the likelihood of its success. Existing technological inventions also provide signals of the value of potential opportunities in the technological domain and thus serve as guideposts that guide future search. However, sometimes inventors make discoveries that are (or at least appear to be) quite remote from existing inventions, and sometimes these discoveries turn out to be important, path-breaking innovations. We were thus interested in the following question: How do inventors explore these uncharted technological regions?

We built upon work on knowledge networks to identify search processes that are more likely to yield discoveries that are distant from existing inventions. These processes include using *long search paths* that cumulate a series of unobserved steps on a technological trajectory, use of *scientific reasoning*, and use of *distant recombination* that fuses disparate knowledge domains. We hoped to be able to identify systematic differences in the search processes leading to outlier inventions using a unique—and very large—data set on outlier patents. Consistent with the idea that use of scientific reasoning might facilitate discoveries in uncharted technological space, we found that outlier patents are significantly more likely to cite scientific articles, and that outlier patents are more likely to be associated with university assignees. The effect of a university assignee was particularly large: accounting for the other search processes in our full model, the odds of a university assignee patent being an outlier were 24% higher than the odds for a nonuniversity assignee, which is a large and meaningful difference.

We also found evidence supporting the idea that unusual inventions might be the result of distant recombination: inventors or inventor teams who had patented in a greater breadth of fields in the 10 years leading up to the focal patent were much more likely to produce outlier patents, even when controlling for number of inventors on the inventing team and their degree of prior patenting experience. These results differ from those of Kaplan and Vakili (2015), who found a negative relationship between breadth and novelty using their text analysis approach. We do not, however, believe that these findings are necessarily at odds with one another. Their measure of breadth was about the content of the patent based on the subclasses of the patents cited by the focal patent (and it is important to note that not all prior art citations come from the inventing team—roughly 40% of prior art citations are added by patent examiners as reported by Alcacer et al. 2008). Our breadth measure is based on a broader construct of technology domains (vs. their use of subclasses), and our measure captures the prior experience of the inventing team, which may or may not be reflected in the prior art citations of the focal patent. Kaplan and Vakili's (2015) work also focused on a single technology domain, which may have significantly influenced the likelihood of observing fusion of technology domains.

One of the most important findings in our data was that the maximum breadth of any single individual on the team had a particularly strong influence on the probability of an outlier patent, highlighting the fact that synthesis of diverse knowledge might be facilitated by having an individual on a team whose experience is broad—the renaissance woman or man (Melero and Palomeras 2015). This has extremely important implications for managing inventor teams, which we will discuss at greater length in the "Contributions and Implications for Managers" section.

This paper also pioneers the use of a novel measure of outlier patents. Aharonson and Schilling (2015) motivated and conceptualized the computation of geodesic distances between binary vectors of mainline subclasses; however, this is the first application of the method for identifying outlying innovation. Our measure of an outlier is based on the underlying technological components of the innovation at the time of application rather than either the subsequent forward citations of the patent (an outcome measure of success) or the prior art included in the patent (backward citations as a proxy for knowledge drawn upon). Research that has used prior art citation measures of novelty (such as an unusual combination of backward citations) suffers from issues of sequential interdependence (i.e., you can only cite what has come before you) and biases from patent examiners adding cites to patents (Alcacer and Gittelman 2006). Using the patent subclass system ensures "historical consistency" as USPTO updates how the patents are classified based on the current understanding of the technological field (Fleming 2001). Our measure improves our ability to conceptualize the extent of novelty in a spatial manner (as opposed to "familiarity" based on the frequency of combination of subclasses) and enables us to introduce relative concepts between patents, such as adjacency and neighboring technology. We believe this measure will prove useful for many future studies, as it is not constrained by domain, yields discrete identification of outliers and a continuous measure of technological distance, and is readily updated to account for technology evolution.

Finally, by drawing on a cognitive network model of search, we developed a simple and visual way of thinking about how inventors can arrive at hitherto unknown areas of the technological landscape. Although there has been considerable work assessing the importance of exploratory versus exploitative search, there is relatively little work (either conceptual or empirical) that examines how inventors achieve exploratory search. We hope that this framework will prove useful for many future studies.

