HEY, MAYBE YOU CAN HELP ME WITH THIS:

CHANCE ENCOUNTERS,

GEOGRAPHIC PROXIMITY,

AND INNOVATIVE COLLABORATION

A DISSERTATION

SUBMITTED TO THE FACULTY OF THE

UNIVERSITY OF MINNESOTA

BY

KEITH ELIOT PENNINGTON

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

ADVISOR: J. MYLES SHAVER

JUNE 2021

ACKNOWLEDGEMENTS

The guidance of others shaped my dissertation in significant ways, and I want to give appropriate thanks to everyone who helped along the way. And, my ability to progress through my Ph.D. program has been supported, even carried at times, by many, many more people in multiple different ways, and they are no less worthy of acknowledgment. But, I feel I have too many people to thank, and I need to thank them too intensely, that I find it humbling to attempt to write an honest acknowledgments section for fear of leaving people out or diminishing how important they have been to me.

However, some specific thanks are easy to prioritize because of their sheer magnitude. First and foremost, I thank my advisor Myles Shaver. He initially encouraged me to work on what I thought was an exciting but challenging research topic and set high standards – especially in terms of clear thinking about accomplishing research – that I aspire to meet for the rest of my career. I also thank the members of my dissertation committee, Mary Benner, Evan Rawley, and David Knoke, who helped and shaped my dissertation – and research style – immensely.

Classes by Mary, Myles, Andy Van de Ven, Shaker Zahra, Harry Sapienza, Joel Waldfogel, Aseem Kaul, and Janna Johnson shaped my early experience in the program and formed my view of what makes good research. In addition, working with Sunasir Dutta and Alex Wilson gave me an understanding of how fun research can be! I had what I think of as a healthy skepticism of how I might fit into this profession until about halfway when I decided this is a great profession to be in, and I could start seeing myself thriving as a strategy Ph.D. I'm happy I stuck with it, and I would not have without those

mentioned here and my fellow students at the time. I also want to thank everyone not yet already mentioned who gave me guidance and helpful suggestions with my research, teaching, job search, and career advice, both inside the department and outside. I hope to one day replicate the same amount of kindness, patience, and thoughtfulness in the help that you gave me.

I also want to thank all of my friends throughout the last six years, precisely because sometimes you gave me the most critical support when I needed it most. But there's too much to mention, and if I list all events I consider essential, at some point, readers will begin wondering if I did any work myself. I trust you know I appreciate you if you're reading this. Those beers at Republic (& Insight, &…), walks to Izzy's, Sea Salt trips, runs by the river, impromptu coffee breaks, hallway chats, desserts at Cossetta, lunches designed to escape campus, all were a core part of my experience here, and I know I'm lucky to count you as friends. Many of you are like family now, and you know who you are. (Some are even more officially like family than others – Jeeny's the godmother to my daughter Meena who was born during the program, and I get the feeling I'll get in trouble if I don't mention her here.)

And, of course, I could have done none of this without the unwavering support from my wife, Vishaka. I know that me pursuing a Ph.D. has been a considerable and sometimes challenging experience for you as well, and one you weren't expecting to need to put up with. Thank you for indulging me.

While Meena was cute, she was not any help at all.

For Jackie, why not

ABSTRACT

We know that proximity can influence collaborations, but it is unclear how proximity affects collaborator choice. Collaborator choice is important because it determines what knowledge is recombined, shaping the very nature of an innovation. In my dissertation, I consider both the context and mechanisms of proximity to explore the effect of proximity on collaborator choice. My first essay explores the context of proximity between knowledge workers' residences, a setting that excludes some potential proximity mechanisms and modestly generalizes to between-firm spillovers. I show residential proximity predicts collaboration, as prior literature shows within a workplace. Notably, I show that only same-gender pairs seem influenced by residential proximity. This finding is potentially counterintuitive because, here, proximity leads to less diverse collaborator choices. I suggest homophily prohibits between-gender interactions outside an office from being substantive enough to increase awareness and influence collaborator choice. In the rest of my dissertation, I focus on the proximity mechanism of chance encounters as theoretically distinct from proximity. I predict and show that chance encounters should generally increase the variance of collaborator choice and diversity of knowledge in innovation. I also show that chance encounters do not seem to increase productivity. Instead, when chance encounters are less available, people turn to their prior contacts, limiting recombination potential. To empirically invoke variance in chance encounters instead of proximity, I develop and validate a measure using the flu to proxy for fewer chance encounters, borrowing from epidemiology findings that show when people fear a disease, they can engage in social avoidance behavior with proximate people.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1, INTRODUCTION: A FRAMEWORK FOR THE STUDY OF PROXIMITY AND COLLABORATIONS

MOTIVATION

The simplest version of how I communicate my research is that when physically proximate people decide to collaborate, I ask, who participates? Why? And what are the outcomes? I study the interrelated questions of when and where chance interactions happen between proximate people, with whom do they happen, and how collaborations and innovations that start *because of* proximity differ from collaborations that start from other ways.

Small interactions between people often begin with similar words to the first line of the title of this dissertation: "Hey," someone asks, seeing a colleague at the watercooler, "maybe you can help me with this," referring to a technical problem they are facing. The colleague is asked, in part, because they are physically proximate to the person with a technical problem: they see each other at their shared watercooler. As the relationship grows, more knowledge is shared, and ultimately, they collaborate and produce an innovative new technology.

Like I illustrate at the watercooler, collaborations can happen by chance, involving no planning from the individuals involved. But collaborations only happen this way sometimes because only some encounters grow, and only some encounters happen at all. We don't notice when people don't have such successful serendipitous meetings, so we don't know what we miss. Worse yet, we don't even know if we should be missing something.

We have plenty of anecdotal evidence that chance encounters can spur innovation, so popular opinion is that more chance encounters should be good for innovation. We may have missed a lot as a society, for example, if not for a meeting at a coffee shop in Berkeley a few years ago. The recent Nobel prize winner Jennifer Doudna and her collaborator describe in some detail how "it was serendipitous" how the early development of the gene-editing technology CRISPR all began, ultimately setting Doudna on an important new path of science and technology. A few quotes tell the story succinctly:

> *"Thinking back on it, it really began … when I moved my lab from Yale to Berkeley, I had a conversation with Jillian Banfield, who told me about evidence for a bacterial immune system called CRISPR" – Jennifer Doudna* (Public Affairs UC Berkeley 2020)

> *"Every now and again, Jennifer and I met (usually at the Free Speech Café) and talked about collaborating." – Jillian Banfield* (Public Affairs UC Berkeley 2020)

Douda moved her lab, and while at Berkeley, she talked to Banfield, another researcher in a different area. "It was serendipitous" how they met, Banfield stated (Public Affairs UC Berkeley 2020). This serendipitous meeting set Doudna in a new direction that ultimately changed science. Her work enables scientists to edit genomes; that's what the name CRISPR now means to people who know the medical technology. In terms of recent scientific development, CRISPR is a massively big deal; it is not surprising to many in the field that Doudna received a Nobel prize for her contribution.

There are other high-profile examples of proximity and chance in the development of important innovations. A conversation over burgers with Nobel laureate Paul Lauterbur led to the development of the MRI (Spice 2003). Katalin Kariko, the pioneer of mRNA technology used widely in COVID-19 vaccines, used chance encounters with the people around her to develop and support her work (Crow 2021). The world would be in a different place without these technologies.

It's easy for people and the popular press to think that this topic is a settled fact: we understand that proximity increases chance encounters between people, leading to innovation. However, we do not fully understand the effects of proximity, and it is not clear that proximity always leads to the positive outcomes some may expect. Proximity can induce social processes and knowledge processes that are difficult to empirically study. From what we know from recent studies, proximity plays a complicated role in collaborations by acting through different mechanisms, and these mechanisms can have very different potential outcomes on innovation (Catalini 2018). It is not clear that proximity should always encourage innovation despite what popular business press – and leaders of firms – seem to believe. However, incomplete information does not stop firms and managers from making policy decisions in attempts to be more innovative, likely motivated by examples like I gave above. That organizations find the idea of increasing proximity-induced collaborations to increase strategic innovation introduces my first broad motivation for my dissertation.

Practical motivation: Firms try to manipulate chance encounters

There is a trend of designing workspaces to encourage chance-type social interactions between employees, the idea being employees will work with each other more and ideas will spread more freely (Hess 2017). Apple's Apple Park in Cupertino exemplifies this trend. Steve Jobs' last project with Apple was to help design the company's new headquarters, and Jobs was known for trying to make serendipitous events happen, so the whole facility is designed to make people have chance meetings. At Pixar, Jobs even seated employees with the longest possible walk to the bathroom to have more opportunities to interact with other employees (D'Onfro 2015). (Presumably, these interactions mostly happened on the way back.) Jobs believed in the power of chance encounters, and he's not alone.

From this example, you can see that such "unplanned" encounters are "unplanned" by the employees. Chance encounters may be influenced by planning at other levels, like the CEO Jobs at Apple, but the events themselves are unplanned by the people engaging in them. A cottage industry of innovation consultants supports firms trying to make similar decisions to Jobs' efforts at Apple Park. This trend currently has little hard and generalizable scientific basis, but such changes may still be effective for many firms. However, in some cases, such efforts may be counterproductive. Better understanding the mechanisms and social processes of proximity will be key to developing better strategies for organizations who want to use proximity to induce innovation.

Practical motivation: Remote work

Working from home (or working from "anywhere") is another recent trend that began after I started my dissertation but is even more motivating than the previous trend. The COVID-19 pandemic changed everyone's lives in many ways, but one of the most salient changes for many knowledge workers is that they began working from home if the nature of their job allowed it. Working from home is a trend that seems to be sticking as we begin pulling out the pandemic, with many leading innovative technology companies switching to working from home long term, like Microsoft (Gilbert 2020). Even Apple CEO Tim Cook says some elements of remote work will persist well into the future, in a quote from the Atlantic Festival Ideas Stage: "I don't believe we will return to the way we were, because we found that there are some things that actually work really well virtually" (2020). In this interview, Cook also notes that he expects innovation to be more of a struggle remotely, but his statement that elements of remote work will persist seems to stand in stark contrast to the efforts of Apple's prior CEO Steve Jobs to encourage chance, in-person encounters.

In some regions, governments are creating policies to make working from home more of a protected right (e.g., in Germany; Meyer 2020), or even requirement (e.g., in the Bay Area; Ingram 2020) into the future. Many knowledge workers may not go back to the office as it was. Without understanding the processes that happen because of proximity, we don't know what effect removing proximity might have. How do we know what we might be losing? Plus, are there settings, job roles, and strategic goals of firms that could make remote work more of a good idea or a bad idea? These timely and

pressing questions further motivate my dissertation. Though I cannot answer this question completely, I provide evidence that I believe can generalize to and inform some of the work from home decisions organizations and policymakers are making right now.

Broader motivations

Other than the narrower immediate and practical motivations listed earlier, better knowledge of how proximity affects collaboration would help shed light on various streams of strategy, innovation, economic geography, and public policy research concerned with physical proximity between people.

Spillover and regional agglomerations of knowledge literature examine knowledge transfer between proximate organizations. One mechanism described – but seldom evaluated to the exclusion of others – is that chance conversations between employees of different but proximate firms can lead to knowledge spillovers between those firms, which could be competitors. In this field, the common citations provided for evidence of chance conversations leading to spillovers are usually limited to ethnographies without quantitative approaches (Jacobs 1969, Saxenian 1996). Though these books convincingly establish that geographic proximity can have meaningful effects on knowledge dynamics, the qualitative sources do not help us understand the extent of the effects of proximity or the conditions that may enable them. More recent knowledge spillover studies often also cite within-organization and within-facility studies like Catalini (2018) as evidence that chance conversations can lead to spillovers. Catalini's study is quantitative and benefits from a quasi-experimental research design but describes an entirely different theoretical setting and context of proximity than one

where a spillover may occur between firms. Agglomeration literature would benefit from a clearer mapping from the context of proximity, affected mechanisms, and outcomes to highlight where gaps in knowledge exist. Agglomeration literature would also benefit by quantitatively describing the extent and nature of outside-facility chance encounters and mechanisms of proximity because that is the context usually appropriate for studies of between-firm spillovers. I focus on this conceptual mapping and potential outside-facility mechanisms of proximity in chapter 2 of this dissertation.

Organizational structure literature is also concerned with the natural processes arising in informal communication patterns, with an aim to showing how firm structure can influence those patterns. Organizational structure literature is the theoretical side of the first practical motivation I list, where firms attempt manipulations of these natural processes. Firms interested in innovation should also be interested in the control of technology development to carefully match technology developments to core capabilities to be able to exploit their own innovations. Without this control, unconstrained development efforts can lead to potentially harmful projects that do not fit with firm capabilities (Henderson and Cockburn 1994, Tushman and Anderson 1986). The natural collaborations that arise from proximity are less controlled than directed projects with oversight and assigned team roles, and organizations manipulating proximity should be aware of the potential negative effects of this lack of constraint. My research has implications for this research stream as well.

Various other fields like public policy, urban planning, public health, and architecture all have streams of literature around understanding how to manipulate

proximity and the effects of the proximity of people. A common motivation in these streams of literature is to obtain some benefit through maximizing chance interactions (though notably, the epidemiology subfield of public health generally has the opposite goal when considering the spread of infectious diseases through the close contact of people). The study of proximity and chance encounters is far from being contained within the collaboration and innovation space in literature. *Who is near you* and *how you engage with them* determines a great deal of your behavior of interest to many social science fields. Deeper knowledge of the processes proximity allows, and with what effects, has broad implications in many areas.

In the following section, I introduce what we know about the role of proximity in collaboration and innovation within the current strategy, innovation, and collaboration-related literature. I then describe how I build a framework to assess this knowledge systematically. I then show where the chapters of my dissertation fit and complement the growing body of knowledge around proximity and innovation. In this way, I expose the more detailed theoretical motivations for my research to complement the practical and broad motivations listed above.

Introduction to the role of proximity in collaboration and innovation

Creative work is collaborative, and who collaborates can determine the nature of that creative work. I find it helpful to take a recombination perspective of innovation, considering that people have their knowledge "components" as in Fleming (2001). By collaborating with others, they can recombine their existing knowledge components with the knowledge of others to make new innovative recombinations. Therefore, who people

work with and use knowledge from is a centrally important question to innovation production because different people represent different knowledge.

Literature shows that physical proximity shapes the process of recombination by influencing collaborations. Namely, people are more likely to work with those proximate to them for various reasons, and often proximity effectively decides the type of knowledge a person might recombine. Proximity also leads to knowledge spillovers and relationship formation and strengthening through mechanisms related to the collaboration process. Usually, in literature, the three mechanisms are considered as 1) awareness of people and their knowledge increase with proximity, 2) convenience of collaboration increases with proximity, and 3) social relationships tend to increase with proximity. These mechanisms are difficult to untangle from each other empirically and theoretically (a theoretical example of this entanglement is that awareness increases with a social relationship, so the two concepts are related). Because of both empirical and theoretical tangles, we have limited knowledge of the effect of proximity on collaborations at the mechanism level.

While developing my dissertation, I highlight the most influential and recent papers in the study of proximity to get an overview of what is known to literature at the mechanism level. In doing so, I unconver (and later relax) an apparent tension in the literature: some researchers conclude that proximity leads to more diversity in interaction and relationships, and some say proximity leads to less diversity. Observing this apparent conflict leaves me surprised to find that there is no overarching framework to consider the effects of proximity other than to list the three “mechanisms” of proximity I listed

earlier, and I do not see this tension being addressed or expanded upon. In the rest of the introduction to my dissertation, I outline such a framework and use it to analyze what I consider the most relevant and important papers that push our knowledge of proximity further, to show what we know – but also map what we do not, setting a research agenda and situating my dissertation in the literature.

FRAMEWORK DEVELOPMENT

To consider proximity systematically, I begin with a fundamental observation and build on it logically to create a framework that incorporates the abstract mechanisms from prior literature. The fundamental observation is that proximity affects a collaboration because proximate people *decide* to interact. The scope of my study is, therefore, naturally limited by interpersonally determined types of interaction. I believe that proximate people choosing to interact is a theoretically interesting but often overlooked feature in literature around the productive effect of proximity. People could just as easily choose not to interact, or they could interact counterproductively. Something needs to make that interaction happen, and that something is not random, and proximity itself, alone, does not guarantee meaningful interactions.

The view that much of the interaction is self-determined (rather than all assigned working teams) is consistent with observations that the successful production of innovation tends to be open and less hierarchically or organizationally restricted by its nature. Organizations pursuing innovation will not create something radically new or access new knowledge by repeating restrictive processes (e.g., Benner and Tushman 2003). Instead, organizations pursuing innovation can have more explorative innovation

by letting employees use some amount of autonomy. People tend to have a say in who they work with. Even in organizations where people are assigned separate projects, one can imagine the potential for those individual's choices to talk together to make a lasting impact. Suppose their ideas come together in a valuable way. In that case, their small natural interactions – talking over coffee, for example, coming up with a joint idea between divisions – can become more formalized over time and fully supported by organizations that might be eager to take advantage of that serendipitous generation of value.

The three abstract mechanisms from the literature should be varyingly important relative to each other in different contexts, especially given the choices people may make in different contexts. As some recent literature shows, different mechanisms should not have identical effects on the innovative outcome (e.g., Catalini 2018). People making new and diverse connections from increasing awareness should lead to more impactful knowledge recombinations than people working together out of convenience, as Catalini (2018) postulates and infers. But we do not know when proximity will enable specific abstract mechanisms. People do not choose "awareness," "efficiency," or "relationships" – they choose, for example, to engage in conversation, follow-up on something, or ask a specific technical question. When do such choices affect collaborations?

To answer this question, I need to go one level closer to the phenomenon because the abstract level of the mechanisms (awareness, convenience, and relationships) from prior literature cannot easily map to useful contexts. I will from here on refer to the abstract mechanisms from literature as "intermediate outcomes," for simplicity and

clarity, and introduce the idea that these intermediate outcomes have their own, more primitive "mechanisms" that are more useful and cleaner to consider. I list the commonly understood primitive mechanisms in the next section, my goal here is to highlight how my approach to the study of proximity requires less abstract constructs than the intermediate outcomes common to literature.

In my framework, I highlight the main primitive mechanisms from literature and map them through the abstract intermediate outcomes – which remain useful as abstract constructs – to their effect on the nature of collaborator choice. Primitive mechanisms (like observing someone's workstation, having a chance conversation at work, or purely socializing outside of work) have clearer conditions on different types of contexts in which they may be productive and clearer implications of how the interplay between mechanisms and contexts may shape a collaboration outcome than when considering the abstract intermediate outcomes alone. I identify primitive mechanisms as either a part of or a self-determined action or enabling a self-determined decision that could influence a collaboration. When isolated, the potential mechanisms of proximity's effect begin to take less abstract, more recognizable forms having nuance. These primitive mechanisms are still not concrete, measurable mechanisms, but they are easier to assign likely contingencies based on contexts than the intermediate outcomes that exist at a higher level of abstraction, where literature to date has focused.

In short, there should be a large difference in the nature of the innovative recombination and relationship formed if you collaborate with someone you met at the bar instead of with someone you find convenient at work. Proximity can form both

collaborations, but not equally, and not with equal effect. In the next section, I describe my framework in more detail and discuss its application.

Framework for understanding proximity and collaborations

To state the overall logic of the framework: primitive mechanisms affect collaborator choice through different abstract intermediate outcomes. The context of proximity between people can enable different primitive mechanisms, and all are subject to choices made by proximate people. The following figure conceptually illustrates the framework.

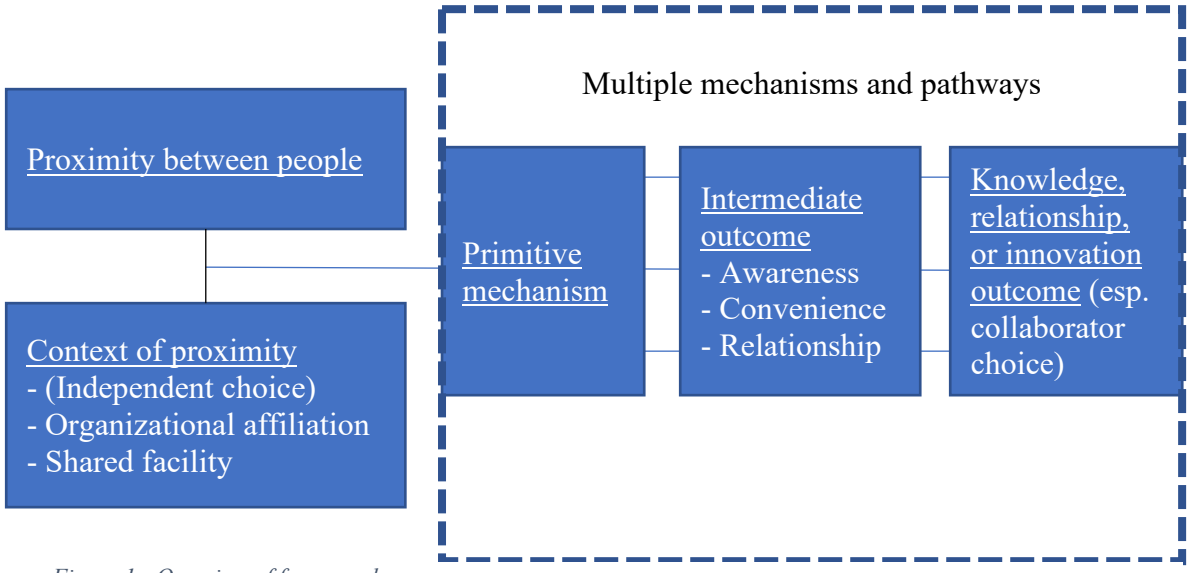

Figure 1 - Overview of framework

Primitive mechanisms that lead to intermediate outcomes

In no particular order, the primitive mechanisms of proximity literature generally describes are: *socializing*, *chance encounters*, *richness of communication*, *the convenience of meeting*, *ease of observation and monitoring*, and *a sense of belonging*.

Socializing outside of work is notably discussed by Saxenian (1996) in her ethnography of Silicon Valley and informal knowledge channels. She describes socializing as a key feature of proximity and a large benefit to the productivity of individuals who locate in Silicon Valley. A recent empirical working paper by Andrews (2020) shows how innovative outcomes that had roots from socializing in bars ended and collaborations changed after bars were closed because of prohibition. He employs a research design exploiting staggered implementations of county prohibition policies, providing causal evidence of the normally highly endogenous and difficult to research process of socializing.

Chance encounters are enabled by proximity. Chance encounters can take several forms, like casually speaking in Allen (1977) and Catalini (2018), but also simple observation of a workspace can convey knowledge, as in Roche et al. (2020). I discuss chance encounters in greater detail in chapter 4 of this dissertation.

Richness of communication is a key feature of face-to-face communication, and being face-to-face is determined by proximity. Storper & Venables (2004) describe how face-to-face communication is a communication technology that transmits several types of difficult to encode information. Proximity can also increase the frequency of feedback,

which I consider another dimension of "richness." These features make face-to-face communication inherently different and richer than other types of communication.

The convenience of meeting generally increases for collocated ongoing projects, enabling people to meet with less frequent planning. But as Catalini (2018) shows, convenience can also lead to suboptimal collaborations through selection.

Ease of observation and monitoring increases with proximity. Arzaghi & Henderson (2008) argue that proximity should increase trust, or at least the ease of observation and monitoring that comes with proximity can effectively substitute for trust between people in different organizations, and this can lead to sharing information and collaborations.

A sense of belonging with others is described by Putnam (2000) as a feature of proximity that can change over time in regions. Kane et al. (2005) experimentally induce a sense of belonging by spending time with others.

This list is not a complete inventory of what could fall into primitive mechanisms, but it contains the most popular primitive mechanisms that I could find in literature relating to innovation and collaboration. In my framework and analysis, I ignore, for example, that proximity can have negative effects on productivity and interpersonal relationships, for which there is good evidence. For example, nearby people can be distracting, as Bernstein (2012) shows how private spaces away from others can improve productivity. Frequently seeing nearby people can also foster feelings of social comparison and develop rivalries, as Lane et al. (2020) show between proximate scientists who do similar types of research. There are also hierarchical nuances to

consider, as Torras & Cummings (2020) show that when an employee is physically proximate to their boss, they change their behavior to limit conversations with other employees, even if work-related, for fear of being watched. I expect these negative effects of proximity should also influence collaborations and innovation, but I do not include them in my framework for analysis because I find no literature linking them to innovative collaborations.

Primitive mechanisms' effects are conditional on their context

If a colleague is unknown or known to be unpleasant, watercooler conversations with that colleague are not likely to be productive if they even happen at all. The person with the question may or may not decide to reach out to those proximity makes available; momentary proximity does not cause collaborations independently. Social norms and expectations, the judgment of the helpfulness or knowledge of the colleague, and relationship factors all influence the final decision to talk (Borgatti and Cross 2003). These factors have multiple antecedents but can also be shaped by proximity over time.

