HARVARD | BUSINESS | SCHOOL

Engineering Serendipity: When Does Knowledge Sharing Lead to Knowledge Production?

Jacqueline N. Lane
Ina Ganguli
Patrick Gaule
Eva Guinan
Karim Lakhani

Working Paper 20-058

Engineering Serendipity: When Does Knowledge Sharing Lead to Knowledge Production?

Jacqueline N. Lane
Harvard Business School

Ina Ganguli
University of Massachusetts, Amherst

Patrick Gaule
University of Bath

Eva Guinan
Harvard Medical School

Karim Lakhani
Harvard Business School

Working Paper 20-058

Funding for this research was provided in part by Harvard Business School.

Engineering Serendipity:

When Does Knowledge Sharing Lead to Knowledge Production?

Jacqueline N. Lane[1,5], Ina Ganguli[2,5], Patrick Gaule[3,5], Eva Guinan[4,5], and Karim Lakhani[1,5]

[1] Harvard Business School
[2] University of Massachusetts, Amherst
[3] University of Bath
[4] Harvard Medical School and Dana-Farber Cancer Institute
[5] Laboratory for Innovation Science at Harvard

Engineering Serendipity:

When Does Knowledge Sharing Lead to Knowledge Production?

Abstract

We investigate how knowledge similarity between two individuals is systematically related to the likelihood that a serendipitous encounter results in knowledge production. We conduct a natural field experiment at a medical research symposium, where we exogenously varied opportunities for face-to-face encounters among 15,817 scientist-pairs. Our data include direct observations of interaction patterns collected using sociometric badges, and detailed, longitudinal data on the scientists' post-symposium publication records over six years. We find that interacting scientists acquire and create more knowledge when they share some overlapping research interests, but are less likely to cite each other's work when they are from similar fields. Our findings reveal both collaborative and competitive effects of knowledge similarity on knowledge production outcomes.

Keywords: knowledge production, innovation, knowledge sharing, knowledge similarity, natural field experiment

1. INTRODUCTION

In 2013, cell biologist William Earnshaw of the University of Edinburgh happened to attend the same academic conference as systems biologist Job Dekker of the University of Massachusetts Medical School and computational biologist Leonid Mirny from the Massachusetts Institute of Technology. By chance, Earnshaw attended a presentation by Dekker and Mirny about their joint work on mitotic chromosomes, during which he became convinced that his lab could provide bench methods to improve Dekker and Mirny's computational models. Earnshaw approached Dekker and Mirny after the talk; their conversation evolved into a three-lab collaboration and a 2018 publication in *Science* (Pain, 2018).

This anecdotal example suggests that serendipitous encounters can play a role in innovation, perhaps as a result of exposing individuals to unfamiliar sources of information that can be combined with their own knowledge stock and lead to new discoveries (Fleming, Mingo, & Chen, 2007; Uzzi, Mukherjee, Stringer, & Jones, 2013). However, the likelihood that serendipitous encounters lead to successful knowledge production remains poorly understood often because the process is uncertain, complex and relatively understudied. In particular, even if an encounter eventually produces new knowledge, this process may occur slowly and only emerge after a number of years (Catalini, 2018; Fleming, 2001). A significant degree of knowledge within organizations is tacit and is not easily transferred (Argote, 1999; Hansen, 1999; Nonaka, 1994). Knowledge creation via new collaborations can be costly (Boudreau et al., 2017, Dahlander & McFarland, 2013), requiring coordinated effort, alignment of incentives, establishment of trust, generation of creative synergy, and "matching" criteria, such as personality and scheduling compatibility (Azoulay, Ding, & Stuart, 2009; Catalini, 2018). Even if knowledge transfer or creation is successful, knowledge tends be slow to diffuse through interpersonal networks (Fleming et al. 2007; Singh 2005).