We were conservative in our expectations of being able to find evidence for the knowledge search processes described here; it is difficult to directly measure the knowledge search processes that lead to a patent, the processes described here are not mutually exclusive (i.e., multiple processes may be at work simultaneously), and some outlier inventions are the result of serendipitous discovery or may arise through processes we have not yet identified. Furthermore, inventors may draw from work that they do not cite, and they may have experience that is not evident in patents. All of these factors made it less likely that our models would yield significant evidence for the processes we identified. However, despite these obstacles, we obtained preliminary evidence that (a) the search processes that lead to outlier inventions are often different from those that lead to nonoutlying inventions, and (b) we can identify individual search processes useful for exploring uncharted technological domains.

Future Research Agenda

Our objective here was to generate theory about how inventors explore uncharted technology space and to subject our ideas to some preliminary empirical investigation. Our findings are encouraging and suggest that these arguments warrant more in-depth assessment. There are a number of ways that future research could broaden our knowledge of existing and potential knowledge search processes, deepen our knowledge about how and why such search processes are used, and give us a better understanding of when they are likely to be useful.

First, it might be useful to try to tease out the temporal effects within a given firm or within a given

Published in Organization Science on April 01, 2020 as DOI: 10.1287/orsc.2019.1328.
This article has not been copyedited or formatted. The final version may differ from this version.

inventor's inventive career. It would be interesting, for example, to see whether a firm that switches between processes reaps concomitant differences in patent outcomes, and what the benefits or costs of those outcomes are. There may also be interesting relationships between the use of scientific reasoning and distant recombination. For example, Gruber et al. (2013) found that inventors with a scientific education are more likely to generate patents that span technological boundaries, and Barirani et al. (2015) found that the use of basic science moderated the relationship between distant recombination and "basicness" of invention in Canadian nanotechnology. Second, it might be possible to gain a richer perspective of the search processes described here by using a multiple case study approach with a matched case design where outlying patents and nonoutlying patents from the same field and time period are compared in more depth.

Third, an experimental design could also be useful. For example, to examine the long search path argument, an experimental setting could prime the use of search paths of different lengths by giving teams significantly different time periods to generate a solution to a problem. The resulting solutions could be compared for their novelty. Semantic analysis of recordings of conversations held within the team could also be used to attempt to trace the search path, so that path lengths could be compared across different teams. The role of team diversity and its decomposition between the individual and collective diversity could also be assessed in a laboratory setting. In a laboratory setting, we could examine not only how the structure of diversity influences outcomes but also how that influence plays out. Do generalists dominate the ideation in a team, for example, or do they primarily help the team form a transactive memory system that helps them to synthesize their collective ideas? When is having a generalist valuable, and does having some degree of expertise overlap among team members obviate the value of having a generalist? There are many interesting questions to be asked and answered here.

Contributions and Implications for Managers

Our results also provide direction for managers who wish to encourage exploration of uncharted technological terrain. First, bringing together people with more diverse experience on the inventing team may increase the likelihood of teams pursuing and patenting breakthrough technology. Notably, breadth of disciplinary experience may not manifest in direct application on the focal patent in the form of a citation to prior art, as studied by Kaplan and Vakili (2015). Rather, we provide suggestive evidence that breadth of experience shapes the thinking, exploration, and identification of potential recombination that results in a technologically distant patent. We show that breadth operates at both the team level, which lends itself to implications for how teams should be staffed, and the individual level, which may speak to hiring decisions of whom to bring into the firm. Our data suggest that having a generalist with broad experience on the team is particularly valuable for generating outlier inventions, offering one of the most straightforward ways managers may be able to improve the likelihood of successful exploratory search.

Second, several of our interviews highlighted the crucial role of bringing on a younger member to the team, sometimes a graduate student, who was given the latitude to take an anomalous finding and probe for its potential value or application. Although we were unable to test for the age range within the team, anecdotally we find support that a range of age on the team allowed for outside-of-box thinking—either in the identification of a technologically distant idea or in the willingness to pursue it. For managers, it would be prudent to think of diversity not only in technological knowledge but also in the complementarity of mental elasticity and expertise that can reside within a team.