The primitive mechanisms mentioned in the literature are usually described as taking place in certain contexts – and these contexts may well enable them. We do not know if a study may have a similar effect in a different setting, but the context of a study specifies the scope of that study's contribution. Most papers do not examine whether the context is responsible for the mechanism, because most papers do not even mention mechanisms on this level. But some general simple distinctions between contexts known to lead to voluntary communication can be helpful.

Specifically, literature shows the following contexts should increase rates of communication between people: if the people have a shared organizational affiliation (Borgatti and Cross 2003), if they are within a shared workplace (Allen 1977), if they share social factors which make them more likely to communicate (Kleinbaum et al. 2013), or if they have the required level of individual absorptive capacity – if they can understand what the other person might be able to tell them (Nooteboom 2000). I omit social factors and individual absorptive capacity from being a regular analysis component because of limited studies in these areas, though several papers mention them as likely boundary conditions.

Variance in the context of organizational affiliation is present when combining the two separate but related literature streams I consider for analysis: literature on regional knowledge spillovers and literature on interpersonal collaborations. The study of spillovers from the agglomeration literature is especially interested in knowledge exchange between firms, so there is a significant amount of research in the proximity of different firms, where employees from different organizations may share knowledge. On the other hand, collaboration literature generally assumes people have no organizational constraints and focuses more on other interpersonal or knowledge factors. Theoretically, interactions between people in different organizational contexts should be different from those within the same organization. For this reason, I add an indicator if a study is between or within an organization.

Likewise, the context of interaction within a shared facility or elsewhere can determine the nature of that interaction and how proximity may influence outcomes. For

example, Saxenian's (1996) description of knowledge sharing is outside of a shared facility, while Allen's (1977) description is between workers within a facility – we can imagine the nature of the conversation, the type of information shared, and nature of relationships formed in each case may be very different based on the two different settings. Social forces like homophily are likely more salient in settings where socializing occurs, like outside the workplace.

Most papers do not make social factors or individual absorptive capacity clear enough to use as an element of the context of proximity for analysis, so I do not include them. A future, fuller picture of proximity contexts should include these factors, and will likely require drawing from literature in different areas.

Intermediate outcomes of proximity: awareness, convenience, relationships

As I stated earlier, the shorthand for the three abstract intermediate outcomes of proximity from literature are awareness, efficiency or convenience, and relationships. They can each affect collaborator choice differently. Awareness of others and their knowledge can enable new connections to new knowledge, and the others do not. However, the other intermediate outcomes can still shape collaborator choice, and with that choice, the nature of the knowledge used in the recombination. The table below shows the general pathways these mechanisms connect to collaborator choice in literature.

Table 1 - Table mapping mechanisms to collaborator outcomes

Primitive Mechanism	**"Intermediate" proximity outcome**	**Collaborator choice outcome**
Chance encounters, **Socializing**, **Observation**	**Awareness** of proximate people as potential collaborators and awareness of their knowledge or expertise	Collaborator choice set includes proximate options that would be excluded if not proximate
Ease of meeting, **Richness of face-to-face communication** (e.g., frequent available feedback, non-codifiable information)	**Efficiency** of collaborations from both convenience and richer communication	Collaborator choice may shift toward convenient proximate collaborators with increasing expectations of productivity
Sense of belonging, **Ease of observation and monitoring**, **Richness of face-to-face communication** (e.g., nonverbal cues), **Socializing**	**Relationship** formation or strengthening; feelings of trust and affiliation increase with proximity (or trust can be substituted)	Collaborator choice may shift toward trusted and liked potential collaborators, as well as social connections

Primitive mechanisms like chance encounters, socializing, and observation likely increase awareness of others and their knowledge, giving new collaborator options and new knowledge to recombine.

The convenience of meeting and richness of communication increase collaboration efficiency, which can affect collaborator choice if collaborators expect and are motivated by increased efficiency before starting the collaboration.

A sense of belonging, ease of monitoring, and socializing (again) can build relationships where trust and affiliation can be included as factors for collaborator choice. Trust and social connections are generally determined by repeated interactions over time and are shaped by some social tendencies like homophily. Not all types of interaction lead to trust and social connection, and those that do are not guaranteed to do so with any given people. Social and trusted connections are not limited by proximity, but those proximate are likely stronger connections.

Noting induced variation can clarify the contribution of a study

When describing *how* proximity affects collaboration or innovation, most prior literature lumps the mechanisms of proximity all together for understandable empirical reasons. Researchers can precisely measure proximity, and it is often impossible to know the true origins of an innovation down to the level of individual conversations that started it. Besides, as I stated earlier, the mechanisms of proximity can all be interrelated – one might ask a friend for advice because they're nearby, but that friend might *be a friend* specifically because they're nearby, so it can seem meaningless to try to separate the specific reason they started working together if both reasons come from proximity. Innovative collaborations that happen because of proximity also take such an endogenous process it can be unclear what a theoretical counterfactual to a given collaboration could even be. For this reason, most prior research has been observational rather than attempting to establish causality.

Two foundational examples of observational literature in the study of proximity are Allen (1977), describing communication patterns in an R&D facility, and Saxenian

(1996), describing informal communication networks in Silicon Valley. Their books do not present a counterfactual or make a causal argument, but they do give us evidence of the importance of the nuances of physical proximity between people, both within and between firms, respectively. Only recently have researchers attempted to implement causal inference techniques in creative and experimental ways, and these studies largely support Allen's and Saxenian's observations but also serve to highlight how little we still know from quantifiable evidence.

How and where studies with empirical evidence invoke variance in their sample is an important ingredient to understanding the contribution of empirical studies. For this reason, I include in my analysis how a study's research design invokes variance. I do this by dividing studies into two major categories: those that invoke variance in the *proximity* between people and those that invoke variance in the *behavior* of proximate people. Each category has different implications.

I also note if authors claim their variance is plausibly exogenous. For such papers, effective exogenous "treatments" should relate to the outcome variables and otherwise be appropriate given their described mechanism. In other words, researchers can't randomize friendships – but researchers can effectively induce exogenous variation in some types of interactions that might lead to friendships if people so choose.

The *choice* to engage with others is an important theoretical consideration in my analysis, and is also an empirical consideration. In any real-life setting, all the interpersonal processes induced by proximity are subject to the people's choices. But not all research considers people or proximity in this way. To accommodate helpful studies

that directly induce choices, rather than observing a natural choice, I indicate the study as in an “assigned” vs. “natural” setting. In “assigned” settings, people are assigned conditions to network at an event. In “natural” settings, proximity outcomes are determined by people's choices to engage with those around them. “Assigned” studies can show us interesting insights because of the level of control over the research design, but they differ from studies of proximity by removing the theoretically necessary component of individual choice.

Framework summary

To reiterate the framework's logic: collaborator choice is affected by abstract intermediate outcomes, which are determined by potentially context-dependent primitive mechanisms. Collaborator choice outcomes that are especially relevant to innovation production are increases or decreases in the variance and diversity of collaborators. Both an increase and decrease in the variance of collaborators are potential outcomes of increasing proximity.

LITERATURE REVIEW USING THE FRAMEWORK

In the following section, I use the framework developed from the previous sections to assess and summarize notable research in the area and compare them to the essays of my dissertation. I compare them to my dissertation instead of including analysis for each paper because this framework aims to introduce the contribution of my research. However, a fuller analysis of each paper should also highlight opportunities for further research that may be interesting more broadly, so I plan on expanding this introduction to

a review paper in the future by systematically describing each paper on its own terms, and by including work in other areas on social factors and individual absorptive capacity.

Paper selection criteria note

Interpersonal mechanisms of proximity have been claimed in research at least since Marshall (1920), usually lumped together. In this exercise, I am interested in papers focused on untangling these mechanisms and intermediate outcomes rather than showing the net effects of some unseen arrangement of them. Therefore, I limit the scope of papers to those that discuss *how* proximity affects collaborations, only including papers that explicitly discuss or attempt to show something about the mechanisms that proximity may work through. I, therefore, exclude many firm- or region-level impactful and relevant papers that evaluate the effect of proximity of people on innovation, knowledge exchange, and innovation (e.g., Sorenson and Stuart 2008).

Framework table

Table of notable prior literature on the mechanisms through which proximity acts on collaboration. Unless otherwise noted, X = condition met, O = condition explicitly not met, - = condition not directly addressed or NA, parentheses () indicates implied. This table is not exhaustive; management and strategy studies are prioritized for inclusion.

Table 2 - Table of notable prior literature

	Context of proximate people			Induced variation		Type	Intermediate outcome			Outcome		
Paper	Processes of proximity can naturally occur over time (vs assigned)	Within or Between organizations	Within shared Facility or Elsewhere	Exogenous (or exogenous Influence of) **proximity**	Exogenous influence on **behavior** with proximate people	Nature of interaction determined by proximity	Knowledge access & awareness	Convenience	Social relationship development	Relationship outcome (Not necessarily a collaboration)	Knowledge transfer outcome (non-collaboration)	Innovation or collaboration outcome
Chapter 2	**X**	**W**	**E**	**I**	**-**	**Social**	**X**	**O**	**X**	**Homophily**	**-**	**X**
Arzaghi & Henderson (2008)	X	B	E	I	-	F2F netw.	X	X	X	Trust	X	-
Kleinbaum et al. (2013)	X	W	F	I	-	Social	-	-	X	Homophily	-	-
Saxenian (1996)	X	B	E	-	-	Social	X	-	X	-	X	-
Roche (2019)	X	-	(E)	I	-	-	X	-	-	-	-	X
Crescenzi (2016)	X	-	(E)	-	-	F2F work	-	X	-	-	-	X
Jacobs (1969)	X	-	E	-	-	Social	-	-	-	Diversity	X	-
Vakili & Zhang (2018)	X	-	-	-	X	Social	-	-	X	Diversity	-	X
Reagans (2011)	X	W	F	-	-	Social	-	-	X	Homophily	-	-
Chapters 3 & 4	**X**	**W**	**(F)**	**-**	**X**	**Chance**	**X**	**-**	**X**	-	**-**	**X**
Catalini (2018)	X	W	F	X	-	Chance	X	X	-	-	-	X
Andrews (2020)	X	-	E	-	X	Social	X	-	-	-	-	X

Boudreau et al. (2017)	O	W	F	-	X	(assigned)	X	-	-	-	-	X
Lane et al. (2020)	O	W	F	-	X	(assigned)	X	-	-	Competition	-	X
Lee (2019)	X	W	F	X	-	-	X	X	-	-	X	-
Allen (1977)	X	W	F	-	-	Chance	X	-	-	-	-	-
Kabo et al. (2014)	X	W	F	X	-	Chance	X	-	-	-	-	X
Torras & Cummings (2020)	X	W	F	-	X	F2F work	-	-	-	-	-	-
Roche et al. (2020)	X	B	F	X	-	Observe	X	-	-	-	X	-
Chown & Liu (2015)	X	W	F	X	-	Inf	-	-	X	Power	-	-

Conclusions from analysis, as it relates to this dissertation

Several findings emerge from this analysis and situate my dissertation in the literature. As I mentioned before, studies show that proximity can both increase (Catalini 2018, Jacobs 1969, Lee 2019, Roche et al. 2020) and decrease (Kleinbaum et al. 2013, Reagans 2011) the diversity of interactions. When viewed through the framework, I see this apparent disagreement as a feature of proximity conditional on the choices people make to collaborate, rather than tension in the literature. When people make choices to engage with others, they can tend to display homophily in those choices. To theoretically connect proximity to the increase or decrease of diversity of collaborators, we should evaluate the contexts and mechanisms of proximity where people are more likely to display homophily. Chapter 2 of this dissertation explores this connection, but I want to do more because I believe variance in homphilic tendencies between proximate people is a potentially rich direction for innovation research.

Chapter 2 finds that the demographic diversity of collaboration partners decreases in a context outside of a shared facility, consistent with a socializing and homophily story as in Kleinbaum et al. (2013). Kleinbaum shows that homophily is responsible for some social relationships between proximate people within a facility context, measured by email volume between the pair, but does not relate these connections to strategic or innovative outcomes. I show that homophily can be a major factor in collaborator selection outside of a facility, showing that all collaborations likely formed outside the facility are the same gender. Regrettably, due to data limitations, I can

only infer that socializing versus other mechanisms occurs, but I am confident that "socializing" would describe most interactions outside of the firm.

But I also find that these connections decay when proximity is removed within pairs, so these do not seem to be the type of deeper social relationships that are generally expected to persist after physical separation (Agrawal et al. 2006). I may then be seeing the effect of socializing through the intermediate outcome of *awareness* and not through the intermediate outcome of *social relationships*, both of which are theoretical options supported by prior literature. Through induction, chapter 2 gives novel evidence that socializing because of proximity may have shallow effects on relationships but meaningful impacts on awareness.

Further distinguishing my second chapter from similar work, I only select workers at the same facility. Having only potential collaboration partners at the same facility focuses and clarifies my contribution by eliminating the possibility that working together is more efficient or convenient if proximate – these benefits would be constant because they would work at their shared facility. I, therefore, remove from consideration the potential mechanisms that lead to convenience and efficiency that may come with proximity. The mechanisms leading to convenience and efficiency are plausible and likely responsible for the effects of proximity in studies like Crescenzi (2016) that do not consider the facilities of inventors.

Also, by selecting workers from the same organization, trust and the need for monitoring that are central to the story of Arzaghi & Henderson (2008) become

unnecessary. Chapter 2 is theoretically distinct from that paper when seen through this analysis.

The question remains if the increase in homophily in collaboration partners I observe with increasing proximity indicates anything about the nature of the recombined knowledge in that collaboration. If demographic and knowledge diversities correlate, this study could show how socializing might lead to less novel outcomes. However, if the proximity mechanisms operate through awareness in this context, as I earlier inferred it might, it could still increase the diversity of knowledge by expanding the set of people for consideration. I do not test these ideas in chapter 2 because of data limitations. Future work should examine the interplay of the context of proximity outside the firm versus inside the firm, the tendency of homophilic collaborations between each, and the novelty of the recombinations of knowledge used.

Furthermore, homophily can also lead to the societally undesirable outcome of intensifying the exclusion of already marginalized groups. Future research in this area should show if there is an optimal place for the proximity of employees for both innovation and social equity motivations.

Chapter 4 works with the measure developed in chapter 3 to explicitly consider the effect of proximity on the variance and diversity of knowledge used in an innovative collaboration, specifically as the novelty of an invention. Chapter 4 focuses on the fleeting mechanism of chance encounters, like at watercoolers or coffee stations.

I assume that most chance encounters that can mature and develop into innovations likely happen at a shared workplace and within an organizational context, and these assumptions align with most collaboration literature.

By exploiting variance in local flu, I attempt to vary one behavior of proximate people – chance encounters – while holding the other potential effects of proximity constant. Chance encounters are likely seen as entirely optional and decrease by someone practicing social avoidance. Socializing and planned meetings may also decrease, but I expect people engaging in social avoidance will see chance encounters as especially optional events to avoid. And chance encounters are especially theoretically interesting because they should only increase the intermediate outcome of awareness, and increased awareness should lead to more variance in knowledge sources and more novel innovation. My chapter 4 bridges the full framework by isolating the primitive mechanism of chance encounters and connecting the phenomenon both theoretically and empirically to an effect on the novelty of the innovative output of the collaboration, something no other study has done. I also present a counterfactual to collaborations formed by chance encounters – collaborations tend to be with prior contacts when not from chance encounters – and show that these counterfactual collaborations may be efficiently produced, but less novel.

Of literature most closely related to chapter 4, only Catalini (2018) arguably shows that different proximity mechanisms can affect the nature of innovation. My results align with his, by in large. But our approaches are entirely different. I benefit from a large sample and a completely different way of inducing variance that does not

introduce the confounding influences of the rest of proximity changing. I only vary the behavior of proximate people to isolate the mechanism of chance encounters, I do not vary proximity. Catalini varies proximity and relies on an inductive argument that additional search costs before a move leading to an especially impactful collaboration give evidence that chance encounters may be affecting those collaborations by increasing awareness. But the quantity or nature of chance encounters between the impactful and non-impactful collocated collaborations do not vary in his study. I first show the mechanism of chance encounters may be affected by an exogenous treatment measure (in chapter 3), independently of the effect that change might have had on an innovation, giving stronger evidence that chance encounters are responsible for my findings.

On a surface level, Vakili & Zhang (2018) seem to make a closer empirical theoretical and empirical approach to my chapter 4. They also observe an increase of the diversity of knowledge in innovations by invoking a change in behavior proximate people. However, they propose that policy shifts at a state level can change the propensity of individuals to engage with more diverse people. In their study, the policy leads to increasing diversity of knowledge independently, regardless of proximity (they appropriately do not claim to study proximity).

Any study is only as strong as its assumptions, and mine contains steps where I posit or argue my interpretation rather than giving evidence, due to data limitations. While researchers cannot directly study the effect of proximity on collaborations with a level of detail that preserves choice and clearly separates each potential mechanism, a

cumulative body of knowledge is being built by varying research designs that can relax various assumptions in turn. My dissertation adds to this body of knowledge.

CHAPTER 2: IT MATTERS WHERE PEOPLE LIVE: INNOVATIVE COLLABORATIONS AND THE PROXIMITY OF INVENTORS' RESIDENCES

(Coauthored with Myles Shaver as a separate manuscript.)

CHAPTER SUMMARY

We examine if geographic proximity of where knowledge workers live affects the likelihood that they collaborate on patents. We argue that the underlying mechanisms for why proximity in the workplace affects collaboration extend beyond the workplace. Therefore, we expect that residential proximity enhances workplace collaboration. We test this prediction by examining patenting outcomes and inventor residential location in the Warsaw Indiana orthopedic device cluster – a setting that provides many research design advantages. We find the likelihood of collaborating decreases 0.3% with each kilometer coworkers live apart. This effect is robust after controlling for other factors that might spuriously lead to this result. Documenting residential proximity as a novel within-firm agglomeration measure has the potential to explain different innovation outcomes between firms and across regions.

An extensive body of research documents that geography shapes firms' knowledge and innovation through knowledge spillovers between geographically proximate firms (e.g., Fleming, King, and Juda, 2007; Jaffe, Trajtenberg, and Henderson, 1993). As a result, agglomeration effects establish the foundation of many studies in strategic management (e.g., Alcácer and Chung, 2007; Shaver and Flyer, 2000). To further understand the importance of innovation and agglomeration effects, and the conditions where they are most prominent, we investigate a micro-level mechanism (i.e., information of individuals' locations within a region) that can lead to agglomeration benefits.

Specifically, we examine if the proximity of inventors' residences affects the likelihood that they collaborate. We investigate residential locations because existing research shows inventors' workspace location within their workplace affects collaboration (Allen 1977, Van Den Bulte and Moenaert 1998, Catalini 2018). Because individuals spend a significant amount of time outside of where they work, we believe that non-workplace interactions can shape work-related outcomes, as well. We examine the empirical relationship between where inventors within a company live and the probability that they collaborate on patents, with an attempt to isolate the relevant mechanisms that promote this outcome.

By examining the orthopedic medical device cluster centered on Warsaw, Indiana between 2001 and 2015, we find the location of inventors' residences impact the likelihood that they collaborate on a patent. The likelihood that a pair of individuals who work at the same facility collaborate on a patent application decreases by 0.3% for each

kilometer the pair lives apart. This effect is robust once controlling for other factors that might spuriously lead to this relationship. Further investigating this relationship, we discuss how social factors, notably gender, appear to play a role in determining this effect.

Finding that employees' residential location affects innovative outcomes contributes to the literature in two notable ways. First, we document a novel *within-firm* agglomeration measure: the distance that employees reside from each other. Our finding is consistent with existing literature that studies increased collaboration between individuals based on the proximity of their workspace. The notable extension is that we demonstrate that geography outside of the workplace affects work-related outcomes. Second, by isolating this individual-level measure of geographic location, we provide a mechanism to explain different agglomeration outcomes across regions and companies within a region. Namely, regions with similar bases of talent might have different innovation outcomes because of regional characteristics that affect where people live. These can include geographic features (e.g., land shape and bridges) or socio-economic features (e.g., cost and availability of housing). Likewise, where a company locates within a region and where their employees live has the potential to create differing agglomeration effects within regions.

In the next section, we motivate our study by situating it in the existing literature. We then describe the history and unique characteristics of the orthopedic medical device cluster in Warsaw, Indiana, the setting of this study. From here, we describe the research design challenges in assessing residential locations and innovation outcomes, and how

the setting overcomes many of those challenges. The following sections document our empirical strategy and results. The final sections draw implications and conclude.

THEORETICAL CONTEXT

Existing research argues that encounters between individuals, facilitated by geographic proximity, are an important determinant of regional agglomeration effects (e.g., Rosenthal and Strange, 2004, Hall, Lecuona, & Cummings, 2017). The reason proximity affects knowledge sharing is because proximity lowers collaboration costs, which we can classify as search costs or execution costs (Catalini 2018).

Search costs

Search costs restrict the set of individuals that a searcher considers to have helpful knowledge. Such costs include network limitations (a searcher knows a limited number of people), cognitive limitations (a searcher might not recall all individuals in their network), knowledge limitations (a searcher has to know what people in their network know), and absorptive capacity limitations (a searcher has to understand that what people in their network know could be helpful to their current problem). As a result, lowering search costs increases the likelihood that a searcher considers someone to have potentially helpful knowledge, which in turn increases the likelihood of collaboration (Boudreau et al. 2017).

Physical proximity is one means to lower search costs through observation and chance interaction. Observation is simple; more proximate individuals see each other more often, which increases network inclusion and presence of mind. Individuals need not interact for the searcher to be more aware of the other through observation, although

the amount of information the searcher can gather from observation is limited.

Compared to observation, chance interaction results in greater information transfer. Chance interactions are encounters where people communicate without planning to, and these interactions happen more frequently with proximity. Such interactions can increase the searcher's knowledge of what others' know, as well as increasing network inclusion and presence of mind. Over time, chance interactions might increase a searcher's ability to understand how those nearby might have helpful knowledge to their own technical problems.

Observation and chance interactions occur at individuals' workspaces; however, they also occur at third locations such as a coffee station within the workplace. Field experiments show this effect robustly, describing common walking paths (Kabo et al. 2014) and that collaboration can be affected by nuanced factors like manager locations (i.e., people might not want to walk past the boss's office, Hall *et al.*, 2017).

Chance encounters also occur outside the workplace. Like the effect of common walking paths demonstrated within facilities, research shows neighborhoods with more connectivity likely result in increased serendipitous encounters (Roche 2019). Encounters at common third spaces like grocery stores, coffee shops, schools, or sporting events are more likely for people who live near each other due to their joint proximity to the space or inclusion in an administrative boundary like a school district. A searcher's encounters with people who have potentially relevant knowledge should include more encounters with those who live nearer their own home because of such overlaps driven by their residential proximity. As a result, residential proximity should affect search costs

and collaboration.

Execution costs

Execution costs are barriers that keep a person searching for knowledge from actually accessing knowledge in other people and from completing a collaborative project (Catalini 2018). Whereas search costs primarily reflect searcher awareness, execution costs consist of practical and social considerations. Practical barriers to sharing knowledge may include conflicting motives (e.g., they work for rival firms) or convenience (e.g., it is burdensome to meet in person). But social barriers and organizational norms can also prevent knowledge transfer (Borgatti and Cross, 2003). Thoughts of violating organizational norms like bypassing hierarchy or established communication channels can elicit fear of repercussions, preventing contact. A searcher may also fear damage to his or her group's reputation by appearing weak. Such organizational and social barriers limit collaboration.

Proximity lowers execution costs because of increased convenience meeting in person (Catalini 2018), increased affinity for those who live in the same area (Turner et al. 1987), an increased likelihood of social relationships (Agrawal et al. 2006), and increased trust (McEvily et al. 2003).

Although the convenience of meeting is an important proximity consideration in within-facility studies, we expect it plays a minor role outside of a shared workplace. We expect that knowledge workers do not often plan on meeting outside of the workplace to do work related to their employer. Rather, they tend to perform research-related meetings in their work facility.

Alternatively, we expect that residential proximity increases a sense of in-group belonging, where attitudes become more positive toward others in the perceived group (Turner et al. 1987). Increased affinity toward others in an in-group may lower social costs among people who are in the same neighborhood, or school district, or another local area of which they are both informal members due to their residential location. Lowering perceived social costs increases the chance that two proximate people collaborate.

Residential proximity also increases the likelihood of a social relationship between people, and a social relationship should lower nearly all of the mentioned search and execution costs (Agrawal et al. 2006, Kleinbaum et al. 2013).

Clearly, multiple mechanisms outside the workplace could lead to an effect of proximity of residences on collaborations. These mechanisms are assumed to exist in the literature (e.g., Agrawal *et al.*, 2017; Rosenthal and Strange, 2004) and are generally considered all possibilities without differentiation. Untangling these mechanisms directly would require a prohibitively deep level of data on individuals, their thoughts, and their relationships, so we do not attempt to untangle mechanisms in this study. Our goal is to see if the relationship between proximity and collaboration exists on the level of the actors and their residences.

In summary, we expect that many of the underlying mechanisms that decrease search and execution costs for collaboration within a workplace hold in the context of where individuals within a workplace live. We turn to investigate this possibility in data. To do so, we choose a setting well-suited to isolate this effect. We examine a relatively geographically isolated cluster within a technologically innovative industry.