The question thus arises of when serendipitous encounters are more likely to facilitate useful idea exchange and knowledge production. Previous literature tends to focus on the factors that may increase the amount or intensity of short-term communication and interaction between individuals, rather than the implications of serendipitous encounters on knowledge production (e.g., Allen, 1977; Kleinbaum, Stuart, & Tushman, 2013). Given that geographic proximity increases the likelihood of serendipitous encounters (Catalini, 2018), recent research has begun to examine the effects of organizational redesign on knowledge production (Catalini, 2018; Fang, Lee, & Schilling, 2010; Lee, 2019). Although it is possible that organizational redesign and greater geographic proximity between individuals can increase knowledge production, most studies do not directly address how knowledge sharing between two individuals affects the path-dependent nature of the knowledge production process (Carlile, 2004; Hargadon & Sutton, 1997). Since prior work tends to focus on a single knowledge production outcome (Phelps, Heidl, & Wadhwa, 2012), often on the factors that impede knowledge transfer (Carlile 2004; Hansen, 1999; Szulanski, 1996; Uzzi, 1997), it is unclear how knowledge sharing may also affect knowledge creation and diffusion. Hence, we know relatively little about whether serendipitous encounters between individuals actually lead to meaningful idea exchange that can be applied to one's own tasks, or whether these interactions can spark new collaborations and broader diffusion of ideas over time. The goal of our work is to address this gap in literature by examining to what extent a systematic relationship exists between serendipitous encounters and knowledge production outcomes.

We posit that an effective encounter hinges not just on the serendipity of the interaction but also on the existence of common knowledge that individuals use to make sense of each other's specialized knowledge and connect it to what they already know. Common knowledge enhances the ability for new knowledge to be assimilated into the concepts, objects and patterns that are already present in people's cognitive structures (Bower & Hilgard, 1981; Cohen & Levinthal, 1990). When new information is related to prior knowledge constructs, it is more readily absorbed, integrated, and applied in new settings (Carlile, 2004; Kogut & Zander, 1992); without sufficient prior knowledge, individuals may have difficulty integrating new knowledge (Rosenkopf & Almeida, 2003). In this paper, we argue that there are two essential characteristics to people's common knowledge: *field similarity* in their educational backgrounds and training, and *intellectual similarity* in their interests, passions and pursuits. Serendipitous encounters are more likely to be effective when people share some knowledge similarity in both their field and intellectual interests.

To systematically investigate the relationships between serendipitous encounters and knowledge production requires a rich longitudinal dataset that directly observes exogenous face-to-face knowledge sharing between two individuals and subsequent knowledge production outcomes. To draw causal inferences, we designed and executed a natural field experiment at a research symposium on advanced imaging in early 2012 at an elite academic medical center. We chose this setting due to the centrality of knowledge production to their organizational and individual performance goals and the prevalence of knowledge sharing norms (Dahlander & McFarland, 2013). We created exogenous variation in opportunities for information-rich, face-to-face encounters between cross-disciplinary scientists interested in applying for an internal grant program promoting the development of advanced imaging solutions to address an unmet clinical need. Using electronic sociometric badges (Kim, McFee, Olguin, Waber, & Pentland, 2012), we collected a unique dataset of fine-grained, live interactions between pairs of scientists who were randomly assigned to be in the same symposium room for knowledge sharing. We then combined this with a rich longitudinal dataset on long-run knowledge production outcomes from the scientists' written publications over a six-year period, where we linked scientists' pairwise keywords, collaborations and forward citations to examine the extent that the encounters led to knowledge transfer, creation and diffusion, respectively. This design enabled us to causally and systematically identify the relationships between knowledge sharing and knowledge production to explain why some serendipitous encounters resulted in knowledge production and others did not.

We find both cooperative and competitive effects of serendipitous knowledge sharing encounters on knowledge production. On one hand, knowledge sharing led to greater knowledge transfer and knowledge creation when people had some overlapping intellectual and field interests, respectively. Although knowledge sharing had limited impact on knowledge transfer among pairs with either few or many overlapping interests, we observe a two-fold increase in knowledge transfer (of Medical Subject Headings or MeSH keywords) for moderately similar pairs, suggesting an inverted U-shaped relationship between knowledge overlap and knowledge transfer. Similarly, we find that knowledge sharing between pairs with some common knowledge led to 1.2 more coauthored peer-reviewed publications, compared to a 0.98 decrease in copublications among high knowledge overlapping pairs, which suggests that the knowledge sharing intervention reduced the search costs associated with forming synergistic collaborations. By comparing the knowledge drivers of near-term collaborations and long-term collaborations, we find that common knowledge is more critical for the former. On the other hand, knowledge sharing reduces knowledge diffusion between people from the same field, with interacting pairs citing each other between 2.8 to 6.7 times less than non-interacting pairs from more distant fields.