Third, managers should create opportunities and incentives for teams to make greater use of academic research in their innovative efforts, as well as deeper engagement with the scientific community, which is less driven by the potential commercial value of its endeavors. The developments in the basic sciences may allow for technological breakthroughs in distant areas of the landscape in a way that relying on prior patents does not. Although we have highlighted the role that academic research plays in enabling outlier patent creation, several of the inventors mentioned being influenced by seemingly unrelated sources, including how the natural world solves related problems. Managers ought to welcome nontraditional sources of inspiration into problem-solving conversations. These novel sources of information from distant domains may contribute to creating outlier patents in relation to preexisting technologies and may also contribute to helping organizations distinguish themselves in relation to their competition. If the majority of organizations are relying on similar sources of information from which to create recombinant solutions, embracing diversity of team composition and of sources of inspiration may provide a competitive advantage in creating breakthrough innovations.

Finally, our results highlight that managers should give teams the freedom to follow unusual paths of inquiry when they stumble upon an atypical finding. Several of the inventors we interviewed noted that an unexpected or counterintuitive finding sparked the search that ultimately yielded an outlier patent, and the value of being able to pursue an unusual path, as articulated by inventor James Murto: "The other thing that helps is if you work for a company that doesn't dictate how you develop a product. If they leave you to solve problems without totally narrowing it down, it opens up your opportunities and the way you look at things."

Conclusion

The purpose of this exploratory paper is to develop a better understanding of how inventors purposefully explore uncharted technological space, where greater opportunities of untried technological combinations might yield important breakthroughs. We extend research in this domain by taking a cognitive perspective, and by doing so, we bridge the existing literatures on innovation, landscapes, and cognition. We theorize about three specific search processes (long search paths, scientific reasoning, and distant recombination) and present qualitative and quantitative evidence that suggests these processes facilitate innovation in areas of the technological landscape which are distant from what is well understood. This research presents a novel framework on which future research can build to further our collective understanding of the important processes that underlie exploration.

Acknowledgments

The authors gratefully acknowledge the help and support of Sam Arts, J. P. Eggers, Lee Fleming, Deepak Hegde; the participants of the 2016 Organization Science Winter Conference, the 2017 Winter Strategy Conference, the 2017 Strategic Management Society Conference, and the 2017 Academy of Management Conference; and participants at seminars at New York University, IE Madrid, and Drexel University. The authors also thank Felipe Csaszar and three anonymous reviewers for their thoughtful comments and guidance throughout the review process.

Appendix A. Measuring Technology Distance and Outliers

For each patent, we create a vector of binary indicators of the 9,864 mainline subclasses. This vector represents a patent's technology position (there are a possible $2^{9,864}$ potential technological positions). The vast majority of patents are filed in technology positions in which there are already other patents (i.e., their vector of mainline subclasses is the same as those of other patents). A smaller number of patents are filed in positions that have no other patents, but are adjacent to occupied positions (i.e., their vector has one difference from the nearest occupied technology positions). We use these adjacencies to create a network of technology positions that enables us to measure the technological distance from every position (and the patents in it) to every other position. Outliers have no technological adjacencies at the time of application (i.e., their vectors have at least two differences from every other existing patent); they are isolated technology positions. For example, consider the binary vectors for the following four patents (only the first eight mainline subclasses are shown; the remaining would have "0s"):

a. 0 0 0 1 1 1 0 0 . . .
b. 0 0 0 1 1 1 0 0 . . .
c. 0 0 0 0 1 1 0 0 . . .
d. 0 0 0 1 1 0 0 0 . . .

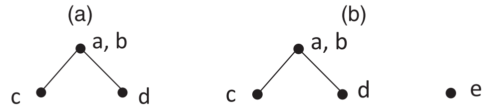
Patents a and b are in the same technology position. Patents b and c are adjacent, and Patents b and d are adjacent. Patents c and d are at a network distance of two from each other. In network form, they can be represented like panel (a) of Figure A.1. If this were the universe of patents, and a new patent was filed with the following vector,

e. 0 0 1 0 1 0 0 0,

it would be an outlier that was at a distance of two positions from the nearest occupied position. In network form, it can be represented like panel (b) of Figure A.1.