WARSAW INDIANA ORTHOPEDIC DEVICE CLUSTER

The small town of Warsaw, Indiana, is a global center of orthopedic implant and instrumentation development. This region has a long history in medical device innovation, starting with the humble creation of the low-tech splint manufacturing industry in a hotel room by traveling salesman Reverend DePuy in 1895. The orthopedic industry grew organically, mainly in place, into an innovative, multi-billion-dollar industry (Bormet 2001).

The location of DePuy's founding of DePuy Manufacturing (currently DePuy Synthes, a Johnson and Johnson company) was not a strategic choice considering local knowledge: there was no relevant industry nearby, and no research centers. The location was likely motivated by the location of the local county sheriff's daughter, whom DePuy married the year after settling in the town. The industry has since evolved from rudimentary splints to sophisticated orthopedic implants, an evolution reflecting consistency with musculoskeletal medical treatment areas but reshaped by advances in medical technology. The town still boasts its status as the "Orthopedic Capital of the World" with local companies representing roughly a third of the global market share in orthopedic devices and 60% of the global hip and knee replacement market. DePuy itself is the global market leader with 20% market share of the $46.6 billion orthopedic market in 2015. In the period of this study, the three of the five largest orthopedic firms in the world, DePuy, Zimmer, and Biomet, were all headquartered in Warsaw.[1]

The Warsaw Indiana region is relatively unique in its isolation and concentration

[1] In late 2015 Zimmer and Biomet began a merger to become the single entity "Zimmer Biomet" – but are separate for the duration of this study.

for also being an actively innovating knowledge cluster, and this stems from its history. The cluster's development is a story of localized path-dependency and opportunism rather than strategic choice because the area has no significant connections to outside research facilities, universities, or otherwise complementary sources of specialized knowledge or technology. Zimmer Manufacturing was founded in 1927 by an early employee of DePuy Manufacturing, and Zimmer's first facility was less than a mile away from the DePuy facility. Biomet was founded in 1977 by former employees of Zimmer, also within the town. The initial choice of location within the town for each was likely driven merely by the locations of the founders, which itself was determined by their prior employers. A growing concentration of knowledge workers and a local system of suppliers, manufacturers, and other service providers that sprang up in support of the industry likely helped Zimmer and Biomet decide to stay in town. Today, the R&D facilities of these three companies remain all within a short distance of each other in Warsaw. The supporting manufacturing ecosystem has a reinforcing effect on the isolation and concentration of the industry in this small and isolated region.

A 2009 study of the region documented that the orthopedic industry employs about 6,500 people in Warsaw, Indiana – nearly half the town's population (BioCrossroads 2009). Such a high portion of employment in one industry shows that the orthopedic industry is not only concentrated in town, but the town is isolated to the orthopedic industry.

The rural location presents over an hour drive to the nearest cities of South Bend or Fort Wayne, neither of which are large cities. At a regional development roundtable,

area leaders expressed the sentiment that the geographic isolation negatively affects their ability to attract skilled employees. An excerpt from a report on that roundtable:

> *"One company executive said that he often brings in a candidate for an engineering position, makes a job offer, and then has to tell the prospective employee that his or her new home will be a one-hour drive to the airport, mall, and closest auto dealership. Far too often, such factors become deal breakers for successful recruitments."*- (BioCrossroads 2009)

The remoteness likely has a filtering effect on the local labor pool as described. While the isolation of the region is empirically convenient for study, it also limits the generalizability to those who select into a remote environment. However, the inventors in this study are similar in many ways to research scientists and product development engineers in more cosmopolitan cities in significant ways. Orthopedic product development requires a high level of knowledge of relevant anatomy and biomechanics, materials, and advanced manufacturing. Inventors here have generally received a high level of formal education outside the region and are taking part in industry-leading innovation in a technology-intensive field.

Innovation in the orthopedic industry

The orthopedic industry is competitive, and product offerings within a product category (e.g., total knee replacement systems) may have only minor differentiation between firms. However, minor points of difference are often the basis of competition between similar products. In this industry, differentiation is usually in surgical

instrumentation affecting ease, speed, or accuracy of the surgical procedure. Because small differences can result in products that command higher prices, nearly all orthopedic firms engage in research and development efforts. Competitors can easily observe and copy each other's product characteristics, and the desire to protect improvements results in a high patenting rate in the industry.

Research scientists and engineers are organized within orthopedic companies by product divisions that correspond to musculoskeletal treatment areas (e.g., knees, hips, etc.). Between divisions, employees have a similar educational background in biology, engineering, and material science. Engineers and scientists moving between divisions during their career are common, and moves can sometimes bring a new perspective on instrumentation in a new application. A quote from an interview we conducted with an experienced engineering manager at one of the major companies in the area illustrates:

> *"When I changed over to knees (from shoulders) I could see right away that a solution from shoulders could work in the knee procedure with some adjustments. That shoulder solution resulted in a few patents when we worked it out in the knee group."*

This quote illustrates a few points, one is that some instrumentation knowledge can be useful between divisions. The quote also suggests that the divisional structure may be imposing search costs between siloed divisions, preventing valuable knowledge transfer. It would be easy to imagine that had there been lower search costs between divisions the knee group should have had the shoulder instrumentation insight much earlier, and it would not have required an individual's mobility between departments to

transfer knowledge effectively.

The silo effect is well known to senior engineering managers like the one quoted above. He describes a brainstorming session common at his firm when a team has a technical problem:

> *"If we have a problem, if we're stumped, that means to me that the solution isn't in the group. So, what we do is we invite people from hips, from clinical, from other departments – we're looking for people who aren't as locked into our way of thinking about the problem. They'll come by for a big session, we sometimes work it out that way, or we don't. Usually, we don't, but it's at least a start, and we've got some other people thinking... Ultimately I'd say it's a very organic process how we find a solution."*

In this example, those who get invited to brainstorming sessions seem to be determined by who the organizer knows and who is available. When asked who he knows in R&D functions at his firm, the manager claimed to know all 70 employees engaged in research and development in his division, but only about 10-15 in each of the other divisions.

He could know people from other divisions for any reason, internal or external to the firm itself. The reasons for inviting any specific individual, as we posited, could be heightened awareness, affinity, or relationship related to the proximity of their residences. He might invite his scientist neighbor since he saw her that morning (increasing awareness of the scientist lowering search costs); he might invite his running club friend

who works in the machine shop, since he remembered a conversation that made him think the machinist could help (increasing awareness of the knowledge of the machinist lowering search costs); he might also invite a senior engineer from a rival division, since they both coach little league together and this rival division engineer seems nice (lowered perceived social costs from a potential rival).

We present the brainstorming session invitation list as a specific example of one of the many ways that proximity might lower collaboration costs. But we find it helpful to use a specific example to illustrate how different collaboration costs can be affected by proximity of residences in concrete ways. A similar collaborator selection process occurs on a smaller scale much more often. Pairs of people within and across divisions and teams have varying levels of collaboration costs, many of which may be related to proximity.

EMPIRICAL CHALLENGES

Quantifying how residential proximity affects work outcomes presents several empirical challenges. Foremost, common measures of innovative collaboration rely on patent filings; however, it is difficult to find co-inventors' residential addresses from these data. We chose a region and industry for our empirical analyses, in part, because it overcomes this data availability concern. Although the commonly used bibliographic information from patents does not include inventors' residential addresses, the Image File Wrapper documents associated with patent applications, provided by the USPTO, do. Moreover, many firms do not ask inventors to submit their home address for patent office correspondence; however, the firms in this study did. Therefore, we can assess inventors'

residential addresses by manually coding scanned files in the USPTO Image File Wrapper.

Having a data source, the next challenge is that other systematic factors that affect geographic clustering of inventors' residences might also affect collaboration. Again, our empirical setting mitigates some of these concerns. Of particular concern is that people may cluster by shared knowledge type or shared employment history, either of which would affect their likelihood of collaboration. If there is evidence of this type of clustering, it will present a competing explanation for why proximate people collaborate more.

The Warsaw Indiana orthopedic device cluster has no other related industry nearby so we do not expect residential clustering by type of knowledge. People clustering by knowledge is a potential confounding factor in other regions with related industries, especially if there are groupings of specialized firms within or near the region (e.g., steel processing plants near automotive industry centers, or universities near pharmaceutical clusters). People may remain near their former employer when changing jobs to a different sector but may continue to work with people who live around them because of their shared knowledge base. Because there is no nearby industry related to orthopedics, we do not expect this effect to confound our inference.

In addition, there are no major universities that provide orthopedic device industry research and education in the region. Therefore, most people engaged in R&D in Warsaw Indiana have moved from outside the region. Furthermore, the firms' R&D facilities are geographically close together, and variation in home locations offer nearly identical

commutes for each residence to each possible R&D facility. So, in this setting, there is no clustering of residences by a shared employer between or within industries. Finally, the surrounding rural area presents no significant geographic barriers to the spread of home residences, resulting in a significant variance in residential locations. A network of country roads crisscrosses the flat landscape, and people often use country roads to commute more directly than highways; therefore highways are not a major constraint for commuting considerations.

While not insulated from every non-random reason for clustering, the Northern Indiana orthopedic cluster is a setting that mitigates many major empirical concerns mentioned earlier.

RESEARCH DESIGN

We consider collaboration as individuals choosing to work together through a search process subject to search costs that we expect vary with distance between residences. Assuming that residence location choice is made independently of the residences of potential collaborators at their firm, greater distance between residences of potential collaborators will lower the likelihood of collaboration.

To assess this relationship, we consider the set of all unique pairs of inventors who worked at a firm during the year of an invention application as potential pairs of inventors that could have appeared on that patent application.[2] In this way, the sample of potential collaborators is restricted by those at that firm at the time of patent application,

[2] A patent with 3 co-inventors (A,B, and C) would have three pairs of co-inventors (AB, AC, and BC).

and in a role that is relevant to R&D. This design does not consider any variance from between firm-specific residence patterns or shifts in residential location patterns over time because the comparison group updates each year and is unique for each firm.

The dependent variable is a dummy variable that takes the value 1 when a pair of potential collaborators appear as inventors on a given patent application. According to guidance provided by the USPTO, all inventors listed on a patent application must have been necessary for the conception of the invention, and there must have been "…some element of joint behavior, such as collaboration" (USPTO 2017, sec. 35 U.S.C. 116). Therefore, a collaboration between any two inventors on the invention is a necessary condition for that invention.

Research design advantages compared to related studies

To identify our test, we rely on within-region variation. That is, we measure the proximity between the residences of potential inventors within Warsaw, Indiana region. Related studies, which test for proximity and collaboration tend to use across-region variation. That is, they capture whether or not potential inventors live within the same commuting region (e.g., Agrawal, Kapur, and Mchale, 2008; Crescenzi, Nathan, and Rodriguez-Pose, 2016). Our research design better matches the source of variance to the underlying theoretical mechanisms because reduced collaboration costs should vary depending on an individual's day-to-day travel within their region. Therefore, within-region proximity is a more appropriate source of variation to invoke when studying proximity and collaboration.

Sample Construction and Data

Construction of Sample

We identify 1913 patent applications from the orthopedic industry cluster in Northern Indiana over the 2001 to 2015 period. We include patent applications that 1) had at least one inventor in commuting distance from Warsaw, Indiana, and 2) were initially assigned to one of the three major firms in the area (DePuy, Zimmer, or Biomet). We limit inventors to those with Indiana addresses and exclude inventors with extreme commutes to the Warsaw facility. We define an extreme commute as 150km or about 93 miles.[3] We removed only 32 addresses by this threshold when applied to all addresses in Indiana. Those removed were from central and Southern Indiana, with a notable cluster of 8 addresses in West Lafayette – which were likely residences related to Purdue University (and represent an unreasonable commute to Warsaw of about two hours). No pattern of distances from Warsaw indicated that the 150km threshold artificially drives the results.[4]

[3] We base this on the census definition of an "extreme" commute being 90 minutes or more (Jiles 2007) and assessing speed limits in the area.

[4] Concern that more distant individuals may not be employees is further alleviated through a robustness check using self-reported employment information from LinkedIn in a later section.

(Region)

(Detail)

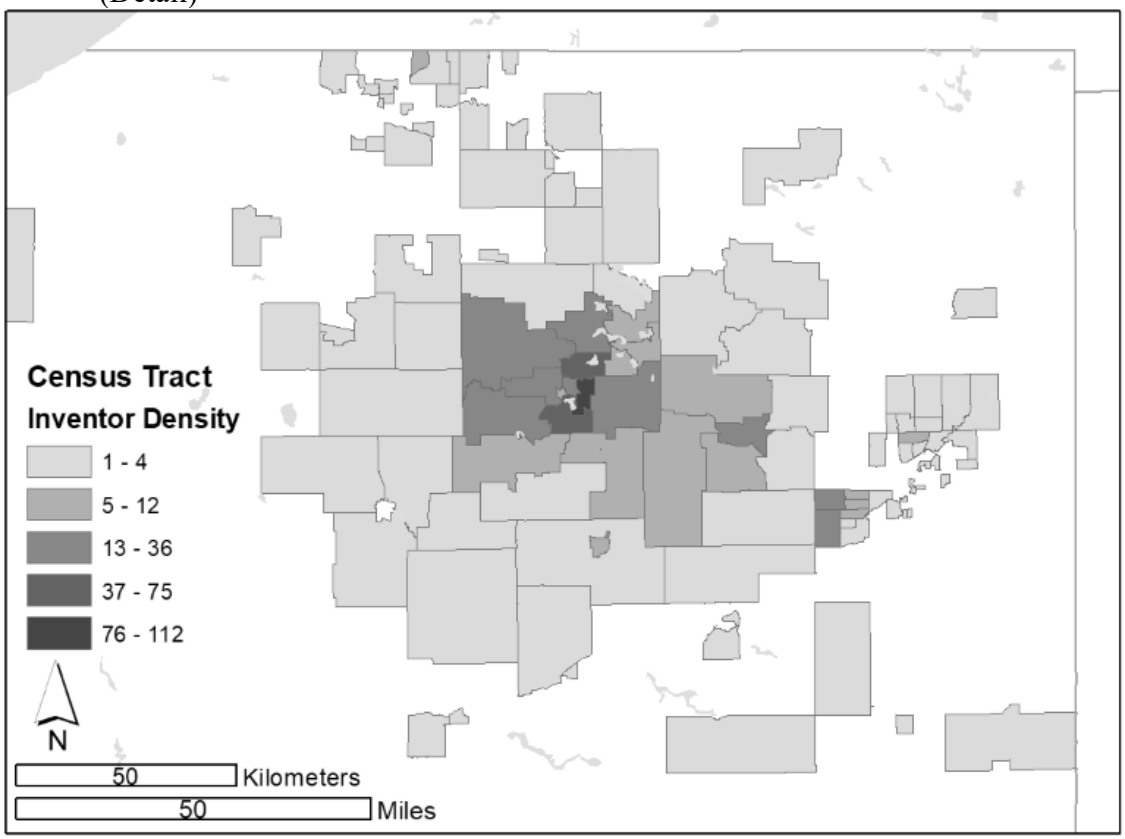

Figure 2 - Distribution of residences of inventors

Inventor residences. Census tracts containing inventors are shaded. Warsaw is in the center concentrated region, as are all R&D facilities. Smaller tract size indicates higher overall population density of that tract, darker tracts indicate more inventors at the Warsaw firms.

Rather than using existing disambiguation crosswalks, we manually disambiguated all inventor names to be confident that all inventors are unique individuals, double-checking all inventors at repeated addresses or with potentially similar names. Our disambiguation reduced the number of seemingly unique individuals in the region from an original number of 917 to 702 unique inventors. These 702 inventors lived at 752 unique addresses over the 15-year range. The discrepancy comes from some inventors moving and some collocating at the same residential address – including at least one married couple working for rival firms.

Because some people move residences within the 15 years, we attempted to obtain at least one instance of an inventor's address for every year an inventor has any patent application. There are 5187 application-inventor data points, a value corresponding to an average of 2.7 inventors from the region per patent application. Of these potential application-inventor data points, we found a contemporary inventor address that year for 5008, meaning that we were able to get a yearly updated residential address for an inventor in over 96% of the patent applications. We applied the closest address in time we had for an inventor for the few instances where a P.O. Box or business address was provided instead of a residential address.

About ten percent of inventors move residences at some point during this study. We consider moves occurring between years, so an inventor is only considered living at one location each year. Most inventors did not patent multiple times each year, so using a finer resolution of the timing of residences is not feasible. Having contemporary residential street addresses, we have a much more robust and detailed address dataset

than other inventor addresses commonly used in research which are only collected at the regional level and often include nonresidential addresses like workplaces and post office boxes.

Construction of Choice Set

To identify meaningful differences in proximity between people, the comparison choice set needs to include all other feasible options of pairs. We assume that any employee engaged in inventing around the time of a given patent application was a possible inventor. Doing so, we assume all other inventors in the firm were potential coinventors on any patent application from that firm regardless of the content of that patent application. As we described previously, the orthopedic industry divides engineers and scientists with similar technological backgrounds into product areas. Therefore, the assumption that any inventor might have appeared on any patent in the firm is not as potentially problematic as it may be in other industries.

If the assumption that any person could be on any application is violated, including additional unreasonable options would increase the number of unreasonable false pairings in the counterfactual group. False pairings would lead to artificially lower standard errors but would not bias the result unless there is a systematic association with residence location.

For each patent application, we assign a set of all possible pairs of inventors we estimate are working in a role that may lead to patenting at the assignee company that year. The set of pairs repeat for each application (and change for each year and firm) to feed a conditional logit regression model, where we condition on the patent application.

The average choice set has 122 inventors, corresponding to 7,381 pairs.

The choice set for any patent application contains inventors that have patented recently because they are the people most likely to be in roles related to patenting at the time. Not every inventor patents every year, so we included inventors that appear on any patent application within two years of the focal application (i.e., we create a five-year window around each patent year). Inventors likely spend some amount of time at a firm before patenting and likely do not leave immediately after filing one. Estimating the choice set can result in conflicts between extrapolated ranges between firm changes or address changes of an inventor. To settle conflicting ranges, we used both ranges for firm employment because people can be added to a patent after they have left a company if they played a role in starting that project. We made a clear division in the middle of the application dates while extrapolating addresses. In cases where an even split was not possible, we gave preference to the earlier address as a convention.

Evaluation of Choice Set

We validate the use of a five-year window using self-reported LinkedIn employment ranges for a subset of inventors. In order to do this, we create a restricted set of 456 inventors for whom we could code much more accurate employment date ranges using hand-collected data from their public profiles on LinkedIn. The LinkedIn inventors' self-reported employment ranges inform our extrapolated ranges as well as restrict the sample in further tests to show that our estimates are reasonable in the population studied. Apart from estimating employment ranges, we use self-reported employment information to increase confidence that inventors are employed at the firms and are not external

entities like consultants or doctors that would have different collaboration processes. We expand on how we use LinkedIn data directly in a later section.

To evaluate the appropriateness of the estimated five-year employment window, we compare various ranges of window duration to employment ranges and count years where the estimate matches LinkedIn and those where it does not. We code a "false positive" when the estimated range predicted employment in a year for an inventor but LinkedIn reports that the firm did not employ them at that time. False positives may inflate the significance of results, so we use this information to choose an estimate that minimizes it. We also consider total classification error, both false positives and false negatives, since minimizing false positives alone should result in the smallest possible set, removing an unreasonable amount of data.

Estimated ranges using a five-year window produce the best balance of minimizing both the total classification error (12.2%) and false positives (23.2%) while maintaining a reasonable amount of usable data. A three-year window had a larger total classification error (13.3%), a smaller false positive rate (19.2%), but introduced a 33% reduction in true positives compared to a five-year window. The reduction of true positives indicated that we would discard a large amount of reasonable data with a three-year window for an only marginal improvement of false positives. A seven-year window had a lower total classification error (11.8%) but introduced significantly more false positives (26.2%) than the five-year window.

We expect a certain number of false positives because inventors are often added to patent applications after they leave the firm if they played a role in the development of

that invention. We examine how many inventors are included in patent applications at a date after they claim to have left employment with that firm to determine a reasonable level of false positives. Of all inventor-application observations that have self-reported LinkedIn employment ranges, we find that 19% take place at a time outside of self-reported employment. That so many inventors are included on patents assigned to a company after they are no longer employed at that company indicates that our use of a range to estimate employment is appropriate (rather than using LinkedIn employment directly for all tests) and that our "false positive" rate at 23.2% is not unreasonably large.

Construction of Independent Variables

We geocode addresses to calculate distances between individuals' residences. We do so by inputting the geographic coordinates into Vincenty's formula, which is a geographical distancing method that models the earth as an oblate spheroid and returns the Euclidean distance between two points over the surface of the earth.

Estimation Model

We run a conditional logit model, conditioning on patent application. All patents filed by a company within the same year share the same choice set of potential inventor pairs. Distance in kilometers between the contemporary addresses of each pair is the focal explanatory variable. Every application needs to have at least one pairing (variable ACTUAL=1 for one or more pair) for that application to be used as a group in the regression. Ultimately, 693 patent applications are dropped due to having no pairing in the data, leaving 1257 applications. The dropped applications are either sole-inventor applications or applications where only one inventor lives within the commuting region.

The 693 dropped applications are still useful, however, by providing employment and residence input into potential pairing groups.

INVENTOR RESIDENTIAL DISTANCE AND LIKELIHOOD OF COLLABORATION

The results in table 3, model 1 provide evidence that residential proximity increases the likelihood of collaboration. For every kilometer a pair of potential inventors live apart from each other, the likelihood of them appearing on the same patent application decreases by about 0.3% (the odds ratio for distance in model 1 is 0.997, 1-0.997 = .003, indicating 0.3% less likely each kilometer further apart). 0.3% per kilometer is a meaningful effect in aggregate. If collaborations affected by proximity are necessary for their inventions to exist, aggregate proximity levels should affect firm and region innovation levels.

Table 3 – Results for the likelihood of two individuals collaborating on a patent application.

Model 1 uses purely estimated potential collaboration pair sets, where two years on either side of a patent application form the estimated range that an individual is employed at a firm. Model 2 restricts counterfactuals by removing pairs with individuals who indicated on LinkedIn that they were not employed by the assignee when the application was filed, leaving other estimates. Model 3 shows only those potential collaborations who meet both the 2-year range and gave positive confirmation of employment on LinkedIn for the year of a patent application.

Results are reported as odds ratios, where a value of one corresponds to no effect, below one a negative effect, and above one a positive effect.

Table 3 - Results for the likelihood of two individuals collaborating on a patent application

Dependent variable = collaboration Model number	1	2	3
Distance between inventor residences in km	0.9974 [0.000] (0.000491)	0.9969 [0.000] (0.000558)	0.9961 [0.000] (0.000995)
Observations (Potential Coinventor Pairs x Applications)	12,853,480	7,904,357	1,951,493
Groups (Number of Applications)	1,257	1,107	733
Average Observations (Pairs) per Group	10,225.5	7,140.3	2,662.3
LR Chi Squared (df=1)	29.62 [0.000]	32.93 [0.000]	16.09 [0.000]

Results reported as an odds ratio; standard errors in parentheses; p-values in square brackets.

Discussion of the size of the effect for a firm

While conditional logit results are in reference to a changing choice set and therefore are not able to be interpreted broadly, we can still interpret from the results that a firm with more concentrated employees may be more collaborative. To illustrate the potential size of an effect for a firm, we note that the mean and standard deviation of the distance between the residences of all pairs is 35.3km and 31.5km. Imagine a firm in a similar setting except with a distribution of employees where the mean distance between them is increased by a half a standard deviation – such a company would have the average pair 15.75 km further apart, resulting in the average pair being (0.3%*15.75) 4.7% less likely to collaborate on any given application. Assuming inventions won't exist without collaborations, such a firm would have about 5% less productive knowledge workers, which would be a strategic concern.

If our finding is valid, this would suggest that fewer collaborations occur when inventors live farther apart. We examine if companies with more spread-out inventors have differing levels of innovation. The conditional logit model used above conditions-out across-firm variation. Therefore, assessing if collaboration differs across firms based on the proximity of their employees' residences provides a means to assess the validity of the results. This comparison uses a source of variation not used to derive the estimates. Two of the three firms (Zimmer and Biomet) have similar mean distances between pairs of individual residences (32.0 km and 33.1 km); whereas, the third (DePuy) has a mean distance about 7.4 kilometers further apart (40.0 km). We calculate the mean collaborations each year for each firm, and to account for different inventor population

sizes, we weight them by the number of potential pairs in each firm each year. A simple means comparison shows that the DePuy's more spread-apart inventor pairs are roughly 17% less collaborative than Zimmer's and Biomet's inventor pairs. With only three firms, we do not make claims about the statistical significance of these differences. Nevertheless, these results are consistent with the interpretation that firms with more spread-apart inventor residences may be less collaborative. Unplanned collaborations may be a significant source of innovation for some firms, and others may be missing out.