We aim to make several contributions to the literature. First, we contribute to the literature on search costs and information frictions associated with the knowledge production process

(Agrawal & Goldfarb, 2008; Boudreau et al., 2017; Catalini, 2018; Haas & Hansen, 2007; Hansen, 1999), by focusing on the extent that "engineered serendipity" can mitigate some of these costs. we emphasize the role of knowledge similarity in shaping different components of knowledge production, and over different time horizons. Most prior research has predominantly focused on one component or a single time frame (e.g., Adams, Grant, Black, Clemmons, & Stephan, 2005; Dahlander & McFarland, 2013; Hansen, 1999; Szulanski, 1996). Our work suggests that knowledge dissimilarity remains a barrier to knowledge production over time, as pairs are more productive when they share sufficient common knowledge. However, the magnitude of these frictions are likely to be more difficult to overcome in the short-run compared to the long-run.

Second, we make contributions to the literature on temporary colocation, and the value of conferences, symposia and similar events on the direction of inventive activity on the knowledge frontier and the emergent patterns of collaborative activity (Biancani, Dahlander, & McFarland, 2014; Boudreau et al., 2017; Campos, Leon, McQuillin, 2018; Catalini, Fons-Rosen, & Gaule, 2020; Chai & Freeman, 2019). Prior research has been limited in its ability to observe actual acts of knowledge sharing between pairs, and its implications on knowledge production. Our study indicates that even brief, information-rich encounters at conferences can benefit knowledge production, with the potential to alter people's research directions and collaboration networks.

Third, we make methodological contributions by highlighting the benefits of long-term studies that amalgamate multiple forms and uses of data. Prospective experiments can support multiple lines of investigation involving both near-term and long-term outcomes that may not be possible in retrospective, archival studies and suggests the use of multiple sources of data for unpacking the dynamics of knowledge production.

2. THEORY AND HYPOTHESES

2.1. Knowledge Production and Knowledge Sharing

According to the knowledge-based view of the firm, knowledge production is a multi-component process that involves the transfer, creation and diffusion of knowledge (Grant, 1996; Kogut & Zander, 1992; Nonaka, 1994; Spender, 1996). Knowledge transfer involves the movement of facts, relationships, and insights from one setting to another, and becomes evident when the experience acquired by an individual, group or organization in one setting is applied in another setting (Argote, 1999; Hansen, 1999; Szulanski, 1996). Knowledge creation refers to the generation of facts, relationships and insights to solve new problems (Arrow, 1962; Nonaka, 1994). The emphasis on *new* knowledge is what distinguishes the process of knowledge transfer from creation. Because new knowledge is not held by anyone prior to its creation, it cannot be transferred and applied directly (McFadyen & Canella 2004). Conversely, knowledge diffusion refers to the process through which transferred or newly created knowledge is disseminated and used by other individuals, groups and organizations (Fleming et al., 2007; Singh, 2005). Knowledge diffusion is a critical component of the knowledge production process because it reduces duplication of effort and promotes efficiency by demarcating what is known from what is yet to be explored on the knowledge frontier (Boudreau & Lakhani, 2015).

Critical to the efficiency of knowledge production is the process of knowledge sharing between individuals possessing different types of knowledge (Grant, 1996; Kogut & Zander, 1992; Nonaka, 1994), and their ability to learn from these interactions (Levinthal & March, 1993; Simon, 1991). Common knowledge between partners is important because it enables individuals to absorb the aspects of their knowledge sets that they do not hold in common, or that are unique to each individual (Carlile, 2004; Cohen & Levinthal, 1990; Maurer & Ebers, 2006). Building on prior work, we propose that there are at least two separate pathways through which people develop their knowledge bases: their field and intellectual specialties. Because of the path-dependent nature of knowledge (Hargadon & Sutton, 1997), the overlap in two individuals' field and intellectual specialties may have different implications on knowledge production. Table 1 provides a summary of our knowledge constructs.