Our concept of an outlier patent is temporally conditioned—that is, whether something is an outlier when it is filed depends upon what has been filed before. A patent that is an outlier when filed in 1982, for example, may not have been an outlier if filed in 1994, as illustrated by the biotechnology "bloom" graphically depicted in the Aharonson and Schilling (2015) paper. Therefore, the outlier measure is calculated on a yearly basis, from 1976 forward (these data were obtained from the authors of the Aharonson and Schilling (2015) paper).

Figure A.1. Example of Network Diagram with New Unconnected Node



Appendix B

Table B.1. Additional Interview Information

Name of interviewed inventor (assignee in parentheses)	Outlier patent number	Number of technological steps at time of application	Patent description	Length of interview
Constantinos Karatzas (Nexia Biotechnologies, Inc.)	5780009	2[a]	Direct gene transfer into the ruminant mammary gland	95 minutes (2 interviews)
Surachai Supattapone (The Regents of the University of California)	6322802	11	Method of sterilizing	24 minutes
Jeremy W. Steitz (Moore North America, Inc.)	6422938	11	Pressure seal C-Z fold	16 minutes

Table B.1. (Continued)

Name of interviewed inventor (assignee in parentheses)	Outlier patent number	Number of technological steps at time of application	Patent description	Length of interview
James Murto (individually owned patent)	6689615	11	Methods and devices for processing blood samples	19 minutes
Donald Elbert (Universitat Zurich)	6884628	12	Multifunctional polymeric surface coatings in analytic and sensor devices	32 minutes
Jonathan Stamler (Brigham and Women's Hospital; NitroMed, Inc.)	6255277	13	Localized use of nitric oxide adducts to prevent internal tissue damage	50 minutes
Juhani Soini (individually owned patent)	6361956	13	Biospecific, two photon excitation, fluorescence detection and device	65 minutes
J. Bruce Pitner (Becton, Dickinson)	6576430	13	Detection of ligands by refractive surface methods	65 minutes
Jason E. Gestwicki (Becton, Dickinson)	6576430	13	Detection of ligands by refractive surface methods	20 minutes
Lanny S. Liebeskind (Emory University)	6632805	13	Methods for using water-stabilized organosilanes	75 minutes

Notes. In total, we attempted to track down 93 inventors from 30 patents. We were able to find contact information for 20 inventors on 16 patents, and we conducted interviews with 10 of those inventors (representing 9 patents), for an overall response rate of 50%.

[a]Though patent #5780009 did not appear on our list of most distant patents, the spider goat story had been important in motivating our study, and we had extensive conversations with one of the inventors. We thus include them here.

Appendix C. Robustness Analyses

Model 1 uses a higher threshold for defining outliers (four or greater technological steps away at the time of application, as compared with two steps in the main analyses). This cutoff represents the top 13.48% of the most distant outlier patents. Not only are the results consistent with the main analyses presented in Table 6 of the paper, but marginal effects are actually stronger for more distant outliers.

Table C.1. Robustness Analyses

	Outlier		Distance	
	1	Marginal effects	2	Marginal effects
Articles cited	0.057** [0.008]	0.066	0.012** [0.003]	0.005
Patents cited	0.277** [0.007]	0.390	0.057** [0.003]	0.026
University	0.354** [0.049]	0.354	0.070* [0.021]	0.070
Collective breadth				
Max individual breadth	0.121** [0.004]	0.381	0.015** [0.002]	0.016
Difference in team/individual breadth	0.091** [0.013]	0.062	0.011* [0.005]	0.003
Number of inventors	0.068** [0.006]	0.119	0.010** [0.002]	0.006
Prior patents (10 years)	−0.288** [0.008]	−0.385	−0.030** [0.003]	−0.019
Constant	−4.536** [0.125]		2.493** [0.047]	
Field dummies (37)	Yes		Yes	
Application year dummies (10)	Yes		Yes	
N	1,529,701		120,480	

Notes. Logit coefficients shown with standard errors in brackets. Marginal effects are calculated after the preceding model. Each marginal effect represents a one-standard-deviation increase in the likelihood of a patent being an outlier. The exception is for the independent variable for university, which is a binary indicator for which the marginal effect is the increase from zero to one.

$^{**}p < 0.001$; $^{*}p < 0.05$.

For more distant outliers, scientific reasoning (article citations and university) and distant recombination are stronger predictors on the likelihood of being an outlier.