Alternate choice set determination

To alleviate concerns that the way in which we constructed the choice set includes potential coinventors who were not employees at the firm at the time, we use data from the public LinkedIn profiles of 456 inventors to create different choice sets. We do this because unreasonable counterfactual pairs could potentially drive the results. Inventors who are not employed within the firm would have a lower likelihood of collaboration with inventors within the firm and they potentially live further away from inventors within the firm.

Model 2 uses the same set of potential pairs as Model 1 but drops pairs that erroneously flagged individuals as working for the assignee company. As a result, the number of applications decreases between models 1 and 2. Model 3 uses only the intersection of self-reported LinkedIn ranges and the two-year estimate ranges. The intersection severely restricts the counterfactual choice set and limits it to only those who choose to make their LinkedIn information publicly searchable, which is less representative than using the population of inventors.

The results are robust to the restricted specifications of Model 2 and 3, with the effect of distance maintaining a high confidence level and suggesting an increase in magnitude from Model 1. As a result, we find that the results are not sensitive to these alternative approaches to forming the choice set.

Alternative explanations

We examine multiple alternative explanations that might lead to a spurious relationship between the proximity of inventors' residences and patenting.

Homophily or prior association

One concern is that employees may have geographically clustered residences because of homophily or prior relationships, and this homophily or shard personal history drives collaboration rather than proximity. We assess if homophily and prior association leads to the previous results in four ways.

Prior organizational affiliation. The first indicator of homophily we examine is prior organizational affiliation. For example, two people who are alumni from Clemson University in South Carolina might be more like each other than someone from Case Western Reserve University in Ohio. They may appreciate the same nearby amenities or value an intangible feel of a certain neighborhood because of their shared history. They may also have higher awareness of opportunities in a neighborhood due to collegiate connections. It might be this similarity (or alumni connection) that causes two people to work together and also live near each other. People who have worked for the same prior employer may also share such similarities, so we consider both educational and employment histories as shared organizational affiliation.

Table 4 – Results for proximity determined by prior associations

If the pair is associated through a prior organization, the variable Association takes the value of 1; 10% of pairs are associated. Each pair appears only once in model 4, where the distance between residences is shown not to be predicted by prior association. Model 5 uses the 262 moves to the area by inventors during the period of the study, the alternate person in each pair represents potential influence on moving location. Model 5 uses all people in the area at the time of move in the sample. The Association variable shows positive correlations with distance, indicating that new people tend to live away from those with shared organizational history, rather than closer.

Table 4 - Results for proximity determined by prior associations

Dependent variable = distance (km) Model number	4	5
Association = 1	0.0256 [0.975] (0.7773)	5.5150 [0.000] (0.6015)
Constant	34.396 [0.000] (0.2466)	33.035 [0.000] (0.1170)
Observations (coinventor pairs)	15,567	80,960
F (df = 1 for all models)	0.00	84.08
Prob > F	0.9748	0.0000

Robust standard errors in parentheses; p-values in square brackets.

To test if there is significant residential clustering by organizational affiliation, we return to the self-reported career and academic histories from LinkedIn. If there is any clustering by prior organizational affiliation, the distances between pairs of residences of those individuals should be lower on average when compared to all other pairs. We regress shared-prior affiliation on the residential distance between all pairs of individuals. If pairs with prior affiliations have residences that are closer together, the coefficient of the organizational affiliation variable should take a negative value. We find that about 10% of pairs in the subset of the sample with career histories had some prior affiliation through either university affiliation or prior employer, and indicate each with a binary variable that takes the value of 1 when there is affiliation. Model 4 in Table 4 shows no

association between prior affiliation and distance between residences.

Newcomers moving to area. The previous approach evaluates one static arrangement of residences. To further assess if prior affiliation drives the results, we examine potential influence on the initial residential location choices of newcomers to the area. As stated in the description of the region, most of the knowledge workers in the region moved there from outside due to the lack of local education options. We find that to be true in the subsample with career histories described above (only 6 of the 453 inventors with career histories that we could evaluate seemed to originally be from the Warsaw area). From the career histories, we know the year in which an inventor moved to the region, regardless of patenting activity. Using this information, we test if people moving to the region live closer to people that they may already know or are associated.

The logic behind this test rests on the observation that if newcomers to the area are influenced by other inventors that they may collaborate with, newcomer location choice will reflect this. If this occurs, we expect newcomers' first residential choices are closer to those with potential prior connections. Using distance between each pair as the dependent variable, prior connections should be negatively related to distance. This test is different from the cross-section reported earlier in that this test takes the perspective of the individual mover and each unique year one can move to the area presents a different layout of possible influence.

Because we have residential information for 2001-2015, we restrict our analysis to that period. We assign the first address from the application address data as every inventor's first address upon their move. 262 inventors moved to the area during the

period of the study, while 191 were already in the Warsaw area by 2001. These 191 potentially influenced the 262 that moved during the period of the study, and those of the 262 that moved earlier in the period potentially influenced those who moved after.

We use a set of all pairs of any known established inventor in the region and the newcomer. The benefit of considering any inventor in the region rather than only within firm inventor is that prior connections may influence residential location regardless of hiring firm. We use the shared organizational affiliation indicator from the previous test.

Table 4, model 5 presents the results for the newcomer moving test. The results are not consistent with this alternative explanation because they show that people avoid those they may already know by an additional 5.5 kilometers when moving to the area.[5]

Frequency of collaboration. Although we do not find evidence of residential clustering and prior association, some people with prior associations and who live near each other may be more frequent collaborators. Should this be the case, then prior association could still lead to the observed relationship between residential proximity and patenting. We examine this possibility by including affiliation indicators and their interaction with residential distance to predict collaboration. This test uses a conditional logit model as in the main effect test but uses the subset of the sample with career histories. If affiliation drives the result, the odds ratio of the interaction of distance and should be significantly lower than one and the main effect of distance should diminish. Models 6-7 in Table 5 present the results of this test. In model 6 we show that prior affiliation increases the rate of collaboration between two inventors by about 50%. The

[5] Results are consistent with other specifications restricted to only within-firm area residents or residents who have had overlapping time at prior organizations.

interaction between affiliation and distance is the variable of interest to see if some proximate affiliated inventors are driving the effect. We see the odds ratio of the interaction term does not significantly differ from zero, and the main effect of distance alone is not diminished[6]. This test does not support the alternative explanation.

[6] To test that the interaction result is not an artifact of a nonlinear model, we tested the same specification with a linear probability model (LPM). We find the magnitude of the main effect is unchanged, but that the interaction term is negative and significant at the 0.05 level. This significant LPM result does not change our interpretation because the main effect is not reduced.

Table 5 – Results for the likelihood of collaboration with homophily measures and proximity

Model 6 includes an indicator variable of prior association through education or employment and shows that a pair with prior association is about 50% more likely to collaborate than those pairs without prior association.

Model 7 includes an interaction term between association and distance to verify that associated pairs are not driving the main effect of residential distance. The interaction term does not have statistical significance and the distance term does not meaningfully reduce in magnitude, so associated pairs do not seem to be driving the main result. Models 6 and 7 use a reduced sample of those pairs with known prior affiliation histories.

Model 8 includes the difference in block group household incomes of the addresses from the 2010 census. There are fewer observations in model 5 due to a few addresses without census info in 2010, otherwise the full set of inventors is used.

All results are reported as odds ratios, where a value of one corresponds to no effect, below one a negative effect, and above one a positive effect.

Table 5 - Results for the likelihood of collaboration with homophily measures and proximity

Dependent variable = collaboration Model number	6	7	8
Distance (km)	0.9934 [0.000] (0.0011)	0.9939 [0.000] (0.0011)	0.9984 [0.002] (0.000515)
Associated pair = 1	1.4981 [0.000] (0.1158)	1.6342 [0.000] (0.1858)	
Distance X Associated		0.9969 [0.309] (0.0030)	
Difference in block HH income (in '000s)			0.9957 [0.000] (0.000759)
Observations (Coinventor Pairs x Applications)	1,592,898	1,592,898	12,839,086
Groups (Number of Applications)	659	659	1,257
Average Observations (Pairs) per Group	2417.1	2417.1	10,214.1
LR Chi Squared (df = 2 model 6, df=3 model 7)	66.39 [0.0000]	67.45 [0.0000]	61.62 [0.0000]

Standard errors in parentheses; p-values in square brackets.

Socioeconomic clustering. Socioeconomic clustering of residences presents another competing explanation where homophily may drive the observed main result. If senior inventors earn more than junior inventors, tend to work together more often, and live in the same (i.e., more expensive) neighborhoods, this could produce our central finding. Similar clustering would occur with junior-level employees in less expensive neighborhoods. For this reason, we are concerned that our finding could reflect status in the organization and associated income level.

To test if socioeconomic status drives the result, we estimate the socioeconomic status of inventors by matching their residential address with the census household income estimate for their residence's block group. Block groups provide the highest resolution of publicly available income estimates. We include the difference between estimated household incomes for each inventor pair as an explanatory variable. The result in table 3, model 8 shows that there is a small effect of socioeconomic status and a minor reduction in the magnitude of the main effect of residential proximity when a socioeconomic difference is included. Socioeconomic clustering may be occurring, then, to some degree, but it does not eliminate the effect of proximity. The results show that for every thousand-dollar difference in household income, the likelihood of a pair collaborating goes down by about 0.4% (1-.9957 = .0043). To calibrate this, consider that a senior engineer might have a yearly salary of $130,000, and a junior engineer might earn $55,000. Using our findings, the likelihood of a senior engineer collaborating with a junior engineer is about 32% lower than their likelihood of collaborating with another senior engineer (130-55 = 75, 75*.0043 = 32%).

The preceding sections demonstrate that while a seemingly plausible alternative explanation, homophily – at least in all the ways we can measure it – does not appear to be driving the between residential proximity and collaboration.

Inventors moving within sample. To further demonstrate robustness of our demonstrated effect of residential proximity leading to collaborations – as opposed to the reverse causation – we examines individuals' moves within the region during the time of the study. Specifically, we test same-inventor pairs in which there is a change in residential distances over time (i.e., at least one of them moved). If residential proximity has no effect, we would expect to find no difference in collaborations of pairs before and after any moves within the region. However, if inventors collaborate more when they live nearer each other versus when they live farther apart, this supports our argument that residential proximity matters to collaborations. The fixed effect on the pair in this test effectively controls for homophily by having no within-pair differences on any observable or unobservable characteristics before and after a move.

Moves within the region are not random, and we could not identify any instruments in the region or sample to claim that some moves are somewhat random (i.e., no neighborhoods became more or less attractive because of development projects, there were no natural events, administrative boundary changes, or anything else we could identify that we could argue could exogenously influence home locations). If future collaborators influence future residential locations their influence would be an endogeneity concern for this test. But we believe future collaborators are not major influences on home location compared to more exogenous factors like price and

availability, so this endogeneity concern is limited.

To separate before and after effects within-pair, we create two observations per pair of inventors: one before and one after either of the inventors move within the region. For the first test, we assign a binary value to whether a pair collaborates in the before or after period; pairs that either never appear together on a patent or appear together on a patent in both before and after periods are dropped from analysis. Table 6 model 9 presents the results of a within effect of pairwise distance, where distance predicts collaboration. Each pair that experiences a shift in proximity is 1% more likely to collaborate per kilometer they are closer, regardless if before or after the shift. This test shows that some pairs stop working with people they used to live near, and they also may start working with people after they become more proximate.[7]

Table 6 – Results for the likelihood of collaboration, fixed effect on pairs with movers

Model 9 uses a logit estimator to model the likelihood of collaboration changes within pair when the distance between the pair changes. This model only includes two observations per pair: before and after the move. Results show that a pair is 1% more likely to collaborate per kilometer they are closer, regardless if before or after the shift in proximity.

Table 6 - Results for the likelihood of collaboration, fixed effect on pairs with movers

Dependent variable = collaboration	
Model number	9
Distance (km)	-0.0117 [0.001] (0.00340)
Observations (coinventor pairs X period)	1,326
Observations per pair	2
LR Chi Squared (df = 1)	13.50 [0.0002]

Robust standard errors in parentheses; p-values in square brackets.

[7] A count model comparing the frequency of collaboration in each time period gives consistent results, but is not reported.

Central concentration of residences

Figure 1 exhibits that inventor residences tend to concentrate toward the center of Warsaw, Indiana. Inventors who live in the center live closer to each other; however, they also live closer to their workplace. It might be that a pair's proximity to their facility increases the likelihood of collaboration, because they may be more committed to their work, or spend less time commuting to work. To test if proximity to facilities drives the result, we measure the average distance of each pair of potential inventors to their facility. If both live near the facility, then the value will be low. If shared proximity to the facility is driving the result, the facility proximity variable should explain collaborations rather than the proximity of residences. Table 7, model 10 adds the facility proximity variable to the specification in model 1. The coefficient estimate of this variable is not statistically significant ($p=0.315$). Proximity between residences, rather than proximity to the facility, predicts collaboration. As a result, this alternative explanation appears not to be material in our setting.

Table 7 – Results for the likelihood of collaboration with the distance to the center of town

Model 10 includes the average distance of the pair to the center of town (near where all R&D headquarters are located).

All results are reported as odds ratios, where a value of one corresponds to no effect, below one a negative effect, and above one a positive effect.

Table 7 - Results for the likelihood of collaboration with the distance to the center of town

Dependent variable = collaboration	
Model number	10
Distance between inventor residences in km	0.9967 [0.000] (0.000845)
Average distance of pair to town center	1.0014 [0.315] (0.00143)
Observations (Coinventor Pairs x Applications)	12,853,480
Groups (Number of Applications)	1,257
Average Observations (Pairs) per Group	10,225.5
LR Chi Squared (df = 2)	30.61 [0.0000]

Standard errors in parentheses; p-values in square brackets.

Collaboration, gender, and proximity

Homophily and social factors are shown to influence innovative output and collaboration networks (Almeida et al. 2015, McPherson et al. 2001, Uzzi 1997). We do not have detailed information about the inventors' demographics except through their names. As a result, we use inventor's names to assess their gender.

We hand-code each inventor as male or female (and cross-reference with LinkedIn when available) to explore a gender-based social cost effect. Of 702 inventors, we coded 610 as male and 92 as female. We code pairs as same gender (i.e., MM or FF) or across gender (i.e., FM). Gender homophily among same-gender pairs may lower collaboration costs between that pair (Ruef et al. 2003). We investigate this possibility.

Table 8 – results for the likelihood of two individuals collaborating on a patent application with gender information.

Model 11 includes an indicator for the pair being the same gender. Model 12 includes an interaction term between the same gender indicator and distance of residences.

All results are reported as odds ratios, where a value of one corresponds to no effect, below one a negative effect, and above one a positive effect.

Table 8 - Results for the likelihood of two individuals collaborating on a patent application with gender information

Model number	11	12	13
Distance between inventor residences in km	0.9974 [0.000] (0.000491)	1.0009 [0.388] (0.001075)	1.0009 [0.388] (0.001075)
Same gender indicator (pair is MM or FF)	1.2839 [0.000] (0.0507)	1.4858 [0.000] (0.0860)	
Same gender X distance		0.9957 [0.000] (0.001202)	
Pair is MM			1.4880 [0.000] (0.0862)
Pair is FF			1.3702 [0.152] (0.3009)
MM X distance			0.9957 [0.000] (0.001209)
FF X distance			0.9884 [0.053] (0.001202)
Observations (Coinventor Pairs x Applications)	12,853,480	12,853,480	12,853,480
Groups (Number of Applications)	1,257	1,257	1,257
Average Observations (Pairs) per Group	10,225.5	10,225.5	10,225.5
Degrees of freedom	2	3	5
LR Chi Squared	71.99 [0.0000]	84.34 [0.0000]	90.26 [0.0000]

If same-gender pairs collaborate more often than expected given the gender distribution of their options, then this is evidence that homophily may have decreased collaboration costs between some same-gender inventors. We test this by adding a same-gender indicator to each pair to our main model. Model 11 in table 8 shows that there is a base rate of same-gender pairs collaborating about 28% more often. There is, then, a tendency for same-gendered people to work together. Because our main effect of proximity is unaffected when we add the same-gender indicator to the specification suggests that same-gendered pairs do not have a pattern of living near one another (i.e., gender is unrelated to the distribution of residences).

We expect serendipitous encounters to lower the costs of collaboration if the encounter results in an actual social interaction. Homophily between two similar people should increase the chances that an encounter escalates to a social interaction. Social interactions that lower costs of collaboration may then be partially contingent on homophily. Because the chances of such encounters decrease with distance, we create an interaction term between the same-gender indicator and distance in model 12. We code interaction as zero if the same-gender indicator is zero (indicating an FM pair): otherwise, it takes the value of the distance between the pair.

The estimates in model 12 show a strong proximity effect for same-gendered pairs, with the original distance variable effectively having no influence. These results demonstrate that same-gender pairs drive the distance result in the previous analyses. In other words, the proximity of residences only matters for same-gendered pairs. To test that the result is not an artifact of the interaction term in a nonlinear model, we tested the

same specification with a linear probability model (LPM). We find consistent results with respect to significance levels and direction of effect. The magnitude of effect is not given as an odds ratio, but is consistent with the difference in odds in the conditional logit model once transformed by dividing by the base rate odds (the odds of any pair collaborating per application is roughly 1/2700).

To further investigate this finding, we break the same-gender variable into FF and MM pairs. We represent these results in model 13. The coefficient estimates demonstrate that the proximity effect is primarily between MM pairs. This is not surprising because MM pairs constitute the bulk of the sample. The relatively smaller proportion of female inventors to male inventors is likely driving down the significance of FF pairs. We find an effect of proximity between FF pairs (at $p<.053$ significance level) at a magnitude close to that of the MM pairs.

DISCUSSION AND CONCLUSION

Our work extends existing research that investigates how employee proximity within the workplace affects collaboration by demonstrating that the proximity of employees' residences affects collaboration. We identify that proximity of people outside the workplace is a strategic concern and show that high-resolution data is necessary to appropriately evaluate the mechanisms through which proximity affects collaboration outside the workplace.

Proximity outside the workplace is a strategic concern

Knowledge effects from proximity between people outside the workplace are of strategically interest for the following reasons. First, similar phenomenon to what we

describe is argued to be responsible for knowledge spillovers between geographically proximate firms. Interactions between people that lead to knowledge spillovers between firms necessarily take place outside of a shared workplace. Our results suggest that proximity between the employees of firms, not only proximity between firms themselves, should predict knowledge spillovers between those firms. The mechanisms of proximity in a between-firm scenario are subject to greater costs due to conflicting organizational affiliations and will therefore not likely be as measurable as what we study. Nevertheless, proximity-reduced collaboration costs should result in the increased sharing of knowledge between employees of different firms. Although we are unable to directly measure spillovers, we are able to isolate the effect of proximity outside of a shared facility, which is a necessary step toward characterizing how spillovers occur.

Second, another reason proximity outside a workplace matters to firms is that the residences of employees are uncontrolled by firms. As such, they are an external influence on collaboration networks, which shape the internal direction of innovation development. The communications and collaborations we describe are unplanned by firms, and outcomes of these collaborations may be either good or bad for the firms. We only witness beneficial collaborations chosen to be developed and patented, but patents likely represent only a small portion of the potential innovations initiated this way. Undirected activities may instead result in the organic development of a new technology that undermines existing firm competencies or resources; any development of this type of technology may be stopped and hidden by the firm. Better understanding these mechanisms outside the workplace is important. Although firms cannot control outside

interaction, they can undertake actions that take into account these effects. For example, firms may consider the potential effects of employee residential distribution when evaluating facility location choices. Firms might also choose to increase employee services and programs (e.g., in-house daycare, a company-sponsored garden, sports clubs) at facilities in areas especially at risk of negative spillovers.

Empirical considerations support that residential proximity matters

The orthopedic cluster in Northern Indiana is uniquely isolated and concentrated, so organizational affiliations are not potentially driving the proximity effect directly or indirectly. There is no direct effect of organizational affiliation because we only consider collaborations between people employed at the same facilities. Our robustness check with self-reported employment ranges from LinkedIn increases our confidence that each inventor is associated with a specific facility during their employment. Furthermore, there is no indirect effect of prior organizational affiliations because of the concentration (and otherwise lack) of relevant facilities. Although our research design may be useful in other regions, most innovating regions would likely have indirect effects of prior organizations on residences that could drive any proximity effect. We believe we identify a uniquely well-suited region and accompanying method for the study of collaboration mechanisms where facility locations and organizational affiliation are controlled for. Therefore, our study is able to capture proximity mechanisms like serendipitous interactions outside of the workplace in a more plausible way than other research designs.

We attempt to isolate interactions outside of work by only testing for an effect of residential proximity between people who work at the same facility to identify outside-of-

the-workplace effects, something that is not commonly done. Geographic proximity of inventors on a regional scale should correlate with their shared workplaces, so we should consider workplace locations if we want to learn about collaborations or potential knowledge sharing mechanisms that happen outside the workplace. Controlling for shared membership of an organization, as is sometimes done, does not remove a proximity effect caused by facility in multi-facility organizations. The concern that a shared facility drives the apparent effect of proximity is overcome by considering potential collaborators within each facility, as in our research design.

The search and execution costs of collaboration we describe can be lowered by proximity through various mechanisms. Because these costs and mechanisms are different, they do not all have the same relationship with the scale of proximity. Mechanisms that lower search costs and social execution costs require that people see and encounter one another in regular day to day activities, which can only be meaningfully considered with a high-resolution measure of proximity. The mechanism of repeatedly seeing someone requires that preferred third locations are in common, e.g., they have overlapping preferred grocery stores, cafes, bars, dog parks, school-related functions, etc. Given how the many options of these third locations – even in smaller cities – a personal preference for one is likely highly determined by proximity to the residence of the person. This person likely considers travel time to each similar option: if given a choice to go to the same grocery store chain location either 5 minutes or 10 minutes away, they would seldom choose the farther option, all else equal. The difference between 5 and 10 minutes is a short geographic distance, but a meaningful distance for

these decisions. This resolution is lost and distorted when researchers use regional measures for proximity, so we use data with the granularity that matches the scale of the mechanisms we research.

We use the distance between residences to measure proximity, a measure of distance with the granular resolution a person uses to make the type of decision of where to spend time outside of work and the home. The distance between residences is a much more fine-grained measure than an indicator of being within the same commuting region or the distance between centroids of commuting regions. Researchers often use regional measures for proximity as a necessity due to data limitations, but regional measures are unreasonably coarse when considering mechanisms that lower search and social costs like serendipitous encounters. Using a measure appropriate to an individual's experience allows us to employ a search and social cost argument to explain our results with plausibility.

Gender and proximity

Our findings concerning gender suggest that cross-gender social costs persist even when potential collaborators live proximate to each other. Gendered social costs may prevent many organic, unplanned collaborations between male and female inventors. Our interpretation is consistent with the idea that the firm does not plan the collaborations influenced by residential proximity, because the social costs to collaboration should not be salient between two people of different genders assigned to collaborate by the firm.

The low number of women in the sample may be responsible for the finding that there is no baseline homophily effect between female pairs. However, no baseline FF

effect is consistent with literature stating that women do not tend to work together as much as men (Benenson et al. 2009). Research shows that women will sometimes avoid working with each other, especially in male-dominated industries, due to fears of reinforcing negative stereotypes (Duguid 2011). Given the relative number of female inventors to males in the sample, this is a male-dominated industry. Women avoiding each other may be weakening an underlying a separate homophily effect, though not to a high degree since there is no evidence of a negative effect of FF pairs.

Limitations & Conclusion

Although there are many advantages of our study's research design, we recognize that our efforts to demonstrate causal identification are limited in that there is no exogenous shock or quasi-experimental research design. A perfect study would have random assignment of residential locations so we know they are unrelated to future collaborators, but this is impossible. In spite of limitations, we provide plausible findings that residential proximity matters for collaborations. Other limitations include generalizability to other areas. Cultural norms are likely influential in this setting and should not be ignored when attempting to extrapolate findings more broadly. There may be especially strong regional cultural norms at play in this setting that may not exist elsewhere, especially considering the rural setting.

The firm does not plan all collaborations between knowledge workers at a shared workplace, but instead, collaborations result from a loose organic search process. The engineering manager at one of the firms studied that we interviewed described an example of such a process. This example highlighted the importance of an individual

searcher's knowledge of others and the perceived knowledge held by these others: the searcher will likely not invite someone from a different department if they don't have a personal relationship. The example also described a situation where social fears are potentially salient: the searcher will not invite outsiders if they believe there will be social costs to their involvement, like might come from making their department look bad. The example was about a brainstorming exercise on a divisional level, but it is easy to imagine similar costs in common search processes on an individual to individual level.

Frequent observation and serendipitous encounters outside of the workplace can lower these costs, increasing the odds of collaboration between two people. It may seem like we rarely encounter potential collaborators who live near us, but we do see them much more frequently than we encounter those who live further away. These interactions add up, and we can see an effect of proximity outside the workplace.

Our findings show that, in general, proximity outside the workplace can result in strategically meaningful communications. While communication between individuals at the same facility is fundamentally different than between individuals at different firms, the costs of such communications are similar. This conclusion is interesting to the industry cluster, knowledge spillover, and agglomeration literature, which desire a clearer understanding of the mechanisms of agglomeration economies.