2.2. Two Knowledge Bases of Similarity: Field Specialty and Intellectual Specialty

Field specialties are developed through an individual's educational background, skills and training (Bourdieu, 1984; Bechky, 2011; Haas & Park, 2010). In biomedical research, which is our focal context of study, there are multiple medical field specialties, such as neurology, radiology or oncology that focus on investigations of the biological process and the causes of specific diseases. Organizations often structure their departmental and divisional memberships around field specialties (Biancani et al., 2014). In universities and academia, this means that field specialties are often the primary channels through which resources, such as salaries, funding, tenure, offices and laboratory space, are apportioned to faculty (Biancani, Dahlander, McFarland, & Smith, 2018). For these reasons, field specialties provide professional reference groups (Haas & Park, 2010) for members to identify appropriate behavior, rigor of scholarship, career goals and aspirations (Abbott, 2001). In contrast, intellectual specialties are closely aligned to personal interests and passions, that evolve over time. Individuals can have memberships in multiple intellectual specialties depending on their current pursuits and priorities (Caza, Moss, & Vough, 2018). For example, in biomedical research, intellectual specialties include research topics such as aging, addictions, Alzheimer's disease, brain injury, sleep, and stem cells; each of these intellectual specialties is associated with memberships in professional reference groups (Haas & Park, 2010) for people motivated by common problems, methods, trends, and processes. These intellectual interests can be specific to a field's research topics, such as Alzheimer's disease, which is typically studied by neurologists, or span different fields, such as stem cell research.

In short, field and intellectual specialties shape people's identities, language, tastes and affiliations through professional memberships (Haas & Park, 2010). Individuals will likely adopt multiple professional identities depending on their field and intellectual memberships: this means that people from different field specialties can share overlapping intellectual interests, while individuals from the same field specialty can have divergent intellectual interests, because field and intellectual interests are two distinct dimensions of individuals' knowledge constructs.

In distinguishing how field and intellectual overlap shape knowledge production, knowledge similarity describes the distance between the field specialties and intellectual specialties of two knowledge sharing partners. Individuals are more likely to recognize and absorb new ideas when they already have some existing expertise and find it more difficult when the ideas are outside their realm of expertise (Carlile, 2004; Cohen & Levinthal 1990). Put differently, even though distant knowledge tends to be novel and valuable (Jeppesen & Lakhani, 2010, Leahey, Beckman, & Stanko, 2017), people are more likely to experience challenges communicating with one another, and may not recognize the value of external information (Cohen & Levinthal, 1990; Grant, 1996). For example, two neurologists are less likely to experience challenges communicating with each other than a neurologist and oncologist pair because they share more similar educational backgrounds, training and clinical expertise. However, there may be greater opportunities for knowledge production between the neurologist and oncologist pair because their field specialties are more distant. Similarly, although two sleep researchers may communicate and comprehend each other with relative ease because they read and conduct similar research (Dahlander & McFarland, 2013), there is greater potential for knowledge production between an aging and a sleep researcher, who may offer greater utility and value to one another to solve problems related to both sleep and aging processes—provided that they have sufficient common knowledge (Bechky, 2011; Grant, 1996). This suggests that there is an inherent trade-off between the greater ease of conversing and absorbing ideas locally from individuals with many overlapping interests and the greater potential opportunities for novel ideas and knowledge transfer, creation, and diffusion from sharing and assessing the knowledge of those with more distant interests.