Model 2 is an ordinary least square regression predicting the number of technological steps away at the time of application (between 2 and 18) *within* the outlier patent population (N = 120,480). Within this group of outlier patents, these processes predict comparatively smaller effects in how distant the patent is from preexisting technology; however, the direction of the effects is entirely consistent with the theoretical arguments in this paper.

Endnotes

[1] The idea of a technology landscape is derived from Sewall Wright's (1932) concept of a fitness landscape within which evolution occurs. In Wright's conception, each attribute of an organism relates to a dimension of the space, and a final dimension of the space represents the organism's fitness—that is, likelihood of survival and reproduction (Levinthal 1997).

[2] We chose the 1990–2000 time frame so that we could gather 10 years of previous patenting expertise for each of the inventors and complete forward citation data for 10 years after the patents are granted (the data set from which we draw patent class and forward citation data extends from 1981 to 2014).

[3] See https://eml.berkeley.edu/~bhhall/patents.html, accessed March 16, 2020.

[4] The distribution of forward citations is highly skewed; however, this difference is consistent when comparing the median of forward citations between outliers (10) and nonoutliers (8).

[5] We did not interview inventors with the intention of conducting a rigorous qualitative analysis—that would have necessitated finding a much larger set of representative inventors, which is beyond the scope of the current study. Our interviews were merely intended to give us a richer sense of these inventors' outlier discovery process.

[6] Note, we refer here to networks of cognitive associations in the mind, not social networks. For simplicity, we will describe knowledge creation only at the individual level, but these arguments are straightforward to generalize to the transactive memory systems of groups.

[7] Using a base 2 log transformation makes interpretation of the coefficients in a logit analysis more straightforward than using a natural log transformation.

[8] Our results are robust to using a simple count of the material in the other references section, following the approach of Fleming and Sorenson (2004). Notably, with increased precision of the measure from the Patent Citations to Science project (Marx 2019, Marx and Fuegi 2019), we find stronger results with larger marginal effect sizes.

[9] Though this measure is based on only the first assignee, we are reasonably confident this does not introduce bias into the results, as only just over 1% of utility patents are filed with multiple assignees. We are indebted to Juan Alcacer for this statistic.

[10] As expected, the correlation between our breadth measures and prior patents is high, however it serves as an important control of past experience. We ensure that in all of our models the independent variables of interest (collective breadth, max individual breadth, and difference in breadth) have variance inflation factor (VIF) scores below 5.

[11] See https://sites.google.com/site/patentdataproject/Home, accessed March 5, 2020.

[12] Our results are robust to using a dummy indicator for having *any* references to scientific articles, consistent with the approach used by Fleming and Sorenson (2004).

[13] Notably, we are using both citations to academic research and university assignees as proxies for the scientific reasoning. Although the correlation between these two measures (0.26, $p < 0.001$) is not high enough to consider them multiple measures of the same construct, patents with university assignees were significantly more likely to cite at least one article (0.75 versus 0.20, $p < 0.001$) and cited significantly more articles (13.37 versus 1.46, $p < 0.001$). Thus, although we believe it is valuable to include both of these measures of scientific reasoning, the relationship between them slightly dampens the results we obtain for each. The variables are tested separately in Model 2 and Model 3 and tested together in Model 4.

References

Adner R, Levinthal DA (2008) Doing vs. seeing: Acts of exploitation and perceptions of exploration. *Strategic Entrepreneurship J.* 2(1): 43–52.

Aharonson BS, Baum JAC, Feldman MP (2007) Desperately seeking spillovers? Increasing returns, industrial organization and the location of new entrants in geographic and technological space. *Indust. Corporate Change* 16(1):89–130.

Aharonson B, Schilling MA (2015) Mapping the technological landscape: Measuring technology distance, technological footprints, and technology evolution. *Res. Policy* 45(1):81–96.

Ahuja G, Katila R (2004) Where do resources come from? The role of idiosyncratic situations. *Strategic Management J.* 25(8–9):887–907.

Ahuja G, Lampert CM (2001) Entrepreneurship in the large corporation: A longitudinal study of how established firms create breakthrough inventions. *Strategic Management J.* 22(6–7): 521–543.

Alcacer J, Gittelman M (2006) Patent citations as a measure of knowledge flows: The influence of examiner citations. *Rev. Econom. Statist.* 88(4):774–779.