CHAPTER 3: THE INFLUENZA INFLUENCE: THE FLU DIVERTS COLLABORATOR CHOICE AWAY FROM LOCALS

CHAPTER SUMMARY

I introduce local flu prevalence as an exogenous influence on chance encounters during the knowledge-seeking process of inventors. I argue that a social avoidance response to a disease epidemic – a phenomenon well-established in epidemiology literature – decreases the likelihood of chance encounters between geographically proximate people. I establish a measure of local flu prevalence during the early collaboration process as a proxy for lower potential chance encounters affecting an innovation. Supporting the validation of this measure, I find that collaborations more likely formed during average flu prevalence are 1.1% less often with local collaborators than if there was no flu. I develop and test corollaries that support that the effect is consistent with individual choice and opportunity for engaging with non-locals, giving additional evidence that chance encounters are responsible for the relative decline of local collaborators.

INTRODUCTION

To continue my research on proximity, I narrow my focus on the mechanism of chance encounters. As I will expand in chapter 3, chance encounters are especially interesting theoretically. Chance encounters primarily act on collaborator choice by broadening the awareness of others and their knowledge, which should increase the novelty of innovative output of the collaborations that they help form.

In the current chapter, however, I am interested in demonstrating plausible empirical evidence of the effects of chance encounters as they occur in commonplace settings. To best complement current literature in the area, I focus on common natural settings that researchers have not artificially manipulated, and I aim to evaluate chance encounters associated with proximity as a phenomenon apart from proximity. However, I can only study the effect of chance encounters once I am convinced that I can create a research design where chance encounters can vary exogenously *and* where the rest of the endogenous collaborator selection process is unaffected. In this chapter, I introduce a measure using the flu that proxies for a level of potential chance encounter that I argue varies chance encounters more than any other meaningful influence. I then describe how I validate that measure before its use in the following chapter.

GOALS AND EMPIRICAL CHALLENGES

There is not yet large-scale evidence of chance encounters affecting collaboration. Instead, there is evidence that changing proximity has outcomes that chance encounters might best explain. Proximity can indeed increase awareness of others and their knowledge through chance encounters, but proximity also brings other mechanisms with

other effects on collaborations, complicating any study that varies proximity between people.

Furthermore, studies do not generally map the mechanisms of proximity to innovation or collaboration outcomes (Andrews 2020 & Catalini 2018 are notable exceptions). Some empirical studies of proximity measure earlier stages in the process, like conversations, but do not show that these activities lead to meaningful outcomes. Other studies observe outcomes and must guess at the mechanisms that caused that outcome, between all the other proximity mechanisms. The intermediate connections necessary to complete a view of the process remain abstract.

As I describe in my introduction to this dissertation, proximity sets the stage for choices to be made, and those choices determine very theoretically different and interesting outcomes – experimental studies with direct manipulation of subjects removes that choice and with it, a key theoretical feature of proximity's effect on collaborations. I see two goals for a successful research design to study the mechanism of chance encounters on collaborations that I describe in this section.

Goal 1: Observe the effect of chance encounters in a *natural setting*

The design of a study to evaluate chance encounters presents theoretical as well as empirical challenges. The first goal of a effective treatment in a research design that studies chance encounters as a part of proximity is that the research design must theoretically reflect the natural process that people make, because the choices by people is a key feature of any effect of proximity. Not only can researchers not manipulate chance encounters directly in a controlled setting, but it is also a theoretical impossibility

to manipulate chance encounters in a setting that ultimately relies on individual choice determining collaborators. People must make the choice themselves to interact and collaborate. If a study's manipulation simply tells Alex to talk to Blake, that removes the choice from both to engage with each other, and therefore removes a theoretically integral component of the effect of proximity. Proximity sets the stage and gives options, but if we try to understand how proximity works, researchers should not make choices for people and their interactions. While assigning random people to interact might be tempting for research design purposes, random choices do not reflect either reality or the nature of proximity.

Goal 2: Separate proximity from chance encounters

The other primary goal for this chapter is to plausibly isolate chance encounters as a mechanism apart from the broader concept of proximity, empirically. Most prior studies interested in chance encounters study them by varying the proximity between people, where many of the mechanisms and intermediate outcomes of proximity might also change at once (Catalini 2018, Kabo et al. 2014, Lane et al. 2020, Lee 2019, Roche et al. 2020). When all mechanisms may move with proximity, these researcher – including myself in my own first chapter of this dissertation – can only inductively argue from results what mechanism may have had an effect. This omission is because chance encounters are currently unmeasurable themselves, especially in natural and unmanipulated settings. [8]

[8] Ethically, I believe chance encounters should remain unmeasurable in natural settings, as any situation where I can imagine constant measurement of such actions would be extremely invasive. So I believe this will be an issue well into the future and we should not wait for better data to investigate this phenomenon. Ethical issues aside, studies measuring chance encounters directly should be considered in a

Catalini (2018) makes a notable effort at inductively arguing that increasing awareness, through mechanisms such as chance encounters, can lead to more impactful collaboration output. I will describe his approach here, both to highlight what is in my opinion the best empirical example of the study of chance encounters in current literature, but also to demonstrate the potential weaknesses of such approaches when used to isolate mechanisms.

Catalini's 2018 study uses a natural experiment where research labs effectively randomly collocate when they move to new facilities, finding that after they move, some of these newly collocated labs work more often with each other. In the main research design of the study, he varies proximity and with it, who the mechanisms of proximity might be acted on. The inductive argument that chance encounters are likely to increase awareness in some collaborations is only a small part of the paper's overall contribution, but I see it as an important part. Catalini argues that awareness affects the nature of collaborative output by showing that collocated labs generate especially highly cited works if they were labs with especially higher search costs before moving together. The argument goes that higher search costs indicate the disproportionate importance of increasing awareness for these labs, in that they have much more room to grow if they only knew what was out there, as it were.

Higher search costs a priori do not, however, increase the likelihood of chance encounters. The mechanism of chance encounters is theoretically unchanging in the

less-natural state because settings where observation is known people will distort behavior away from natural actions, as in relevant proximity studies with the presence of observers (Torras and Cummings 2020).

setting, and changes in awareness remain entirely abstract (this part of the study invokes variance in the need to be more aware, effectively). Furthermore, the argument that awareness is responsible for the impact – rather than convenience, for example – rests in large part on the nature of the impact itself. More awareness should lead to better output for these labs, and that is the observation. I agree with this logical conclusion, but it is an indirect test that relies on inferring the mechanism from the nature of the output, not testing the mechanism.[9]

Apart from somehow directly manipulating people into chance encounters (but still somehow letting them make their own important choices as if they were not manipulated), a better way to isolate the mechanism of chance enounters is to take a inductive approach as above, with an expected *but theoretically neutral* outcome. A validation of such a "treatment" with a theoretically neutral outcome can infer the mechanism more neutrally than working backward from the expected interesting theoretical relationship. Once established, the treatment can be used as a proxy for testing the mechanisms impact on the more theoretically interesting outcome.

To resolve these goals, I use prevalence of local flu as a proxy for lower chance encounters. I describe this choice below, and how it likely affects chance encounters more than other potential mechanisms. I then validate the use of variance of local flu with the "neutral" outcome of an expected lower local collaboration rate, a choice I also describe below. I develop and test corollaries consistent with testing this approach.

[9] Catalini's study is about proximity and not chance encounters specifically, so this is not a criticism of his approach, but a statement of how much we should read into his findings.

I employ some inferential logic that can push forward our knowledge of chance encounters. But I do not have an ideal research design, as I believe that is an impossible goal for studying chance encounters. Accordingly, with a less-than-ideal research design, however, there is a high burden on the appropriateness of empirical choices, which I explain in the next section.

CONSTRUCTION OF A MEASURE USING THE FLU AS AN EMPIRICAL TOOL

Local flu variance addresses challenges to the study of chance encounters

I use the variation of local flu as a proxy for the variation of potential chance encounters. Local flu prevalence is a new independent variable for literature on strategy and innovation, so I devote the next portion of this paper to demonstrating the validity of this measure. The flu, or more specifically, the social avoidance behavioral response of people to their knowledge and fear of local flu, overcomes many empirical challenges to the study of chance encounters and their effect on innovation, and allows the use of a large set of data to test my hypotheses in the next chapter.

Social avoidance behavioral response to the flu

A major empirical challenge to the study of chance encounters is that chance encounters themselves can't be measured, and certainly not on a large scale. However, economic epidemiology and social epidemiology research have shown that diseases like the flu affect human behavior that can influence potential chance encounters (Ferguson 2007). People engage in various behavioral responses to the fear of disease when they

think that behavior might protect them, including reducing social interaction (Fenichel et al. 2011). Models used to predict the spread of diseases include social avoidance[10] as a factor (e.g., Funk, Salathé, & Jansen, 2010), and social network papers consider social avoidance in response to disease as a factor influencing the reconfiguration of network structure (Gross et al. 2006). Social avoidance behavior driven by the flu, specifically, has been quantified using historical records (Caley et al. 2008) and using modern data (Bayham et al. 2015).

Epidemiology research identifies that social avoidance will occur in a population depending on three conditions: people are aware of the disease, fear catching the disease, and believe that social avoidance will increase their safety. The flu meets these requirements for many people, as the news widely reports the flu when it is present in an area (Young et al. 2008). Furthermore, these reports often contain language meant to appeal to fear to be more sensational (Tannenbaum et al. 2015). Research by Jones and Salathé (2009) shows that fear of seasonal flu is perceived to be high, as high as cancer, heart disease, or pandemic flu (i.e., the "swine flu" pandemic in 2009). Jones and Salathé (2009) find more than half of respondents will shun others they perceive to be sick with the flu, around 20% will avoid gatherings and public places (containing presumably healthy-appearing individuals), but very few will stay home from school or work. The flu is not a huge shock to other aspects of work or life – just potential social interactions,

[10] This term is also known as "social distancing." Because of the COVID-19 pandemic, variants of the term "social distancing" are now commonly known to the public. I use the term "social avoidance" to distinguish a spontaneous and voluntary behavior of an individual to limit in-person social interactions. I do not use it to mean a directive from a government, the requirement to be a certain physical distance from others, to remain at home, or any other of the newer common usages for "social distancing."

making it suitable for study of chance encounters apart from other proximity mechanisms compared to diseases that have much larger impacts on many aspects of life, like COVID-19.

Furthermore, one does not need to engage in social avoidance to experience fewer chance encounters when the flu is prevalent. When anybody engages in social avoidance, they lower the chance of potential chance encounters for anybody near them. The effect of the flu on chance encounters therefore directly affects those who are socially avoidant and a has a potentially large indirect effect on everyone otherwise proximate to them.

Validation of measure: variance in chance encounters

To test the validity of using flu as a proxy for lower potential for chance encounters, I first show that the flu lowers collaboration between proximate inventors. The logic is that only proximate collaborators should be affected by fewer chance encounters, but work and life, as described above, goes on. When potential chance encounters decrease, local collaborations should decrease, conditional on observed productivity, and given total collaborations. Local collaborations are a "theoretically neutral" outcome that I highlight as desirable for a validation test in my statement of goal 2 above.

I begin by describing my independent variable – weeks of local flu during inventive search – and the empirical choices I made in its construction. I go on to describe the dependent variable of the validity test – local collaborators. I then describe the specification of the validity test, results, and outline and describe tests of expected

corollaries and robustness checks that support my interpretation that chance encounters drive the results. I then proceed to describe the use of the independent variable in hypothesis tests, continuing the paper.

Cumulative weeks of local flu prior to a patent application, the independent variable

The flu provides meaningful variance over time with magnitudes that should affect search. Figure 3 shows the variance in flu between states in the U.S. during the dates used for this study. Each month is given a value of the total weeks of "widespread" flu in that state during the prior twelve months to an invention patent application, the time I consider the inventive search period and early collaboration process. (I will discuss the choice of a one-year window and its timing in more depth later in this section.) "Widespread" is the term from the United States Center for Disease Control, or CDC, indicating the maximum reported level of flu in a state. The "widespread" measure is appropriate for this setting because people are aware of the flu through news reports, which usually indicate if the flu is "widespread" in a state. The designation "widespread" does not indicate the meeting of an overall numeric threshold (that could concentrate in one area) but instead indicates that there are outbreaks of flu cases covering more than half of the state's regions[11].

Figure 3 demonstrates that the variance in the independent variable is significant and meaningful, and potentially different than what a reader might expect given their own experiences with the flu. Cumulative local flu exposure over twelve months can be quite

[11] https://www.cdc.gov/flu/weekly/overview.htm

long or can be inconsequential. While the onset of flu is generally seasonal and usually in mid-winter, the actual start varies significantly year to year and between geographies, which also influences this variable (this figure does not show timing).

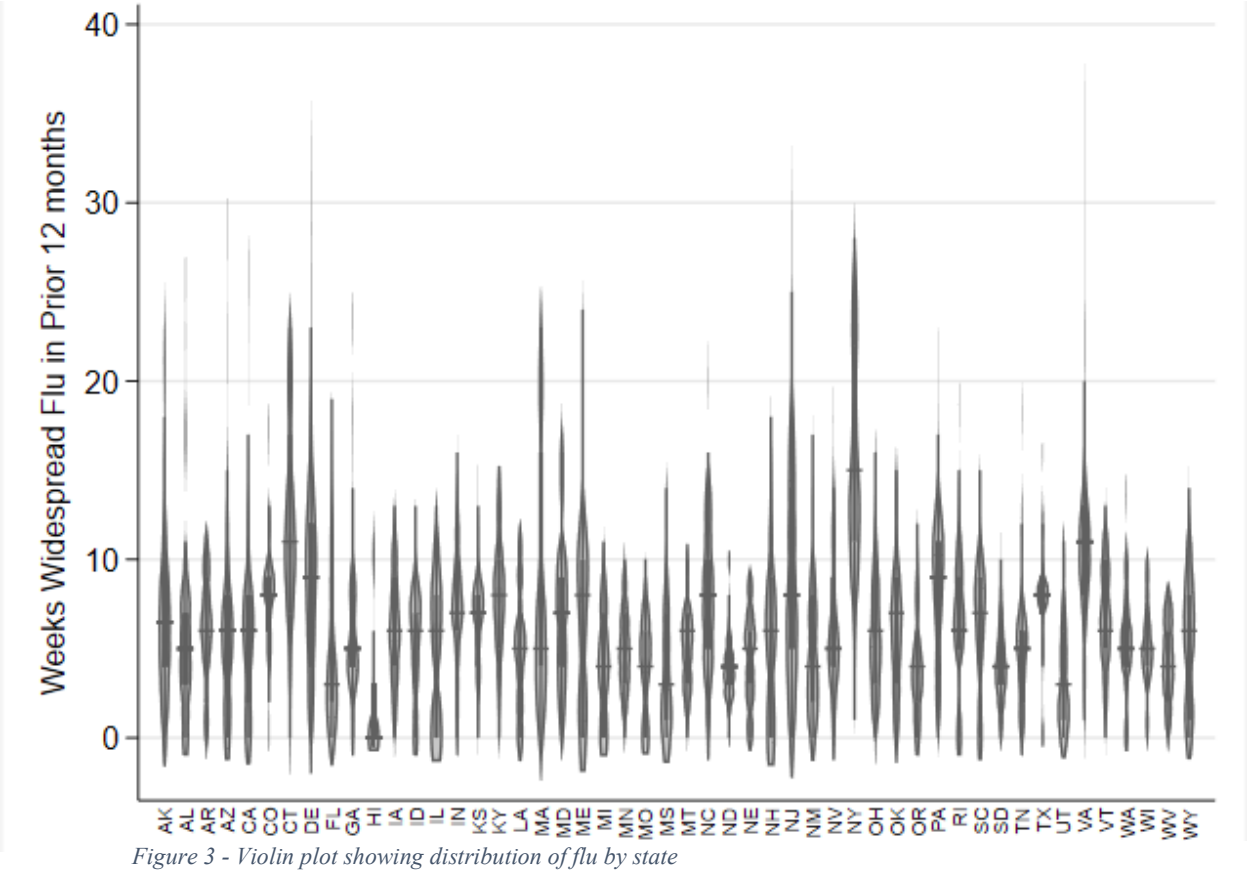

Figure 3 - Violin plot showing distribution of flu by state

Figure 3 – Violin plot showing distribution of flu by state. The data represents the range of the used sample, August 2004 – December 2015, where each month corresponds to one data point for each state. The data represents flu exposure (weeks where flu is considered "widespread," the highest level from the CDC) that state experienced in the twelve months prior to the given month.

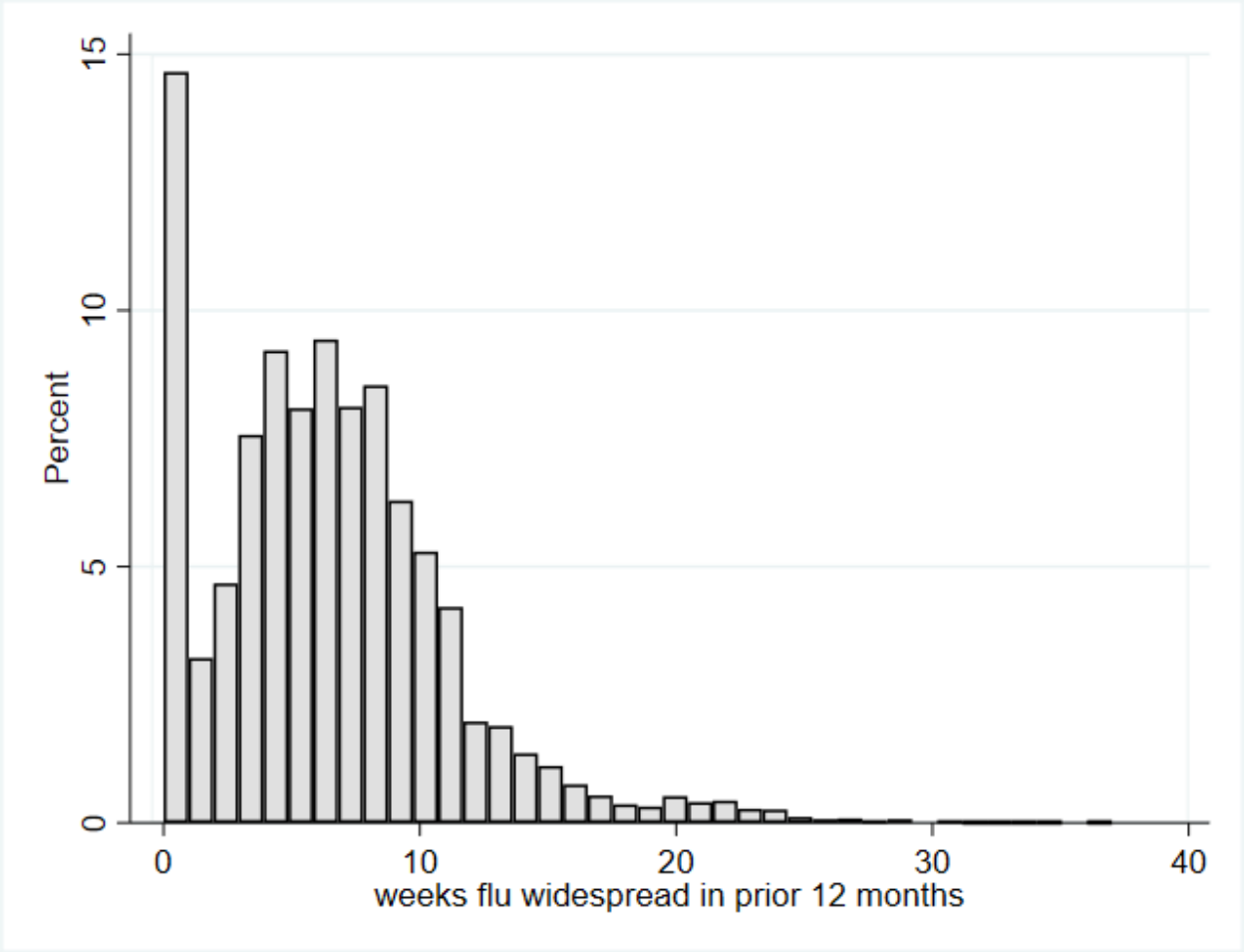

Figure 4 - Histogram showing flu distribution in sample

Figure 4 – Histogram showing distribution of weeks of flu in prior twelve months in sample, with pooled state values. Over 14% of state-month combinations in the sample show zero flu exposure in prior 12 months. The average weeks of prior flu in 12 months is 6.2, with a standard deviation of 4.9 weeks. The maximum value is 37 weeks of flu in the prior 12 months.

Figure 4 shows a histogram of the same values as Figure 3 with all state data pooled to give a more general overview of the variance in flu duration. Over 14% of state-month combinations in the sample show zero flu exposure in the prior 12 months. The average weeks of prior flu in 12 months is 6.2, with a standard deviation of 4.9 weeks. The maximum value is 37 weeks of flu in the prior 12 months.

I illustrate two examples of the construction of the independent variable in Figure 5. The timelines in Figure 5 show example inventors in either Indiana or Michigan with three patent applications, each with varying exposure to the flu. The time windows used to represent the inventive search process are three boxes labeled A, B, and C. Vertical lines represent weeks of widespread flu in that state; the sum of weeks of flu contained in those boxes makes the independent variable. The bold dashed vertical lines represent the patent application date. The pink area plot represents a continuously calculated independent variable for all dates in that geography, the cumulative prior flu that an invention applied for on that date would have been exposed to during its inventive search period. The Indiana example timeline shows patent A was likely influenced to a large relative degree by social distancing due to the flu, with a value of 15 weeks. Patent B in Indiana was potentially influenced less, for ten weeks. Patent C in Indiana was not influenced by the flu at all. I include a Michigan timeline for contrast with the same patent application dates. Data for this example is taken from two similarly sized and geographically proximate states, Indiana and Michigan, to represent the ordinary type of variance in the sample.

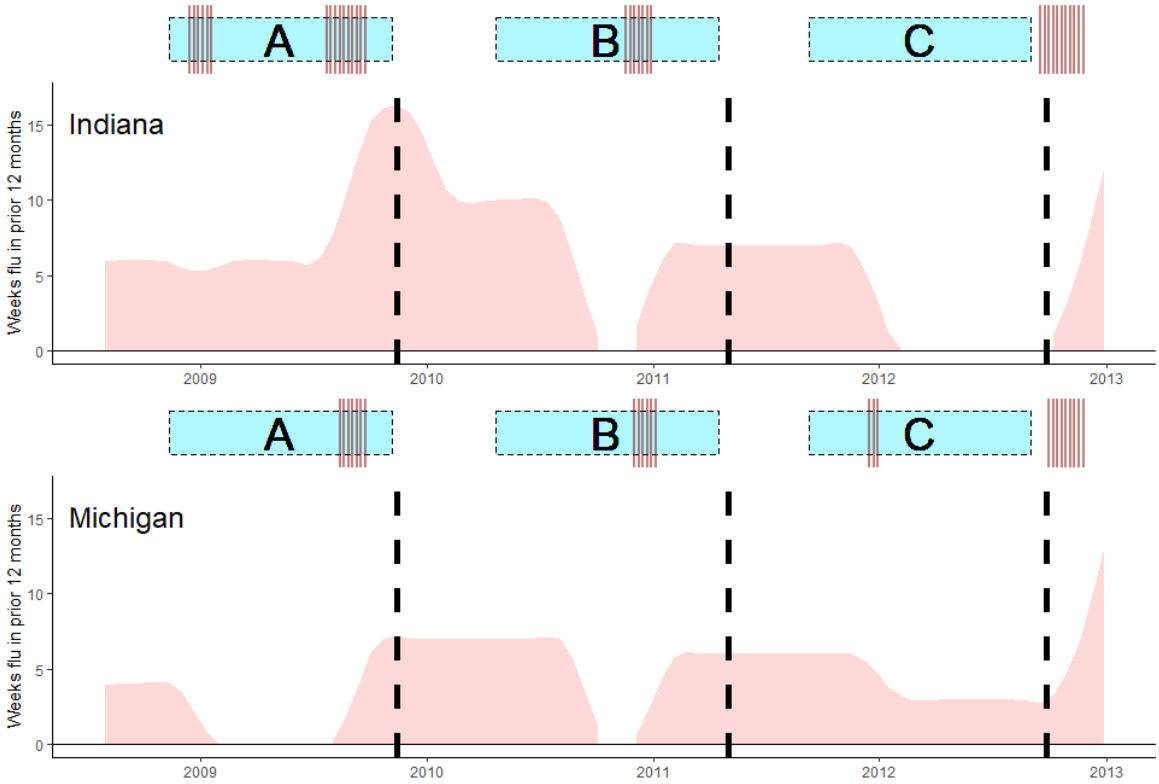

Figure 5 - Timelines demonstrating the construction of the independent variable

Figure 5 - Timelines demonstrating the construction of the independent variable.

Time window choice

I choose a twelve-month window of flu exposure for a couple of reasons. One reason is that there is no potential bias from seasonality. Seasonality bias may occur in a lopsided window if the search is affected by a time of year; that time of the year may be doubled up or skipped with a lopsided window. Another reason is that twelve months likely represents the inventive search time that an inventor experiences in much of the sample of inventors, and I describe evidence of the timing of that search process in greater detail below. In cases where the search process doesn't fit within that window but extends into the past significantly more distant than twelve months, these cases should only add noise and not bias the result. Even such cases may also be affected to a degree during the twelve months prior, as work on projects likely continues until application dates.

We, as researchers, do not have much information that indicates when inventors are doing the work that leads to patent applications. We know that work was done before the application date, but there is little research into when the inventors were actually working on that project. Penner-Hahn & Shaver (2005) note that in interviews, pharmaceutical executives stated that their companies apply for patents within 2-3 months of discovery. However, the pharmaceutical and biotechnology industries are known for patenting very aggressively (Adelman and Deangelis 2005) and likely represent the speedy end of the spectrum. Furthermore, this example does not tell us when the inventors involved were working on the project – discovery possibly triggered the process of involving others who later become coinventors.

Legal guidance for determining coinventors gives explicit instruction that shows that coinventor relationships accurately capture the ongoing process of potentially small – but meaningful – interactions of knowledge transfer. The setting of this study involves patent inventors not only because of data availability and the history of using patents as measures of knowledge with strategic importance (Jaffe and Trajtenberg 2002) but also because of the legal requirements to carefully record the sometimes cloudy reality of knowledge transfer on a micro-level. The United States Patent and Trademark Office (USPTO) definition of coinventorship provides illustrations of this process. The Manual of Patent Examining Procedure, Section V: Requirements for Joint Inventorship states, "…there must be some element of joint behavior, such as … hearing another's suggestion at a meeting." The fleeting micro-interaction of listening to someone at a meeting can result, if the contribution is sufficient, for the person who had the suggestion at the meeting to be a coinventor with little other involvement. Not only would they be included on the patent as an inventor, but they would be *required* to be listed on the patent as an inventor if the assignee wishes that patent not to be easily challenged. Conversely, the leader of a project that ultimately results in multiple patent applications may *not* be listed as an inventor on any of them if the leader did not materially contribute to the idea behind the patents. Because patents are valuable legal documents where failing to follow the rules carefully may destroy their value, firms follow such rules very closely. Being a coinventor on a patent is, therefore, an excellent indicator of a person's involvement in the knowledge and ideas of that patent, however small, at any point in time during the

patent's technological development. And the development of an invention can last up until the final drafting of the application.

Geographically local collaborators, the validation test dependent variable.

The geography of coinventors measures the geography of the search of an inventor. I consider a coinventor as "local" to another inventor if they list their residences as cities within 10 kilometers of each other. Chance encounters have a limited geographical range, so this is an appropriate scale to consider locals with the potential to be affected by chance encounters. In the sample of inventors, being local likely indicates the two are working at a shared facility, which is probably the location where inventors find most serendipitous knowledge.

To see if an inventor has fewer collaborations with locals or if they work with fewer local coinventors, I run two models with different versions of the dependent variable. The first counts each patent an inventor-coinventor pair appears on (as a collaboration); the second counts unique coinventors.

Estimation method of validation test

I use a Poisson regression with fixed-effects on a panel of 406,664 inventors of multiple patents from August 2004 to December 2015, dates that correspond to available flu data from the CDC. Inventors drop from the sample for having no variance depending on the model and variables used. I use fixed effects on inventors because I want to observe the change in the search behavior of individuals while ignoring cross-sectional variance. To remove state effects that may occur if an inventor switches states, I also include state indicators (if inventors do not change states, the fixed effect on the inventor

will absorb all state variation). I include year-month indicators to remove any overall trends in local coinventing patterns over time. While not identical, states tend to experience flu around the same time, so year-month indicators likely absorb some variation from the flu. Therefore, the variance in the independent variable driving the results is the additional variance from an unusually long or short flu in a state given the rest of the country at the time. This indicator is conservative as it likely removes a large amount of potentially meaningful variance to an individual inventor's exposure to the social avoidance effects of flu. Still, it should remove any remaining potential concerns of the measure containing seasonality or global trends that might affect the search processes. As a result, the correlation between states likely dilutes the magnitude of the effect and makes the magnitude difficult to interpret directly. However, in this validation test, I am less concerned with finding a specific magnitude of the effect, rather than showing a significant effect indicating the flu can be used as a proxy for lower levels of potential chance encounters in my sample.

Using local collaborations as the dependent variable, I also include a total count of inventor's collaborations that month as a control; when using local coinventors as the dependent variable, I include a count of unique coinventors that month as control.

Evidence of variance in chance encounters

Results supporting the use of the flu variable as constructed are in Table 9. There is less local collaboration during periods more likely affected by the social avoidance effect of flu. Model one shows local collaborations decrease with the flu. Model two uses unique inventors each month and shows a similar result. I use both specifications to show

that not only locals are worked with less frequently, but that inventors work with fewer local people. Further results (not shown) from logged-linear regression models of the same variables support that the Poisson estimation method is not driving the result.

Table 9 - Poisson fixed-effect regression results of the test validating the use of cumulative weeks of prior flu as proxy for decreasing opportunity for chance encounters

Dependent variable:	Local collaboration count	Local coinventor count
Model number	1	2
Weeks of widespread flu in prior 12 months	-0.00116 [0.000]	-0.000653 [0.013]
	(0.000223)	(0.000263)
Count of collaborations (each coinventor repeated by patent)	0.00542 [0.000]	
	(0.000012)	
Count of unique coinventors (not repeated)		0.0923 [0.000]
		(0.000199)
Inventor FE	YES	YES
State FE	YES	YES
Year-Month FE	YES	YES
Observations	1,338,871	1,337,471
Inventors	177,273	177,111
Avg observations per inventor	7.6	7.6
Log likelihood	-1660536.2	-1238768.7
Degrees of freedom	188	188
Wald Chi Squared	292579.5 [0.0000]	225051.8 [0.0000]

Standard errors in parentheses; p-values in square brackets

It is difficult to describe the result in economically meaningful terms, especially because of how year-month indicators remove correlated variance between states. However limited, I can interpret the result from Model 1 in the following way: for every week extra that state experiences flu than the average in the country that month and the average for that state, inventors in that state are about 0.1% less likely to collaborate with a local. To obtain a more readable finding, I transform results by average exposure versus no exposure (remember that over 10% of the sample had no exposure to flu so this is not an unreasonable boundary). Viewed this way, I show that there is a 1.1% decrease in local collaboration with *average flu* exposure versus *no flu* exposure.

Limitation: local collaboration frictions vs. chance encounters

More strictly interpreting my results, I show that the flu seems to increase *local collaboration frictions*. The frictions I expect to invoke with my treatment are consistent with ones that would decrease chance encounters, but I still do not confidently see an effect from chance encounters. I also have no hard evidence that the flu reduces chance encounters but does not affect planned meetings, such as the other proximity mechanisms. Consequently, the degree to which a reader believes that I describe the effect of the flu as acting on collaborations mainly through chance encounters is as strong as their belief in the support I presented earlier for the appropriateness of the flu as a measure. Robustness tests that more convincingly isolate mechanisms are difficult to imagine and will likely require exceptional data or creativity. And, due to the difficult nature of the phenomenon to study, I expect future tests will also rely on argument in key stages of the research design as I do here, just perhaps with a different set of assumptions.

We can only be confident in our knowledge of the effect of chance encounters and proximity on collaborations by triangulating research from multiple directions and multiple research designs that relax different assumptions.

Corollaries and robustness tests

To add support that my proposed mechanism of social distancing due to flu drives my findings, I develop and test two corollaries and one robustness check that all show the effect varies in predicted ways across different situations. I do not design these corollaries to eliminate the possibility of alternate mechanisms, which I believe would require different data. Instead, I design these corollaries to support the idea that chance encounters are responsible for the observed effect using the data available from patents. All tests support my interpretation of the finding. Detailed descriptions of the tests are in the next section, and their summaries are in the table below.

VALIDATION COROLLARIES AND ROBUSTNESS CHECKS

Heterogeneous distant opportunities to collaborate

Not all inventors have the same opportunity to collaborate distantly. Many may be effectively stuck with locals due to a lack of opportunity to talk to distant people. Those with more opportunities to collaborate with and reach out to others outside their geography should do so more when they experience a lack of opportunity for chance encounters from flu-related social avoidance when compared to those who are more obligated to keep their search for knowledge local, through chance encounters or not. Firms that are more concentrated in one region would present their inventors with fewer opportunities to collaborate between geographies. If concentrated firms experience less of

an effect (conversely, if multi-geography firms see more of an effect), then this corollary is supported.

To determine what people might have more distant opportunities to collaborate ex-ante, I look to the inventor distributions of assignee firms as potential collaboration partners. I consider each firm as the disambiguated assignee identifier from the USPTO PatentsView database. I determine each firm as either "concentrated" or "multi-geography" by observing the spread of inventors assigned to that firm; in my first specification, I consider that if 90% (80% and 70% in further checks) of inventors are within one state, that firm is concentrated and has less distant collaboration opportunity ex-ante when compared to firms with less concentrated inventors. About 24% (41% and 50%, respectively, in the additional robustness checks) of the sample is considered "concentrated" by this measure. In another specification, I create a Herfindahl-Hirschman Index (HHI) corresponding to the concentration of inventors per state.

I run the same regression as Model 1 but include the concentration measure and the concentration measure interacted with the independent variable, weeks of flu in the prior 12 months. This interaction term should be positive if more concentrated firms are less affected by flu, conversely, if multi-geography firms provide more opportunity for distant collaboration that may be sought during periods of social distancing. In all models tested, I find positive and significant values on the interaction term, supporting the corollary.

Implied agency of inventors

Not all inventors can freely choose who they work with, and my model of individual search relies on this assumption. Some inventors may have more agency and motivation for making their own decisions about who they approach versus relying on being assigned to working groups, for example. Those inventors with more freedom to choose their coinventors ex-ante should see more of an effect of flu affecting their search.

I test this corollary with two separate definitions of inventors I argue have more individual choice when they choose who to work with ex-ante. I look at both smaller firms and university-affiliated inventors. I will start by describing the test of inventors at smaller firms.

Research shows that smaller firms have fewer internal resources and so rely on their individual networks more for knowledge (Almeida and Kogut 1997, Feldman and Florida 1994, Rogers 2004). If inventors at smaller firms see a greater effect of flu, this is consistent with my interpretation of flu affecting search more where we expect individual agency in search to make more of a difference.

I measure small firms by using an indicator on patent applications where assignee firms indicate if they are a "small" entity. To claim small entity status, the USPTO requires a firm to have fewer than 500 employees (13 CFR 121.801). Organizations claiming small entity status get significantly reduced filing fees, so it is likely that most organizations that meet the requirement claim small entity status. A benefit to using small entity status rather than inventor counts is that it gives us an idea of the size of the organization, rather than only inventors.

I find that inventors of small entities are affected by flu nearly twice as much as inventors of non-small entities, again interpreting the interaction term between small and prior flu. Believing that employees of small entities have more agency and freedom to choose collaborators, this corollary is supported through this specification.

I also test this agency corollary in another specification using university assignees to indicate academic inventors. Academics – especially tenured professors – are likely less constrained by organizational barriers to working with whoever they choose. We should see academics' choices of collaborators more affected by flu than non-academics to support this version of the corollary.

To indicate inventors as academics, I filter organization assignee names by the words "college," "university," or similar terms. I cannot tell whether the inventors are professors or academics themselves. Still, I do know that those inventions were accomplished with an academic-type process because they are assigned to a university, which I consider adequate to represent the academic setting with fewer constraints to collaboration. With an identical test specification to the prior test except including academically affiliated indicators rather than small entity indicators, I show statistically weak results ($P < 0.10$) in the direction where academics are less likely to work with locals during the flu, weakly supporting the corollary.

Timing should match the search period

I believe a twelve-month window is an appropriate representation of the search time before a patent application due to the continuous nature of knowledge-seeking inventors engage in. The twelve-month window also removes seasonal effects making it

empirically convenient, as discussed earlier. However, in the spirit of finding corollaries to strengthen the plausibility of the flu having the observed effect of reducing local collaboration, it would help this discussion to see if the window duration makes a difference where it should. Specifically, shorter periods of expected search time should correspond to flu within that shorter window more than outside that window.

Earlier in this paper, I mention how there is limited information in the literature about when the project work takes place when compared to a patent's application date. But the evidence I give suggests that pharmaceutical and biotechnology industries file for patents aggressively and quickly in the process (Adelman and Deangelis 2005). Penner-Hahn & Shaver (2005) note that pharma executives believe that they file applications within three months of discovery. These papers are the only evidence I have useful to predict ex-ante the appropriate timing window for an industry.

To test the sensitivity of pharma patents to different timing windows, I vary the window in the original specification, create an indicator for the inclusion of pharma codes in the patent technology class, and create an interaction term between pharma patents and the different dependent variables. If there is an additional effect of pharma patents during a smaller window, the interaction term should be negative and significant. I test the same specification with a pharma indicator and interact with the regular 12-month window to ensure the pharma category is not driving the result. I find no significant effect with the 12-month interaction term on that model, but see a significant and negative coefficient on the interaction term between pharma indicator and a 6-month window in the model with the shortened window. The result supports the corollary.

Table 10 - Corollaries and robustness check summaries

Description	Measurement	Findings
Opportunity for distant collaboration		
People with more opportunity for distant collaboration ex-ante should have an additional effect. Those with less opportunity for distant collaboration should see a reduced effect.	If assignee has 70%, 80%, or 90% inventors concentrated within one state, there is less opportunity for distant collaboration. HHI index also indicates concentration. The interaction term of concentration with flu variable gives the coefficient of interest.	On all models, [Concentrated X Flu] interaction term has significant and positive results, supporting the corollary.
Agency and motivation of inventors to choose their own coinventors		
People with more ability or motivation to choose their own coinventors should have an additional effect.	Inventors with patents assigned to smaller firms (using USPTO small entity status)	Small firms show an additional reduction of local collaborations during flu, supporting corollary.
	Inventors with patents assigned to universities	Universities show a weak (P<0.10) reduction of local collaboration during flu.
(Robustness check) Timing of search process should match window		
Aggressively patenting industries with short development time before patent applications should show effects that are more sensitive to shorter windows.	Pharma and biotech patent class technology codes, 12-month window interactions versus 6-month window interactions	No additional effect of pharma class indicators and 12-month window, but large and strong additional effect with a 6-month window, supporting corollary.

Table 11: Summary of test results of opportunity for distant collaboration, corollary 1. All tests use Poisson regressions with fixed effects on inventors patenting with assignees. Positive interaction terms support corollary one, that I am less likely to observe the effect of the flu on these individuals through a measure of local vs. distant collaborations.

Table 11 - Summary of test results of opportunity for distant collaboration, corollary 1

Dependent variable:	Local collaboration count	Local coinventor count	Local collaboration count	Local collaboration count
Model number	1	2	3	4
Weeks of widespread flu in prior 12 months	-0.00318***	-0.00361***	-0.00277***	-0.00481***
Count of collaborations (each coinventor repeated by patent)	0.00513***	0.00513***	0.00513***	0.00513***
Assignee has 90%+ inventors in one state	0.00312***			
(Assignee 90%+) X (weeks widespread flu) interaction	0.134***			
Assignee has 80%+ inventors in one state		0.00278***		
(Assignee 80%+) X (weeks widespread flu) interaction		0.166***		
Assignee has 70%+ inventors in one state			0.00118**	
(Assignee 70%+) X (weeks widespread flu) interaction			0.175***	
Assignee HHI of concentration of inventors in one state				0.00445***
(Assignee HHI) X (weeks widespread flu) interaction				0.372***
Inventor FE	YES	YES	YES	YES
Year-Month FE	YES	YES	YES	YES
Observations	1,255,504	1,255,504	1,255,504	1,255,504
Inventors	181,140	181,140	181,140	181,140
Avg observations per inventor	6.93	6.93	6.93	6.93
Wald Chi Squared	224511.5***	225274.3***	225191.9***	225740.4***

* p<0.05, ** p<0.01, *** p<0.001, two-tailed t-tests.

Table 12: Summary of test results where agency of inventors should be higher, corollary 2. All tests use Poisson regressions with fixed effects on inventors patenting with assignees. Negative interaction terms supports corollary 2 that people with more ability or motivation to choose their own coinventors should have the geography of collaborator more affected by the flu.

Table 12 - Summary of test results where agency of inventors should be higher, corollary 2

Dependent variable:	Local collaboration count	Local coinventor count
Model number	1	2
Weeks of widespread flu in prior 12 months	-0.00173***	-0.00197***
Count of collaborations (each coinventor repeated by patent)	0.00513***	0.00513***
Assignee has “small” or “micro” status	0.076***	
(small or micro status) X (weeks widespread flu) interaction	-0.00205***	
Assignee is university		0.226***
(university) X (weeks widespread flu) interaction		-0.00138+
Inventor FE	YES	YES
Year-Month FE	YES	YES
Observations	1,285,420	1,285,420
Inventors	187,282	187,282
Avg observations per inventor	6.86	6.93
Wald Chi Squared	223764.0***	224281.6***

+ p<0.10, * p<0.05, ** p<0.01, *** p<0.001, two-tailed t-tests.

CHAPTER 4: CHANCE ENCOUNTERS INCREASE INVENTION NOVELTY

CHAPTER SUMMARY

I explore the effect of chance encounters on innovation. Using the flu measure as proxy for lower potential chance encounters developed in the prior chapter, I predict and find that collaborations that are more influenced by chance encounters result in more novel recombinations of knowledge because "serendipitous" knowledge from chance encounters is less constrained by the tendency toward local, exploitative search. During times with fewer chance encounters, inventors substitute where they get knowledge, more often turning to prior contacts and, therefore, limit recombination potential.

INTRODUCTION

Chance encounters – unplanned social interactions between people who happen to be at the same place at the same time – are partially responsible for the flow of technical knowledge by affecting the direction of an individual's search for knowledge (Allen 1977, Catalini 2018). Convincing recent empirical studies show that chance encounters enabled by proximity are responsible for some amount of recombinant innovation (Catalini 2018) and individual exploration (Lee 2019) within organizations. However, these studies show that altering the set of proximate people with whom chance encounters are possible drives collaboration and learning. Existing studies do not tell us how innovation stemming from chance encounters may be different than innovation stemming from other ways of finding knowledge.

Some firms encourage chance encounters between employees, and the current global COVID-19 pandemic has given us a glimpse of a world where chance encounters are greatly limited. We should want to know how such increases and decreases of the chance encounters may impact innovation. With this motivation, and in the spirit of calls to better understand microfoundations of firms' capabilities (e.g., Coff and Kryscynski 2011), I ask, what is the effect of chance encounters on innovation?

If a chance encounter between inventors at a watercooler results in a valuable innovation, the inventors will call it "serendipity," the luck of being in the right place at the right time. A unique feature of chance encounters is chance and apparent luck – illustrating a lack of intention by both parties. Because these interactions are unintended, they are likely less constrained by the cognitive tendencies toward familiar sources and

local search (Levinthal and March 1993, March and Simon 1958), so innovations from chance encounters may be likely to be more novel recombinations than if they do not stem from chance encounters.

To learn the effect of chance encounters on innovation, I develop a theory to generate hypotheses to test an alternative scenario – what happens when potential chance encounters decrease? What if there's nobody at the watercooler? In the watercooler example, the inventors will return to their desks, not knowing what they missed, and will carry on as before. In my theory, inventors have different ways to get helpful knowledge of components to recombine that solve a problem, and inventors switch between these ways depending on their availability. When chance encounters reduce, they are replaced with knowledge that was more deliberately sought. Specifically, I expect more collaborations with prior contacts and especially with repeated contacts when potential chance encounters decrease.

To empirically test my hypotheses, I use local flu prevalence as an exogenous influence on the opportunity for chance encounters during inventive search. Borrowing from Chapter 2, I argue that a social distancing response to a disease epidemic – a phenomenon well-established in epidemiology literature (e.g., Fenichel et al., 2011; Ferguson, 2007) – effectively creates variation in the likelihood of chance encounters between proximate people. There is significant evidence showing that people avoid engaging in social interaction when they're worried that such interactions might make them sick. My research design does not change the proximity between people as other studies (e.g., Catalini 2018), but instead changes the opportunities for otherwise

proximate people to experience chance encounters. In this way, I can estimate what collaborations were more or less likely to have been influenced by "serendipity."

Using a large sample of U.S. inventors of multiple patents and weekly flu data ranging over 11 years from the U.S. Center for Disease Control (CDC), I find evidence that the flu affects collaboration patterns in ways consistent with reduced chance encounters. I use this measure to find support for my hypotheses, showing that novelty decreases when chance encounters decrease, and that prior contacts are more often collaborators.

I organize the paper as follows: first, I review the research on the importance of chance encounters to innovation. I then develop theory on problemistic inventive search and knowledge substitutions, generating hypotheses. I then describe my hypothesis testing methods, results, and conclude.

Inventive search and knowledge worker interactions

Strategy and innovation research commonly considers knowledge workers as a fundamental unit of knowledge in an organization (Simon 1991). How an organization can coordinate knowledge within organizations is a central research question of a long stream of innovation and firm knowledge literature (e.g., Kogut & Zander, 1992). Knowledge flows and innovation in organizations are not just orchestrated by top management; they are also bottom-up processes determined by individuals and their contacts (Burt 2004, Lazer and Friedman 2007).

We understand that innovation and invention arise from inventors recombining "components," also known as "factors" or "ingredients." (Nelson and Winter 1982,

Schumpeter 1939). Fleming (2001) describes the inventing process as an *inventive search* resulting in new recombinations of useful "components." During inventive search, inventors look for components to combine but usually rely on those that are "salient, proximal, and available" (Fleming 2001, p. 110) because of bounded rationality and the tendency toward local search (e.g., March and Simon 1958, Nelson and Winter 1982). Familiarity with the use of components constrains inventive search, so inventive search is ultimately a search for knowledge of the relevancy and uses of components, rather than "components" themselves. Helpful familiarity and knowledge of components can be found in other people, a point Fleming (2001) also makes. The search for knowledge between people is understood as a core process of knowledge flow within organizations (Nahapiet and Ghoshal 1998, Sorenson et al. 2006).

Inventive search is, therefore, partly guided and enabled by the knowledge of other people and the knowledge of what components those people might have familiarity with. Borgatti and Cross (2003) describe similar constraints in their relational framework determining advice-seeking behavior. Physical proximity between people can decrease these constraints by increasing awareness of proximate people and their knowledge (Allen 1977, Catalini 2018). Awareness is thought to increase through the increased likelihood of chance encounters between proximate people (Catalini 2018). Chance encounters are generally considered opportunities for social interaction uniquely available to those in close physical proximity. Boudreau et al. (2017) show that even short social interactions can increase awareness enough to play a large role in determining collaboration, so it is not surprising that proximity is shown to form social

relationships (Chown and Liu 2015) as well as enabling collaboration through lowering search costs (Catalini 2018).

Proximity is not only associated with the formation of collaborations, but the quality of output as well. Catalini (2018) finds that proximity is associated with higher quality output by finding higher citations of scientific output by collocated labs. He suggests that the increase in quality may come from either more effort enabled by colocation or because distant ideas are recombined when people change locations, enabled by increased awareness. Lee (2019) shows that individual exploration increases with proximity to previously physically distant people by measuring exploration as sales employees selling new products after changes in proximity within a facility. Both Catalini and Lee examine microgeography by exploiting exogenous shifts in proximity between people and find that changing proximity can matter significantly to collaboration and knowledge, and both reasonably argue that chance encounters are a likely mechanism. Chance encounters are likely responsible for the major changes in outcomes measured. These papers are notable for their robust research designs, which should draw the attention of innovation scholars to the large and meaningful magnitudes of their findings. The lesson is clear from this growing stream of microgeography literature: taking a bottom-up individual perspective is potentially more important than previously considered, and that chance encounters between people can play a key role in the flow of knowledge and innovation.

The role of chance encounters

Despite this research, we don't know the effect of the knowledge gained through chance encounters on the nature of resulting innovation, versus other options of sourcing knowledge. The type of knowledge gained through chance encounters may be different and may influence the nature of innovations. Prior studies do not show us if knowledge gained through chance encounters is fundamentally different than knowledge gained in other ways, so we are left with a gap in understanding.

I will illustrate this gap with the above microgeography literature because it provides the best evidence of the innovative impact of chance encounters. Let us assume that the increase in quality of scientific output in Catalini (2018) is driven by awareness enabled by chance encounters and has nothing to do with other benefits of proximity, like enabling increased effort. In this case, the quality of output increases because chance encounters are now with members of a newly proximate lab. When the new lab moves in, new recombination potential is unlocked by members of both labs becoming aware of each other through chance encounters enabled by proximity. Using this logic, the newness of the proximate lab matters to the quality of output, not the chance encounters.