Alcacer J, Gittelman M, Sampat B (2008) Applicant and examiner citations in U.S. Patents: An overview and analysis. Working paper, Harvard Business School, Boston.

Arora A, Belenzon S, Patacconi A (2015) Killing the golden goose? The decline of science in corporate R&D. NBER Working Paper No. w20902, National Bureau of Economic Research, Cambridge, MA.

Austen I (2000) Silk and money. *Canadian Bus.* 73(12):42–45.

Barirani A, Beaudry C, Agard B (2015) Distant recombination and the creation of basic inventions: An analysis of the diffusion of public and private sector nanotechnology patents in Canada. *Technovation* 36–37(February–March):39–52.

Bartlett FC (1932) *Remembering: An Experimental and Social Study* (Cambridge University Press, Cambridge, UK).

Bikard M 2017. Creativity "in the air": Simultaneous discoveries in science and their implications. Working paper, London Business School, London.

Brooks H (1994) The relationship between science and technology. *Res. Policy* 23(5):477–486.

Brown S, Eisenhardt K (1998) *Competing on the Edge: Strategy as Structured Chaos* (HBS Press, Boston).

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

Coover HW (1983) Cyanoacrylate adhesives—A day of serendipity, a decade of hard work. *J. Coatings Tech.* 55:59–61.

Csaszar FA (2012) Organizational structure as a determinant of performance: Evidence from mutual funds. *Strategic Management J.* 33(6):611–632.

Csaszar FA, Eggers JP (2013) Organizational decision making: An information aggregation view. *Management Sci.* 59(10):2257–2277.

Csaszar FA, Levinthal D (2016) Mental representations and the discovery of new strategies. *Strategic Management J.* 37(10):2031–2049.

Csaszar FA, Siggelkow N (2010) How much to copy? Determinants of effective imitation breadth. *Organ. Sci.* 21(3):593–801.

Cyert RM, March JG (1963) *A Behavioral Theory of the Firm*, 2nd ed. (Prentice Hall, Englewood Cliffs, NJ).

Dahlin K, Taylor M, Fichman M (2004) Today's Edisons or weekend hobbyists: Technical merit and success of inventions by independent inventors. *Res. Policy* 33(8):1167–1183.

Dosi G (1988) Sources, procedures, and microeconomic effects of innovation. *J. Econom. Literature* 26(3):1120–1171.

Ellis HC (1965) *The Transfer of Learning* (Macmillan, Oxford, UK).

Fahlman SE (1989) Representing implicit knowledge. Hinton GE, Anderson JA, eds. *Parallel Models of Associative Memory* (Lawrence Erlbaum Associates, Hillsdale, NJ), 171–190.

Fang C, Lee J, Schilling MA (2010) Balancing exploration and exploitation through structural design: The isolation of subgroups and organizational learning. *Organ. Sci.* 21(3):621–642.

Fleming L (2001) Recombinant uncertainty in technological search. *Management Sci.* 47(1):117–132.

Fleming L (2007) Breakthroughs and the "long tail" of innovation. *MIT Sloan Management Rev.* 49(1):69–74.

Fleming L, Sorenson O (2001) Technology as a complex adaptive system: Evidence from patent data. *Res. Policy* 30(7):1019–1039.

Fleming L, Sorenson O (2004) Science as a map in technological search. *Strategic Management J.* 25(8–9):909–928.

Fleming L, Mingo S, Chen D (2007) Collaborative brokerage, generative creativity, and creative success. *Admin. Sci. Quart.* 52(3): 443–475.

Gavetti G, Levinthal D (2000) Looking forward and looking backward: Cognitive and experiential search. *Management Sci.* 45(1): 934–950.

Gestwicki JE (2018) Information on training provided via personal communication with authors, September 22 and 23.

Gosline JM, Guerette PA, Ortlepp CS, Savage KN (1999) The mechanical design of spider silks: From fibroin sequence to mechanical function. *J. Experiment. Biol.* 202(part 23):3295–3303.

Gruber M, Harhoff D, Hoisl K (2013) Knowledge recombination across technological boundaries: Scientists vs. engineers. *Management Sci.* 59(4):837–851.