Similarly, Lee (2019) shows that exploration increases because chance encounters occur between new people when workers change locations. While these studies are valuable evidence that chance encounters matter, they do not tell us about the nature of knowledge obtained by chance encounters. They implicitly assume there is a consistent background rate of chance encounters, and that *whom these chance encounters are with* can be manipulated with changes in proximity. We know from other work that the quality

of innovation can increase when working with new people (e.g., Audia and Goncalo 2007). We still don't know what effect chance encounters, themselves, have on the nature of innovation. To address this gap, I ask the simple question: *what is the effect of chance encounters on innovation?*

Anecdotal evidence from firms like Apple shows that some firms are motivated to increase the frequency of chance encounters to increase innovation, designing workplaces to encourage interaction (Hess 2017). Due to the COVID-19 pandemic, many of us are familiar with a situation where few chance encounters occur because many people are working from home. However, existing research does not tell us what types of innovation are encouraged by increases or decreases in chance encounters. Both of these current trends further motivate my research question.

From an organizational perspective, all organizations dealing with knowledge should want to know the relationship between chance encounters and innovation. Firms have heterogeneous goals for innovation. Firms may not want the nature of their innovation to change, or at least they would likely want to know when their actions could affect it. Exploration is inefficient (March 1991), so if Apple's interventions encourage exploration at the cost of exploitation, firms currently valuing exploitation should not copy them. Decisions that affect chance encounters like flexible (or mandatory) attendance policies may also have unintended effects on innovation. The relationship between chance encounters and innovation should be interesting to any organizations and managers involved with innovation or the coordination of knowledge.

THEORY DEVELOPMENT

Theory of inventive search and knowledge substitutions

To answer my research question, I present a model of inventor search as a problemistic search where ways of getting knowledge can be substituted. I develop this model based on the grounded and fundamental model of recombinative inventive search presented by Fleming (2001). I focus on the "search" metaphor of inventive search as individual-level search is described in March & Simon (1958) and numerous subsequent papers as a search for a solution to a problem. Fleming (2001) cites these sources in the original description of inventive search but does not make explicit the connection between search and a problem-solving motivation. My drawing this connection expands and clarifies, rather than contradicts, his view. Fleming (2002) uses the word "problem" to describe the motivation of inventors 17 times in his later paper that provides a rich illustration of inventive search, so I believe thinking about inventive search as problemistic search is both a grounded and consistent with its original presentation by Fleming (2001). Inventive search is ultimately then a type of problemistic search; it is a search for knowledge of useful components, knowledge, or ideas that might solve a technical problem when combined.

I make this distinction of problemistic search to apply other features of problemistic search to develop testable hypotheses. Notably, a feature of problemistic search is that when a problem is solved, the search generally ceases. Therefore, in a model of problemistic inventive search, inventors are ultimately looking for the first, satisficing recombination that presents a solution, not all possible recombinations or

solutions. Also, when a problem remains, a person may adapt their search to different forms until a solution is found (Billinger et al. 2014).

Inventors search for knowledge in different, substitutable ways, or channels, and these channels are subject to availability. When one channel of obtaining knowledge (like chance encounters) decreases, an inventor may take up the slack in other ways. They do so because they are motivated by the problem – not by the number of available solutions. When inventive search is explicitly viewed as problemistic search, it allows us to think counterfactually to develop and test hypotheses to examine the nature of knowledge between sources – if the problem isn't solved with knowledge from one channel, it may be solved by another. Evidence shows that problemistic thinking matches inventors' thought processes (Fleming 2002), so this model of problemistic inventive search with substitutions is a grounded and useful view of inventive search.

Chance encounters are an unplanned way of finding components

Inventors are generally employed because they already have some expertise in the activities I describe as inventive search. Inventors are usually scientists or engineers with formal training in a subject area, and their knowledge of books and resources to consult for additional information is consistent with their training. They also likely have a set of suggested procedures to follow, whether assigned to them for record-keeping or regulatory purposes, taught as a part of training, or created by themselves from experience. Technical problem-solving is routine for these roles; it is often quite literally the job description for any engineer. There are, therefore, more or less *standard ways* of solving a technical problem for these people. These "standard ways" are generally

deliberately progressed through as a routine when an engineer or scientist comes across a technical problem. These steps might involve first trying to solve the problem using familiar knowledge and tools (for example, tweaking computer simulations), then progressing to other tools (e.g., testing prototypes), then revisiting how components might be used, changing manufacturing method, revising the assembly architecture, revising design inputs, and so on. Different roles in different industries have specific nuances and unique steps, and engineers and scientists in these roles know how to replicate their own "standard way."

However, solutions to technical problems are not immediately apparent. Other potential solutions may present themselves when a problem is on that inventor's mind before they have a chance to pursue their standard resources. A chance encounter and conversation with someone that the inventor had not considered asking for advice might present such a potential solution. This chance encounter would break the inventor's routine and divert their attention to the new potential solution they would not have considered without the chance encounter.

Theoretically, of the different ways to get helpful knowledge of components for recombination, chance encounters are uniquely interesting because they are unplanned and undirected by the searcher. Search is usually considered constrained by cognition (e.g., Levinthal and March 1993, March and Simon 1958), as I illustrate above with the inventor's routine, which is determined by their training and experience. But chance encounters are with whoever is nearby, whatever their expertise happens to be. Alternative ways of seeking knowledge are ultimately more planned than chance

encounters, and therefore more thoughtful, and therefore more constrained by the individual searcher's knowledge, training, and routines.

The lack of constraint should spark variety in the type of knowledge of components possible to receive. Granted, knowledge from chance encounters is not entirely unconstrained: the searcher still needs to understand the other knowledge and how it might be helpful (Cohen and Levinthal 1990, Nooteboom 2000). And it is possible that people with homogenous knowledge physically surround some searchers, so there would be no effect for some individuals. But generally, the variation of ideas and knowledge of components from chance encounters should increase versus the standard alternatives. Moreover, because chance encounters are not planned, there are likely planned alternatives available to substitute.

Chance encounters should increase variance in knowledge sources applied to a problem area by an individual inventor. For example, the inventor might talk to a new person with knowledge the inventor never knew about before. Or the inventor might speak with someone they already know but would not have thoughtfully approached concerning the problem they have. Either of these situations would increase the variance in knowledge sources that the inventor is attempting to apply to the problem and variance in knowledge components available for recombination.

When there are fewer opportunities for chance encounters, alternative ways of seeking knowledge will be more deliberate than chance encounters and will be more familiar to the inventor. Conversely, when inventors rely on the standard and familiar sources of knowledge, inventions will be more like past inventions, reducing novelty.

H1. When there is less opportunity for chance encounters, the novelty of innovation should decrease.

Prior contacts substitute for decreasing chance encounters

The "standard way" of approaching a problem also includes people and their knowledge. Individual routines should involve other people. An inventor knows that they got a specific type of advice from a prior collaborator before and so will return to that person for advice when they have a similar type of problem.

Borgatti and Cross (2003) develop a useful framework that describes why people may seek knowledge in prior contacts more than others. Specifically, prior contacts likely know what each other know – this meets the first requirement in the Borgatti & Cross framework. Furthermore, Borgatti & Cross also point out that someone seeking knowledge is concerned with how receptive their contact might be when approached for advice. If the perceived cost is high, they may not approach someone even if they know that they have valuable information. However, an inventor should be aware of how available a prior collaborator may be to patiently listen to a request, which should increase their likelihood of reaching out.

In my model, when chance encounters present a solution, inventors will use it. When there are fewer potential chance encounters, those inventors who would have been content with knowledge gained through chance encounters now have to seek knowledge actively. When seeking actively, inventors may first contact people they used to work with, because they are more aware of those prior contacts' knowledge, and inventors know that they can access the knowledge in prior contacts without significant cost.

H2a. When there is less opportunity for chance encounters, inventors will work with prior contacts more often.

People learn more about others over time and with experience. Repeated collaborations are opportunities to learn about the depth of a contact's knowledge, and increased awareness of a contact's knowledge should increase the likelihood of someone approaching the prior contact. Trust also increases in repeated interactions, and trust improves the transfer of knowledge between contacts (Levin and Cross 2004). Trust further lowers the perceived cost of asking, consistent with the above described relational framework.

Because of increased knowledge of the contact's knowledge and increased trust, when an inventor faces less available serendipitous knowledge, the more likely people to ask for knowledge are their past contacts with whom they have more shared experience.

H2b. The effect of H2a should be greater for repeated contacts.

Implicit in the model of substitution is that, depending on the efficiency of substitutions, overall productivity could go up when opportunities for chance encounters decrease if substitutes are more efficient (which is possible, as exploration is inefficient, e.g., March 91), or chance encounters provide new productive opportunities that wouldn't otherwise exist, so productivity could go down. Because chance encounters are likely not the usual or expected way to solve technical problems, there are likely ready substitutes for at least this source of knowledge, so I do not expect productivity to decrease. The question of a net effect on productivity is ultimately an empirical one, but the model assumes that substitutions may balance net output. To get a fuller picture of the effect of

chance encounters on innovation, I empirically explore any potential effects of chance encounters on the innovative productivity of individuals and the nature of that innovation.

EMPIRICAL CONTEXT

I use the empirical context of Chapter 2, with the measure developed for use in this chapter and the same sample of inventors. The validation test and support from further corollary tests in Chapter 2 make me confident that the flu decreases chance encounters in my sample consistent with epidemiology literature, and that my measure can be used as a proxy independent variable to test my hypotheses.

Hypothesis tests

Because each test has significantly different approaches, in this section I outline each method of test and present results before moving on to the next test.

Novelty

To test Hypothesis 1, I test if patents that had more opportunity to be affected by chance encounters during inventive search are more novel than those with less opportunity to be affected by chance encounters. If those potentially more affected by chance encounters are more novel, this finding supports my hypothesis. Because the flu only decreases potential chance encounters, more flu should be negatively associated with novelty.

I use a popular existing measure of patents generated by assessing the novelty of the combinations of citing patent IPC classes called the "originality index" (Squicciarini et al. 2013). The concept of novelty considers the sources of knowledge recombined to

make a given innovation, if the knowledge recombines in a new way, innovation is said to be more novel (Trajtenberg et al. 1997). The originality index gives a continuous value of the novelty of combinations of citing knowledge sources, so it is an appropriate measure for the novelty of component recombinations from inventive search.

Because novelty is a characteristic of the invention and not the inventor, but the inventors determine geography and therefore the independent variable of cumulative flu exposure, I approach this analysis in two ways. The first is by using patent-inventor data points where the novelty of patent and prior flu indicator is based on inventor geography, using the same set of multiple inventors from the validation test. This test repeats each patent by coinventor in the United States because those are the inventors with geography that can be tied to flu prevalence. In this model, I use year-month indicators to remove overall trends and state indicators to remove state effects. In the second test, I assign an average prior flu variable to each patent by all such coinventors present. I also use year-month indicators in the second test to remove overall trends. Both tests are OLS linear regressions with year-month indicators because the dependent variable of novelty is a continuous value where 0 is not novel.

The results for the test of Hypothesis 1 are in Table 13. Both specifications support the hypothesis, where weeks of flu are negatively associated with novelty, both using inventor-patent data points and average flu on each patent. These results indicate that decreased opportunity for chance encounters during inventive search can decrease the novelty of an invention.

Table 13 - Novelty of knowledge recombination (DV = "Originality index"), OLS models

Model number	Inventor-patent observations 1	Patent observations 2
Weeks of widespread flu in prior 12 months /52	-0.00526 [0.000] (0.00140)	
Average weeks of widespread flu in prior 12 months /52		-0.0152 [0.000] (0.00198)
Year-Month FE	YES	YES
State FE	YES	N/A
Observations	3,334,837	1,275,447
R-squared	0.0104	0.0016
Degrees of freedom	187	137
F Statistic	195.18	13.84

Robust standard errors in parentheses; p-values in square brackets

Repeated collaborations

To test that a lack of opportunity for chance encounters drives repeat collaborations, I test if prior patenting relationships are renewed when there is a decrease in potential chance encounters. If there is a decrease in collaborations with new contacts when inventive search periods are more affected by the flu, this supports my Hypothesis 2a.

To construct a variable for prior connections, I first take the entire history of all inventor-coinventor pairs in my sample. I count how many times each pair has coinvented before each given application. If it is the first date any pair coinvents, the count is equal to zero, and I assign a value of 1 to a new dummy variable "new contact." I sum new contacts by each productive inventor-month, creating a count of new contacts that month. This test follows a similar Poisson regression specification as in model 2 in the validation test, with weeks of prior flu as the independent variable, controlled by a

count of unique coinventors that month, inventor fixed effects to observe changes in an individual, year-month indicators to control for trends, and state indicators to control for switching state effects.

To test Hypothesis 2b, I test that inventions subject to a decreased opportunity for chance encounters during their development have coinventors who share a deeper history of collaboration. If inventions more affected by the flu during inventive search represent more prior working relationships, this finding supports the hypothesis. I create an aggregate sum of prior collaborations with coinventors each productive month as a measure of the prior depth of experience. A pooled count of prior collaborations with all inventors in that month makes the dependent variable, and this test is otherwise specified as the previous test.

The results for the test of Hypotheses 2a and 2b are in Table 14. In Model 3, the test of Hypothesis 2a, the coefficient of weeks of cumulative weeks of flu during inventive search decreases the number of first patents with coinventors, supporting the hypothesis. Inventors are less likely to work with new coinventors, and therefore are more likely to work with prior inventors during inventive search times more affected by the flu. Model 4 tests Hypothesis 2b, and shows a positive coefficient of cumulative weeks of flu during inventive search on the number of prior collaborations the focal inventor had with coinventors prior to that month. This supports Hypothesis 2b, showing that not only are prior contacts turned to (as in Hypothesis 2a), but that coinventors collaborated with during times with less opportunity for chance encounters represent a greater depth of prior working relationships.

Table 14 - Prior relationships with coinventors, Poisson fixed effect models

Dependent variable:	First patents with coinventors	Prior collaborations with coinventors
Model number	3	4
Weeks of widespread flu in prior 12 months	-0.000578 [0.007]	0.00115 [0.000]
	(0.000215)	(0.000051)
Count of unique coinventors (not repeated)	0.1589 [0.000]	0.0314 [0.000]
	(0.00019)	(0.000023)
Inventor FE	YES	YES
State FE	YES	YES
Year-Month FE	YES	YES
Observations	1,640,973	1,663,320
Inventors	241,994	245,249
Avg observations per inventor	6.8	6.8
Log likelihood	-2019054.8	-9752109.8
Degrees of freedom	188	188
Wald Chi Squared	777539.1 [0.0000]	8.92e+06 [0.0000]

Standard errors in parentheses; p-values in square brackets

Productivity

The model of satisficing, problemistic inventive search that I present assumes that productivity is not necessarily affected and that combinations of components that make solve technical problems will be found through inventive search – the only question is where. I include a descriptive test on productivity because it is possible that productivity may increase with chance encounters (as they increase available knowledge of components), productivity may decrease (as they present more novel components, and exploration is less efficient), or there may be no effect (either through a combination of the above effects, or because substitutions are efficient – I will not be able to tell the specific cause).

To test any effect on the productivity of individual inventors, I design a test to see if the timing of when they are productive is affected by cumulative prior flu. All other tests are conditional on innovative productivity, that is, all observations are inventions visible through patent applications, but this test includes all potential times an inventor could have been productively generating inventions but was not. Whatever the association between the productivity of inventors, findings of this test should provide suggestive evidence for one of the scenarios described above.

I create a monthly panel with the same sample of multiple inventors and create a binary productivity indicator with zeros for months without patent applications and ones for months with patent applications. I organize the sample into two panels with different counterfactual sets, one meant to realistically indicate only the potentially productive months of each inventor near times they have patents (the "realistic" panel) and another that includes all months of the sample (the "full" panel). The realistic counterfactual panel includes all months between an inventor's patent application dates and one-year windows before and after their first and last application. The full panel includes all months, regardless of the activity of inventors, and is included for robustness. I use a conditional logit model for each panel where the binary indicator of productivity that month is predicted by the weeks of flu widespread in the 12 months before that month, with inventor fixed effects and indicators for year and month. If the resulting coefficient of the flu variable is negative and significant, then more flu may be negatively affecting innovative productivity levels, perhaps because of less available total knowledge of components. If the coefficient is positive and significant, more flu may be increasing

innovative productivity, perhaps by trading exploration activities for more efficient exploitation activities. If there is no effect, we do not know if this is because of a combination of the counteracting suggestions above, or because substitutions for chance encounters are simply efficient. I make no prediction of the outcome. This test is only present to more fully describe the effect of chance encounters on innovation.

The productivity timing test shows mixed results, depending on specification. Innovative productivity does not seem to be associated with chance encounters with the realistic counterfactual set, shown in Table 15 Model 5, though it can be interpreted as weakly positive (P=0.067). The full counterfactual set shows a more significant positive result, in Table 15 Model 6. I will discuss the possible implications of these results in the next section.

Table 4 – Productivity timing test, conditional logit model

Table 15 - Productivity timing test, conditional logit model

	Realistic counterfactual set	Full counterfactual set
Model number	5	6
Weeks of widespread flu in prior 12 months	0.000355 [0.067] (0.000194)	0.00110 [0.000] (0.000186)
Inventor FE	YES	YES
Month FE	YES	YES
Year FE	YES	YES
Observations	23,005,156	37,131,795
Inventors	271,035	271,035
Avg observations per inventor	84.9	137 (for all)
Log likelihood	-5087054.3	-5656049
Degrees of freedom	23	23
Wald Chi Squared	49721.54 [0.0000]	65879.42 [0.0000]

Standard errors in parentheses; p-values in square brackets

DISCUSSION

Substitutions may be efficient

The results from the productivity test do not test a hypothesis but give suggestive evidence that is contrary to expectations of some economics literature around innovation. Namely, there is an expectation that productivity should increase with the introduction of more available knowledge, generally. I show that, in this case, productivity does not decrease when a source of knowledge is removed. Whether this is because of efficient substitution between channels in the satisficing, problemistic view I present, or because the impact of more available knowledge is counteracted by the relative inefficiency of pursuing exploration (March 91), I cannot say. This observation allows me to make a broader point about the answer to my research question about the relationship between

chance encounters and innovation: chance encounters affect novelty, but not necessarily the productivity of innovation.

Geographically "local" search connections to literature

Because chance encounters are more likely between proximate people, researchers claim chance encounters are responsible for some knowledge spillovers and agglomeration effects in regions (Bettencourt et al. 2007, Saxenian 1996). Further understanding chance encounters has the potential to expand into agglomeration literature. For example, Rosenkopf and Almeida (2003) consider geographically local search a challenge to overcome through mobility and alliances, but my findings suggest that replacing geographically local search with geographically distant search would not necessarily help exploration. More chance encounters – which are necessarily geographically local – may have a similar beneficial knowledge effect to the steps they outline to avoid geographically local search.

Literature also suggests that social networks of knowledge and geographically local knowledge can act as substitutes (Agrawal et al. 2008, Bell and Zaheer 2007, Sorenson et al. 2006). This literature does not isolate chance encounters from other ways of getting local knowledge, but because chance encounters play a role in local knowledge, they may be involved in these findings. Because chance encounters are available to those in physical proximity, substitution between social and local sources of knowledge mirrors my theorizing.

Limitations and Conclusions

This paper comes with limitations. The most substantial of those limitations is that I do not measure chance encounters directly, but instead argue that I can estimate the influence of potential chance encounters on projects already completed. I attempt to address identification concerns with a novel independent variable – cumulative flu during the inventive search – with variation exogenous to other innovation processes. Using the flu, I vary "serendipity" without varying proximity. I validate the use of this independent variable in the setting of this study with multiple tests. But I understand that all specifications remain indirect measures requiring assumptions, and my findings are only as strong as those outlined.

In this paper I develop a theory of problemistic inventive search, including knowledge worker interaction and knowledge substitutions, and I use that theory to generate and test hypotheses about the effect of chance encounters on innovation. I describe that the ways knowledge is sought are subject to availability and can be substituted, and these substitutions impact the nature of innovation. Chance encounters have the effect of increasing the novelty of invention because knowledge found in chance encounters is relatively unconstrained by an inventor's tendency toward local search. When potential chance encounters decrease, inventors rely on more familiar sources of knowledge – specifically, prior contacts. Drawing from the same sources of knowledge decreases the novelty of potential recombination – but might increase efficiency. I ultimately that show chance encounters are a bottom-up influence on the nature of

innovation. These insights are relevant to the microfoundations of innovation and knowledge transfer literature.

At the time of this writing, the return to the workplace after the COVID-19 pandemic fades is unknown, both when and if a "return" will be fundamentally different than practices preceding the pandemic. Many firms claim to retain work from home policies or flexible working arrangements even after the pandemic (Gallagher 2020). The work-from-anywhere phenomenon (e.g., Choudhury et al. 2019) may become a lasting trend in industries where remote work is possible. The effect of chance encounters on innovation should be considered when making return-to-work decisions, when the opportunity to make them should arrive.

CHAPTER 5: CONCLUSION

Organizations and managers manipulate the proximity of knowledge workers to improve knowledge sharing, access, and innovation, whether by moving people within the facility or by making geographic facility location decisions. This effort by organizations and managers to attain knowledge-related goals motivates my dissertation. So, can my research help firms better improve knowledge outcomes? I believe it's helpful. My research offers some empirical grounding to management trends, uncovers the next set of research questions that would better address problems facing managers, and provides some hope for a better-informed future, despite the new challenges the future seems poised to bring.

In this conclusion, I outline what I think are the pieces of knowledge I have uncovered that might be useful to managers and what challenges I have uncovered to current thinking, and tie these conclusions to suggested strategies to improve innovation outcomes. I also discuss how research linking proximity and innovation should inform the future of work because significant indicators show that the nature of knowledge work will change soon, for better or worse.

Current trends

I outline two managerial trends in the introduction: firms attempting to increase chance encounters and informal conversations and the rise of remote work. In my research, I observe the grains of truth to both seemingly opposite managerial trends above.

Casual conversations and chance encounters can increase the novelty of innovation because you might work with knowledge (or people) you didn't know might be helpful in advance, as I conclude from my dissertation. This is what Apple was trying to manipulate through their design of the Apple Campus (Hess, 2017). But also, since COVID has given many a taste of working without proximity, people have found that they prefer remote work and have found ways to make it productive. Many knowledge workers have found that putting their heads down and working without any watercoolers can seem a less costly work process (Duffy 2021). My research suggests that people working from home are likely going back to familiar sources that they know how to work with, and are getting things done, effectively solving problems – and it feels easier because of how familiar they are with the sources of knowledge they seek and work with.

The difference is in the *quality* of the innovation – proximity, when acting through increasing awareness of others and their knowledge, should increase the novelty of innovation by increasing the variance in knowledge sources accessed, breaking routine local search normally targeted at familiar sources.

End of clusters?

A new potential trend is related to the rise of remote work, and is the purported diminishing of the importance of knowledge clusters. After all, if everyone is working remotely, geography should stop mattering to knowledge transfer (other than presenting minor inconveniences like time zone issues). This potential trend is less studied and written about formally because it has not yet materialized.

The dissolution of clusters may not realize as an actual trend, but it is a logical extension of the logic of work-from-anywhere trend, so I discuss this potential trend here and briefly conclude by stating my belief that it is a bad idea, and locating in clusters is still a good strategy for many. I am mostly aware of this potential trend from tweets from prominent venture capitalists suggesting that because of remote work, they say, everyone might as well move to the beach (and abandon the high costs of locating in the expensive Silicon Valley). I believe abandoning knowledge clusters would be a mistake, given the various mechanisms proximity offers that can benefit individuals, firms, and venture capitalists (e.g., monitoring costs lower and trust increases should be especially salient for venture capitalists, as well as awareness and relationship mechanisms). Individual contributors who want to work without interaction might benefit from not caring about where they live, but moving away from knowledge clusters would only hurt those in the business of innovation and collaboration.

Suggested strategies to shape innovation

How can managers improve innovative outcomes, and is there some balance between chance encounters versus other search methods for a given situation? I believe managers and organizations can do much more to optimize and shape the innovative outcomes of their employees, and this balance will depend on the organization's goals. Organizations with different goals should utilize different strategies.

Goals for innovative collaboration

Often in research, manipulations to worker locations are seen as attempts to maximize employee productivity. On its face, “maximizing productivity” seems like a

worthy goal. But if we take my view that employees are engaged in problemistic search and keep working to solve their problems, despite their context, it becomes more apparent that the definition of "productivity" can take different forms. The idea of "productivity" of employees of an organizations should be productive efforts aligned with that organizations goals. I effectively assume workers are consistently working with more or less a consistent amount of effort, at least, but they might change what they're working on and who they're working with. So, does their employer want problems solved quickly and easily? Or do they want a game-changing novel innovation? What one firm might value and see as an increase in productivity, another might see as a distraction or busywork. So, I do not frame the goals as maximizing productivity as I believe that is too vague a term to be helpful.

My dissertation highlights a significant concern in the use of proximity to increase innovation, which informs my first goal an organization should have when attempting to use proximity to help employees' knowledge related processes. When proximity causes innovation, the collaborators are chosen by each other, leading to homophily in collaborations. This homophily is not desirable from two different perspectives. The first is that recombinative innovation should suffer if knowledge sources are not diverse (and demographic diversity correlates with knowledge diversity). The second is that homophily can lead to undesirable social outcomes by excluding and further marginalizing already minority groups in an organization. Therefore, avoiding exclusion and not unwittingly exacerbating discrimination between employees should be the first

goal of an innovating organization for both moral and practical reasons. This goal should apply to any organization.