Hall BH (2008) The patent-name matching project. Accessed October 29, 2019, https://eml.berkeley.edu/~bhhall/patents.html.

Hall BH, Jaffe AB, Trajtenberg M (2001) The NBER patent citation data file: Lessons, insights and methodological tools. NBER Working Paper No. w8498, National Bureau of Economic Research, Cambridge, MA.

Hargadon A (2003) *How Breakthroughs Happen: Technology Brokering and the Pursuit of Innovation* (Harvard Business School Publishing, Boston).

Helfat CE (1994) Evolutionary trajectories in petroleum firm R&D. *Management Sci.* 40(12):1720–1747.

Iddan GJ, Swain CP (2004) History and development of capsule endoscopy. *Gastrointestinal Endoscopy Clinics North Amer.* 14(1):1–9.

Jiang L, Tan J, Thursby M (2010) Incumbent firm invention in emerging fields: Evidence from the semiconductor industry. *Strategic Management J.* 32(1):55–75.

Kaplan S, Vakili K (2015) The double-edged sword of recombination in breakthrough innovation. *Strategic Management J.* 36(10): 1435–1457.

Katila R, Ahuja G (2002) Something old, something new: A longitudinal study of search behavior and new product introduction. *Acad. Management J.* 45(6):1183–1194.

Kauffman S (1993) *The Origins of Order* (Oxford University Press, New York).

Kauffman S, Lobo J, Macready WG (2000) Optimal search on a technology landscape. *J. Econom. Behav. Organ.* 43(2):141–166.

Keijl S, Gilsing VA, Knoben J, Duysters G (2016) The two faces of inventions: The relationship between recombination and impact in pharmaceutical biotechnology. *Res. Policy* 45(5): 1061–1074.

Koestler A (1973) *The Act of Creation* (Oxford University Press, London).

Levinthal D (1997) Adaptation on rugged landscapes. *Management Sci.* 43(7):934–950.

Li GC, Lai R, D'Amour A, Doolin DM, Sun Y, Torvik VI, Amy ZY, Fleming L (2014) Disambiguation and co-authorship networks of the U.S. patent inventor database (1975–2010). *Res. Policy* 43(6):941–955.

Liebskind LS (2017) Information on discovery path provided via personal communication with authors, November 2.

Luchins AS (1942) Mechanization in problem solving: The effect of einstellung. *Psych. Monographs* 54(6):i–95.

Karatzas C (2017) Detailed information on discovery path provided via personal communication with authors, November 16.

March JG (1991) Exploration and exploitation in organizational learning. *Organ. Sci.* 2(1):71–87.

March JG, Simon HA (1958) *Organizations* (Wiley, New York).

Martindale C (1995) Creativity and connectionism. Smith SM, Ward TB, Finke RA, eds. *The Creative Cognition Approach* (MIT Press, Cambridge, MA), 249–268.

Marx M (2019) Patent citations to science. Inter-university Consortium for Political and Social Research, Ann Arbor, MI.

Marx M, Fuegi A (2019) Reliance on science in patenting. Working paper, Boston University Questrom School of Business, Boston.

Mayer RE (1995) The search for insight: Grappling with Gestalt psychology's unanswered questions. Sternberg RJ, Davidson JE, eds. *The Nature of Insight* (MIT Press, Cambridge, MA).

Mayer RE, Greeno JG (1972) Structural differences between outcomes produced by different instructional methods. *J. Ed. Psych.* 63(2):165–173.

Melero E, Palomeras N (2015) The Renaissance Man is not dead! The role of generalists in teams of inventors. *Res. Policy* 44(1):154–167.

Merriam-Webster. Outlier. Accessed October 29, 2019, https://www.merriam-webster.com/dictionary/outlier.

Merton RK, Barber E (2004) *The Travels and Adventures of Serendipity* (Princeton University Press, Princeton, NJ).

Moreland RL, Argote L (2003) Transactive memory in dynamic organizations. Peterson RS, Mannix EA, eds. *Leading and Managing People in the Dynamic Organization* (Lawrence Erlbaum Associates Publishers, Mahwah, NJ), 135–162.

Murto J (2019) Information on invention process provided via personal communication with authors, April 2.

Nelson RR, Winter SG (1982) The Schumpeterian tradeoff revisited. *Amer. Econom. Rev.* 72(1):114–132.