Other innovation goals should vary by the organization. Specifically novelty, efficiency in problem-solving, or carefully matching technology development to internal capabilities or key competencies are potential goals that different organizations might value differently from each other.

Nimble incumbents or new firms may value novelty in innovation because they might be trying to define a new industry or have a greater ability to adapt to new technology than their rivals. Slower moving incumbent firms may value efficient problem solving instead of new technology that might undermine their current technology, if they believe they cannot adapt to technological changes.

However, all firms likely have a goal of exploiting their own creations, so matching technology development to internal capabilities is another goal that they should not overlook in the interest of blind innovation. And there's a risk that, by letting employees collaborate amongst themselves, organizations forfeit some direction, and employees may create technology that might benefit their competitors more than themselves, when they could have created something better matched to the firm's own competitive advantage.

Given these potential goals, I have some strategic suggestions for organizations.

Strategic suggestions for organizations

1. Be clear on goals and capabilities. My most straightforward strategic advice to organizations would be to be clear on their goals and knowledge of their own

capabilities. While this sounds trivial, clarity here will help them when they likely have to make trade-offs. Specifically, organizations should consider how they might handle the development of a novel and unexpected innovation. Many firms claim to value novelty, but I believe many claim it for legitimacy reasons and not as a statement of confidence in their underlying capabilities. Knowledge of goals and capabilities is necessary to implementing any thoughtful strategy.

2. Have an inclusive culture. My following advice also applies to all organizations, primarily to address the problems created by homophily. In my view, when people choose to work with each other, homophily is a bug, not a feature. I understand why homophily occurs; people tend to trust and are more comfortable talking with similar people. However, the increased levels of trust and comfort that may break the ice in homophilic relationships should be extended to all, and this extension requires an open and inclusive organizational culture. I believe such a culture is a crucial ingredient to innovation excellence and should also be seen as nearly a requirement to effectively use informal and social interactions between employees for organizational goals without exacerbating discrimination and exclusion of minority groups.

3. If the organization values novelty, more random mixing. A firm that knows it can both handle novelty and desires novelty would benefit from workers having more awareness of *random* others and the freedom to access the knowledge of *any* other person they can develop intellectual property with. Socializing between departments and roles should increase trust and build on the inclusive culture from point 2. Make inclusive opportunities for anyone to meet anyone. Generally, encourage social events, but not

exclusive types of events that some groups of people might avoid. Consider introducing genuine random assignments for casual conversations at low-stakes events, like weekly coffee mixers before group meetings, or similar. Do not have a desk layout determined by department or working team. Instead, randomly assign working locations, and change them periodically. If there are isolated groups of people, by virtue of being in a different building or similar, be sure to give them access to main areas, or mix them together with the others.

4. If the organization values problem-solving, enable focused search. When organizations view productivity as undistracted employees solving problems more precisely, they would be well suited to help those employees access knowledge relevant to their problem. Such organizations can encourage focused search by including an easily searched knowledge base – for example, listing all knowledge workers and their areas of expertise, so when one worker needs a solution, they can find who might have the relevant knowledge to help them in a low-cost search. Spatially grouping people in technology areas can also accomplish this – they are more likely to find easy solutions to problems when surrounded by experts in their problem area.

5. If the organization has critical resources and capabilities that it wants to build from, design interactions around experts in that area. This suggestion requires careful knowledge of the core competencies and technical competitive advantages of the firm. Because firms want to exploit their technology development, their development should be guided in part by their existing capabilities (e.g., Penrose 1959). It is conceivable that organizations could design interactions to favor the inclusion of those

most familiar with the firm's core competencies to maximize the use of those capabilities and resources. Interactions could be influenced spatially by distributing these people evenly or putting them in central locations where other employees congregate. Firms could also incentivize these prominent people to interact with and meet more new people and try to understand their problems, in the chance that there is a unique matching of the firm's core competency and this problem. These experts do not have to be technical experts if the firm's core competencies are nontechnical; they could be sales, marketing, or whatever function that firm believes is their best internal resource.

6. Smaller firms should carefully locate themselves geographically. My suggestions for organizations assume that organizations have many knowledge workers to be unknown or unfamiliar to each other. I generally consider cases within a large firm because any employees can legally work together uncomplicated by potential property rights issues of co-developed intellectual property. However, smaller firms can still benefit by adapting these suggestions for use within their broader trusted network of external contacts.

Smaller firms, then, should benefit from geographic proximity to anyone they might work with, for any of my advice above to apply in any direction. For a small firm engaging in technology development, isolation seems negative in all cases. Locating in an area full of diversity of knowledge should lead to more novelty in externally developed collaborations, and locating in a knowledge-focused region might help with development in that one knowledge area. Small firms don’t just face a novelty vs. fast problem-solving

tradeoff that I describe for large firms; the wrong location for small firms likely leads to low levels of *both*.

Implications and opportunities from the rise of "work from anywhere"

There are implications from my research on additional challenges from the rise of the office-less "work from anywhere." To state my position clearly, I believe the importance of proximity between people is not going away. Many technology companies claim success at replacing aspects of proximity because of productivity witnessed during the Covid pandemic. And technology has undoubtedly made advances. But I believe it is folly to compare when most knowledge workers were forced to work remotely – during the Covid pandemic – to a future where proximity is an option. I believe that those who go back to the office will benefit.

However, as I stated before, some roles may flourish remotely, and employees in these roles may benefit from the option. When firms decide to include remote work as an option, which many are, they should ensure that the more focused style of directed search that their remote workers will be engaged in aligns with their organizational goals.

Entrepreneurs are developing new technologies attempting to replace the informal benefits of proximity by introducing randomized online networking events and considering them things like "virtual watercoolers" (Bojinov et al. 2021). These efforts may be a step in the right direction toward replacing the awareness that comes from casual conversations at an office, but I believe more research is needed in this area, especially concerning how firms might implement these treatments outside of experimental settings.

I mention this "virtual watercooler" effort because it speaks to the scope of my research. But I believe if one looks to the full suite of reasons proximity can be valuable, it will be exceedingly difficult to replace each one with technology. But perhaps through these new technologies, we will identify more clearly what we're missing and what we can do without. Until a time in the future when work-from-home entrepreneurs have developed and tested much different technology than exists today, I don't think remote work is good for innovative collaboration or most types of knowledge work in general. Those firms trying technology to replace proximity now will be guinea pigs. I admit I will be excited to see outcomes from these guinea pigs, but I would not recommend any organization I care about to volunteer for this role. The learning process will be painful, and possibly it will prove to be impossible to adapt. Perhaps we will learn that the effects of proximity cannot be replaced for primitive, innately human reasons, like when the person down the hall gives you a kind eye after a bad day. I don't see those interactions being replaced with any technology, ever, and those kinds sof interaction play a significant role in trust, relationship development, and ultimately, collaborations.

Implications and opportunities from the rise of artificial intelligence

The landscape of work is shifting to include an increasing role of artificial intelligence (AI). AI might play a role in a future set of technology aimed at replacing proximity to better enable remote work, but I believe AI can also help guide in-person work with the level of technology that exists today.

We often think of the future of AI in its capacity to replace some worker roles, but I believe AI may be able to serve as a function more often undertaken by managers as

part of a solution to optimize the effect of proximity on innovation and even shape proximate interactions to fit organizational goals.

For example, with a database of detailed knowledge profiles of each employee, AI might match employees with a given set of technical problems with others with complementary skills. AI could organize "suggested" contacts for an individual engaged in search to either make the search process fast or increase variance. AI could also implement ideas involving the absorptive capacity of individuals to determine which people might be able to have a productive conversation in a new area.

Using AI like this is not too futuristic. It's easy to imagine such an implementation of AI in a university setting, where AI can match people who might be interested in each other's research without them knowing about each other in advance. For example, I've had productive conversations about my research with urban planning department faculty and with an architect friend of mine. It turns out, the fields of urban planning and architecture think a lot about chance encounters and proximity too, but differently. It was eye-opening and informative to hear their views. However, I faced very high search costs to access their knowledge. I had to know them, personally, to know I could talk about my research with them. If I opened a directory and saw their expertise written out in the different jargon of those fields, I would never try talking to them about research. But AI is good at categorization and matching patterns – and so AI might have found the connection for me. Some automated categorization and matching would have helped me find and understand the relevance of those people and would likely open my eyes to more relevant fields I didn't have the privilege of knowing people in already. Of course, firms,

governments, regional accelerators, and other institutions could also implement AI in this way. I use the university example only because of how easily I could see myself benefiting from it. AI might help show what doors might be helpful to knock on or what publications to read when you're stumped, even if they're ones you would never have thought of.

AI might also be able to optimize how conceptually "far' a searcher can search. A theme in my dissertation is that conceptually, novelty should increase when you talk to new people and use information further from what you're used to. But, this effect has boundary conditions. For example, if I spoke with an expert on botany, I would not get very far; I am a terrible gardener, and I know nothing about plants. You have to understand each other for talking to new people to result in a collaboration. More random interactions are probably pretty inefficient. I could talk to hundreds of botanists and never produce anything from it because I would not learn from them. There is (seemingly, to me, now) minimal overlap in our fields. This boundary condition shows another opportunity for AI tools to assist the matching process in the future. AI might optimize for a searcher, given their absorptive capacity, how "far" out they can productively search. By suggesting people on the edge of that boundary for people in a group, AI could encourage the most novelty possible for that set of individuals.

Nobody knows what shape the future of work will take, but I believe the effects of proximity between people will continue to be a significant influence on collaborations and innovation. I plan to continue research in this area, and I hope my research can help shape how organizations plan, adapt, and thrive, whatever happens next.

BIBLIOGRAPHY

Adelman DE, Deangelis KL (2005) *Patent Metrics: The Mismeasure of Innovation in the Biotech Patent Debate*

Agrawal A, Cockburn I, McHale J (2006) Gone but not forgotten: knowledge flows, labor mobility, and enduring social relationships. *J. Econ. Geogr.* 6(5):571–591.

Agrawal A, Galasso A, Oettl A (2017) Roads and innovation. *Rev. Econ. Stat.* 99(3).

Agrawal A, Kapur D, Mchale J (2008) How do spatial and social proximity influence knowledge flows? Evidence from patent data. *J. Urban Econ.* 64:258–269.

Alcácer J, Chung W (2007) Location Strategies and Knowledge Spillovers. *Manage. Sci.* 53(5):760–776.

Allen JT (1977) *Managing the Flow of Technology: Technology Transfer and the Dissemination of Technological Information Within the R&D Organization*

Almeida P, Kogut B (1997) The Exploration of Technological Diversity and Geographic Localization in Innovation: Start-Up Firms in the Semiconductor Industry. *Small Bus. Econ.* 9(1):21–31.

Almeida P, Phene A, Li S (2015) The Influence of Ethnic Community Knowledge on Indian Inventor Innovativeness. *Organ. Sci.* 26(1):198–217.

Andrews M (2020) Bar Talk: Informal Social Interactions, Alcohol Prohibition, and Invention. *SSRN Electron. J.*

Anon (2020) The Atlantic Festival Ideas Stage - Day 1 (feat. Tim Cook and Stacey Abrams). *YouTube*. Retrieved (May 18, 2021), https://www.youtube.com/watch?v=y06UTlSn2hw.

Arzaghi M, Henderson JV (2008) Networking off Madison Avenue. *Rev. Econ. Stud.* 75:1011–1038.

Audia PG, Goncalo JA (2007) Past Success and Creativity over Time: A Study of Inventors in the Hard Disk Drive Industry. *Manage. Sci.* 53(1):1.

Bayham J, Kuminoff N V., Gunn Q, Fenichel EP (2015) Measured voluntary avoidance behaviour during the 2009 A/H1N1 epidemic. *Proc. R. Soc. B Biol. Sci.* 282(1818).

Bell GG, Zaheer A (2007) Geography, Networks, and Knowledge Flow. *Organ. Sci.* 18(6):955–972.

Benenson JF, Markovits H, Fitzgerald C, Geoffroy D, Flemming J, Kahlenberg SM, Wrangham RW (2009) Males' Greater Tolerance of Same-Sex Peers. *Psychol. Sci.* 20(2):184–190.

Benner MJ, Tushman ML (2003) Exploration, Exploitation and Process Management: the Productivity Dilemma Revisited. *Acad. Manag. J.* 28(2):238–256.

Bernstein ES (2012) The transparency paradox: A role for privacy in organizational learning and operational control. *Adm. Sci. Q.* 57(2):181–216.

Bettencourt LM, Lobo J, Strumsky D (2007) Invention in the city: Increasing returns to patenting as a scaling function of metropolitan size. *Res. Policy* 36(1):107–120.

Billinger S, Stieglitz N, Schumacher TR (2014) Search on Rugged Landscapes: An Experimental Study. *Organ. Sci.* 25(1).

BioCrossroads (2009) Warsaw, Indiana: The Orthopedics Capital of the World.

Bojinov I, Choudhury P, Lane JN (2021) Virtual Watercoolers: A Field Experiment on Virtual Synchronous Interactions and Performance of Organizational Newcomers. *HBS Work. Pap.* 21(125).

Borgatti SP, Cross R (2003) A Relational View of Information Seeking and Learning in Social Networks. *Manage. Sci.* 49(4):432–445.

Bormet MJ (2001) *A History of the City of Warsaw - Preserving the Past for the Future* (Kosciusko County Historical Society, Bourbon, Ind.).

Boudreau KJ, Brady T, Ganguli I, Gaule P, Guinan E, Hollenberg A, Lakhani KR (2017) A Field Experiment on Search Costs and the Formation of Scientific Collaborations. *Rev. Econ. Stat.* 99(4):565–576.

Van Den Bulte C, Moenaert RK (1998) The Effects of R&D Team Co-location on Communication Patterns among R&D, Marketing, and Manufacturing. *Manage. Sci.* 44(11):S1–S18.

Burt RS (2004) Structural Holes and Good Ideas. *Am. J. Sociol.* 110(2):349–99.

Caley P, Philp DJ, Mccracken K (2008) Quantifying social distancing arising from pandemic influenza. *J. R. Soc. Interface* 5:631–639.

Catalini C (2018) Microgeography and the Direction of Inventive Activity. *Manage. Sci.* 64(9):4348–4364.

Choudhury P, Foroughi C, Larson B (2019) Work-from-anywhere: The Productivity Effects of Geographic Flexibility. *SSRN Electron. J.*

Chown JD, Liu CC (2015) Geography and power in an organizational forum: Evidence from the U.S. Senate Chamber. *Strateg. Manag. J.* 36(2):177–196.

Coff R, Kryscynski D (2011) Drilling for Micro-Foundations of Human Capital Based Competitive Advantages. *J. Manage.* 37(5):1429–1443.

Cohen WM, Levinthal DA (1990) Absorptive Capacity: A New Perspective on Learning and Innovation. *Adm. Sci. Q.* 35(1):128.

Crescenzi R, Nathan M, Rodriguez-Pose A (2016) Do inventors talk to strangers? On proximity and collaborative knowledge creation. *Res. Policy* 45(1):177–194.

Crow D (2021) How mRNA became a vaccine game-changer. *Financial Times* (May 12).

D'Onfro J (2015) Steve Jobs Designing Pixar Office. *Business Insider* (March 20).

Duffy K (2021) 40% of Workers May Quit If Made to Return to Office Full Time: Survey. *Business Insider* (June 2).

Duguid M (2011) Female tokens in high-prestige work groups: Catalysts or inhibitors of group diversification? *Organ. Behav. Hum. Decis. Process.* 116(1):104–115.

Feldman MP, Florida R (1994) The Geographic Sources of Innovation: Technological Infrastructure and Product Innovation in the United States. *Ann. Assoc. Am. Geogr.* 84:210–229.

Fenichel EP, Castillo-Chavez C, Ceddia MG, Chowell G, Gonzalez Parra PA, Hickling GJ, Holloway G, et al. (2011) Adaptive human behavior in epidemiological models. *PNAS* 108(15).

Ferguson N (2007) Capturing human behaviour. *Nature* 446(7137):733–733.

Fleming L (2001) Recombinant Uncertainty in Technological Search. *Manage. Sci.* 47(1):117–132.

Fleming L (2002) Finding the organizational sources of technological breakthroughs: the story of Hewlett-Packard's thermal ink-jet. *Ind. Corp. Chang.* 11(5):1059–1084.

Fleming L, King C, Juda AI (2007) Small Worlds and Regional Innovation. *Organ. Sci.* 18(6).

Funk S, Salathé M, Jansen VAA (2010) Modelling the influence of human behaviour on the spread of infectious diseases: a review. *J. R. Soc. Interface* 7(50):1247–56.

Gallagher (2020) *Gallagher Covid-19 Pulse Survey: Work in a New Normal*

Gilbert B (2020) Microsoft will let their employees keep working from home. *World Economic Forum*

Gross T, Dommar D'lima CJ, Blasius B (2006) Epidemic Dynamics on an Adaptive Network. *Phys. Rev. Lett.* 96(208701).

Hall D, Lecuona R, Cummings JN (2017) Walk and Talk:The Interplay of Work Dependencies, Spatial Distance, and Face-to-Face Communication. *Acad. Manag. Proc.* 2017(1):17302.

Henderson R, Cockburn I (1994) Measuring competence? Exploring firm effects in pharmaceutical research. *Strateg. Manag. J.* 15:63–84.

Hess A (2017) *The science and design behind Apple's innovation-obsessed new workspace*

Ingram D (2020) California's Bay Area may require telecommuting, even after the pandemic wanes. *NBC News - Tech News*. Retrieved (May 18, 2021),

https://www.nbcnews.com/tech/tech-news/california-s-bay-area-likes-telecommuting-so-much-it-might-n1240898.

Jacobs J (1969) *The economy of cities* (Random House).

Jaffe A, Trajtenberg M (2002) *Patents, citations, and innovations: A window on the knowledge economy* (MIT Press).

Jaffe AB, Trajtenberg M, Henderson R (1993) Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *Q. J. Econ.* 108(3):577–598.

Jiles M (2007) Estimates About Journey to Work from the 2005 ACS, C2SS, and Census 2005.

Jones JH, Salathé M (2009) Early Assessment of Anxiety and Behavioral Response to Novel Swine-Origin Influenza A(H1N1). *PLoS One* 4(12):e8032.

Kabo FW, Cotton-Nessler N, Hwang Y, Levenstein MC, Owen-Smith J, Ross SM (2014) Proximity effects on the dynamics and outcomes of scientific collaborations. *Res. Policy* 43:1469–1485.

Kane AA, Argote L, Levine JM (2005) Knowledge transfer between groups via personnel rotation: Effects of social identity and knowledge quality. *Organ. Behav. Hum. Decis. Process.* 96(1):56–71.

Kleinbaum AM, Stuart TE, Tushman ML (2013) Discretion Within Constraint: Homophily and Structure in a Formal Organization. *Organ. Sci.* 24(5):1316–1336.

Kogut B, Zander U (1992) Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology. *Organ. Sci.* 3(3):383–397.

Lane JN, Ganguli I, Gaule P, Guinan E, Lakhani K (2020) Engineering Serendipity: When Does Knowledge Sharing Lead to Knowledge Production? *Strateg. Manag. J.*:smj.3256.

Lazer D, Friedman A (2007) The Network Structure of Exploration and Exploitation. *Adm. Sci. Q.* 52:667–694.

Lee S (2019) Learning-by-Moving: Can Reconfiguring Spatial Proximity Between Organizational Members Promote Individual-level Exploration? *Organ. Sci.*

Levin DZ, Cross R (2004) The strength of weak ties you can trust: The mediating role of trust in effective knowledge transfer. *Manage. Sci.* 50(11):1477–1490. (November 1).

Levinthal DA, March JG (1993) The myopia of learning. *Strateg. Manag. J.* 14(S2):95–112.

March J (1991) Exploration and Exploitation in Organizational Learning. *Organ. Sci.* 2(1).

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

Marshall A (1920) *Principles of Economics (Revised ed.)* (Macmillan, London).

McEvily B, Perrone V, Zaheer A (2003) Trust as an Organizing Principle. *Organ. Sci.* 14(1):91–103.

McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a Feather: Homophily in Social Networks. *Annu. Rev. Sociol.* 27(1):415–444.

Meyer D (2020) Working from home could become a legal right in Germany. *Fortune* (October 1).

Nahapiet J, Ghoshal S (1998) Social Capital, Intellectual Capital, and the Organizational Advantage. *Acad. Manag. Rev.* 23(2):242–266.

Nelson RR, Winter SG (1982) *An Evolutionary Theory of Economic Change* (Belknap Press, Cambridge, Mass).

Nooteboom B (2000) Learning by Interaction: Absorptive Capacity, Cognitive Distance and Governance. *J. Manag. Gov.* 4(1/2):69–92.

Penner-Hahn J, Shaver JM (2005) Does International Research and Development Increase Patent Output ? an Analysis of Japanese Pharmaceutical Firms. *Strateg. Manag. J.* 140(May 2000):121–140.

Penrose E (1959) *The Theory of the Growth of the Firm* (Oxford University Press).

Public Affairs UC Berkeley (2020) Nobel laureate Jennifer Doudna: Gene editing could make a better future. *Berkeley News*

Putnam RD (2000) *Bowling alone: The collapse and revival of American democracy*

Reagans R (2011) Close Encounters: Analyzing How Social Similarity and Propinquity Contribute to Strong Network Connections. *Organ. Sci.* 22(4):835–849.

Roche MP (2019) Taking Innovation to the Streets: Micro-geography, Physical Structure and Innovation. *Rev. Econ. Stat.*

Roche MP, Catalini C, Oettl A (2020) *Entrepreneurs (Co-)Working in Close Proximity: Impacts on Technology Adoption and Startup Performance Outcomes*

Rogers M (2004) Networks, Firm Size and Innovation. *Small Bus. Econ.* 22:141–153.

Rosenkopf L, Almeida P (2003) Overcoming local search through alliances and mobility. *Manage. Sci.*

Rosenthal SS, Strange WC (2004) Chapter 49 – Evidence on the Nature and Sources of Agglomeration Economies. *Handb. Reg. Urban Econ.* 2119–2171.

Ruef M, Aldrich HE, Carter NM (2003) The Structure of Founding Teams: Homophily, Strong Ties, and Isolation among U.S. Entrepreneurs. *Am. Sociol. Rev.* 68(2):195.

Saxenian A (1996) *Regional Advantage* (Harvard University Press).

Schumpeter J (1939) *Business Cycles* (McGraw-Hill Book Company, Inc., New York).

Shaver JM, Flyer F (2000) Agglomeration Economies, Firm Heterogeneity, and Foreign Direct Investment in the United States. *Strateg. Manag. J.* 21(12):1075–1193.

Simon HA (1991) Bounded Rationality and Organizational Learning. *Organ. Sci.* 2(1):125–134.

Sorenson O, Rivkin JW, Fleming L (2006) Complexity, networks and knowledge flow. *Res. Policy*.

Sorenson O, Stuart TE (2008) Bringing the Context Back In: Settings and the Search for Syndicate Partners in Venture Capital Investment Networks. *Adm. Sci. Q.* 53:266–294.

Spice B (2003) Nobel Prize for MRI began with a burger in New Kensington. *Pittsburgh Post-Gazette* (October 6).

Squicciarini M, Dernis H, Criscuolo C (2013) Measuring Patent Quality: Indicators of Technological and Economic Value. *OECD Sci. Technol. Ind. Work. Pap.* (03):70.

Storper M, Venables AJ (2004) Buzz: face-to-face contact and the urban economy. *J. Econ. Geogr.* 4:351–370.

Tannenbaum MB, Hepler J, Zimmerman RS, Saul L, Jacobs S, Wilson K, Albarracín D (2015) Appealing to fear: A meta-analysis of fear appeal effectiveness and theories. *Psychol. Bull.* 141(6):1178–1204.

Torras RL, Cummings JN (2020) Where the boss sits matters: Work interactions among employees and physical proximity in the office. *80th Annu. Meet. Acad. Manag. 2020 Underst. Incl. Organ. AoM 2020*. (Academy of Management).

Trajtenberg M, Henderson R, Jaffe A (1997) University versus corporate patents: A window on the basicness of invention. *Econ. Innov. New Technol.* 5(1):19–50.

Turner JC, Hogg MA, Oakes PJ, Reicher SD, Wetherell MS (1987) *Rediscovering the social group: A self-categorization theory.* (Basil Blackwell, Cambridge, MA, US).

Tushman ML, Anderson P (1986) Technological Discontinuities and Organizational Environments. *Adm. Sci. Q.* 31(3):439–465.

USPTO (2017) Manual of Patent Examining Procedure (MPEP). (USPTO).

Uzzi B (1997) Social Structure and Competition in Interfirm Networks: The Paradox of Embeddedness. *Adm. Sci. Q.* 42(1):35–67.

Vakili K, Zhang L (2018) High on creativity: The impact of social liberalization policies on innovation. *Strateg. Manag. J.* 39(7):1860–1886.

Young ME, Norman GR, Humphreys KR (2008) Medicine in the Popular Press: The

Influence of the Media on Perceptions of Disease. *PLoS One* 3(10):e3552.