Nickerson JA, Zenger TR (2002) Being efficiently fickle: A dynamic theory of organizational choice. *Organ. Sci.* 13(5):547–566.

Nonaka I (1994) A dynamic theory of organizational knowledge creation. *Organ. Sci.* 5(1):14–37.

O'Reilly CA, Tushman ML (2016) *Lead and Disrupt: How to Solve the Innovators Dilemma* (Stanford University Press, Stanford, CA).

Pitner B (2017) Information on discovery process provided via personal communication with authors, September 29.

Rivkin JW (2000) Imitation of complex strategies. *Management Sci.* 46(6):745–873.

Rivkin JW, Siggelkow N (2003) Balancing search and stability: Interdependencies among elements of organizational design. *Management Sci.* 49(3):290–311.

Rosenkopf L, Almeida P (2003) Overcoming local search through alliances and mobility. *Management Sci.* 49(6):751–766.

Rosenkopf L, Nerkar A (2001) Beyond local search: Boundary-spanning exploration, and impact in the optical disk industry. *Strategic Management J.* 22(4):287–306.

Rothaermel FT, Deeds DL (2004) Exploration and exploitation alliances in biotechnology: A system of new product development. *Strategic Management J.* 25(3):201–221.

Rulke D, Galaskiewicz J (2000) Distribution of knowledge, group network structure, and group performance. *Management Sci.* 46(5):612–625.

Schilling MA (2005) A "small-world" network model of cognitive insight. *Creativity Res. J.* 17(2–3):131–154.

Schilling MA, Fang C (2016) When hubs forget, lie, and play favorites: Interpersonal network structure, information distortion, and organizational learning. *Strategic Management J.* 35(7):974–994.

Schilling MA, Green E (2011) Recombinant search and breakthrough idea generation: An analysis of high impact papers in the social sciences. *Res. Policy* 40(10):1321–1331.

Siggelkow N, Levinthal DA (2003) Temporarily divide to conquer: Centralized, decentralized and reintegrated organizational approaches to exploration and adaptation. *Organ. Sci.* 14(6):650–669.

Siggelkow N, Rivkin JW (2006) When exploration backfires: Unintended consequences of multilevel organizational search. *Acad. Management J.* 49(4):779–795.

Soini J (2017) History of research agenda provided via personal communication with authors, September 21.

Simonton DK (1999) Creativity as blind variation and selective retention: Is the creative process Darwinian? *Psych. Inquiry* 10(4):309–328.

Singh J, Fleming L (2010) Lone inventors as sources of breakthroughs: Myth or reality? *Management Sci.* 56(1):41–56.

Stamler JB (2018) Information on discovery path provided via personal communication with authors, October 4.

Steyvers M, Tenenbaum JB (2002) The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. Working paper, Stanford University, Stanford, CA.

Stroebe W, Nijstad BA, Rietzschel EF. 2010. Beyond productivity loss in brainstorming groups: the evolution of a question. *Adv. Experiment. Soc. Psych.* 43:157–203.

Supattapone S (2019) Information on experimentation process provided via personal communication with authors, April 2.

Wegner DM (1987) Transactive memory: A contemporary analysis of the group mind. Mullen B, Goethals GR, eds. *Theories of Group Behavior* (Springer, New York), 185–208.

Wright S (1932) The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proc. VI Internat. Congress Genetics*, 356–366.

Madeline K. Kneeland is an assistant professor at Cornell University's SC Johnson College of Business. She received her PhD from New York University's Stern School of Business. Her research focuses on the dynamics and evolution of networks, collaboration, and innovation.

Melissa A. Schilling is the John Herzog Family Professor of Management at New York University's Stern School of Business. She received her PhD from the University of Washington, Seattle. Her research interests focus on creativity and innovation, including platform dynamics, technology trajectories, collaboration, and serial breakthrough innovators.

Barak S. Aharonson is a professor of strategic management and entrepreneurship at Xiamen School of Management, Xiamen University, China, and a member of the faculty of management at the Coller School of Business, Tel Aviv University, Israel. His research pertains to strategy and organizational theory focusing on entrepreneurial activity; knowledge flows within and across geographic agglomerations and networks; and evolution, creation, and diffusion of technology and innovations.