2.3. Knowledge Transfer and Knowledge Similarity

Knowledge transfer is a two-sided process because it depends on the efforts of a source to share knowledge with a recipient and the recipient's efforts and capacity to acquire, absorb, and learn it (Argote, 1999). Because of greater common knowledge, local knowledge found within groups of similar individuals tends to be more easily transferred than distant knowledge spanning group boundaries (Carlile, 2004; Kogut & Zander 1992; Rosenkopf & Almeida, 2003). That said, more distant knowledge may present non-redundant ideas that benefit learning and acquiring new concepts. Knowledge sharing partners from the same field specialty have similar education and training, meaning that they are more capable of absorbing each other's knowledge. However, these ideas are also more likely to be redundant. Likewise, knowledge sharing partners are more likely to have greater potential for knowledge transfer when they are less intellectually similar (Leahey et al. 2017; Uzzi et al., 2013). Divergent interests create a wider pool of knowledge to draw upon, which allows for multiple perspectives and problem-solving approaches that increase the likelihood of novel solutions and discoveries (Boudreau, Guinan, Lakhani, & Riedl, 2016). Based on these reasons, we expect that individuals with less field or intellectual overlap have a greater potential pool of ideas for knowledge transfer, up until a certain threshold of dissimilarity, after which they will lack sufficient common interests to benefit from knowledge transfer.

Hypothesis 1a (H1a). Field similarity has an inverted U-shaped effect on the relationship between knowledge sharing and knowledge transfer.

Hypothesis 1b (H1b). Intellectual similarity has an inverted U-shaped effect on the relationship between knowledge sharing and knowledge transfer.

2.4. Knowledge Creation and Knowledge Similarity

There is a growing view among innovation scholars that specialization has created a "knowledge burden" hypothesis that has made collaborative work combining the increasingly narrow niches of specialization imperative to moving the knowledge frontier forward (Jones, 2009; Uzzi et al., 2013). Scientific discovery is a process that combines individually focused tasks—such as reading, experimentation, and writing—with social interactions through joint sense-making with others that can spark new discoveries (Boudreau & Lakhani, 2015; Latour & Woolgar, 1979).

That said, there are limits in the types of collaborations that can actually form and these are likely to be constrained by social processes, such as preferences for others with shared attributes. Principles of homophily suggest that people are attracted to others who hold similar values because their interactions are more rewarding and less uncertain (McPherson, Smith-Lovin, & Cook, 2001). When individuals share overlapping interests, conversations tend to be more rewarding due to the benefits of security and mutual attraction (Dahlander & McFarland, 2013). Moreover, joint knowledge creation is a long-term investment in a multiplex relationship (Uzzi, 1997), where partners engage in joint problem-solving and spend time together discussing, reflecting and interacting to achieve mutual benefit (McFadyen & Cannella, 2004).

Based on these arguments, new collaborations will require potential partners to share overlapping field interests, such as similarities in their education, training, clinical and/or laboratory experience, which establish common knowledge, and create similar incentives and publication requirements for promotion and tenure (Stephan, 1996). Similarly, potential collaborators are likely to benefit from some intellectual overlap because commonalities in research interests and passions aid with communication and mutual understanding. When individuals share no intellectual overlap, their interests may be too disparate to establish common research interests but when they share a high degree of intellectual similarity, they may be substitutes for each other and offer few complementarities (Dahlander & McFarland, 2013). Knowledge sharing would enable partners with some field or intellectual overlap to identify potential complementarities beyond those available to individuals with complete knowledge overlap. Accordingly, we expect that increasing field and intellectual similarity should benefit knowledge creation up to a threshold, over which there are decreasing marginal returns on greater levels of knowledge similarity.

Hypothesis 2a (H2a). Field similarity has an inverted U-shaped effect on the relationship between knowledge sharing and knowledge creation.

Hypothesis 2b (H2b). Intellectual similarity has an inverted U-shaped effect on the relationship between knowledge sharing and knowledge creation.

Another pathway that may lead to joint knowledge creation relies on prior collaborations and tie persistence (Hasan & Koning, 2019; Ingram & Morris, 2007; Zhang & Guler, 2019). We examine a specific type of tie persistence, namely the likelihood that an elemental collaboration persists to create a complete knowledge product. In science, grant coapplications and copublications represent to two essential but opposing ends of knowledge creation, from idea-generation to final output (McFadyen & Cannella, 2004). Similar to publications, grants are essential to scientists' research productivity and often used as evaluative criteria for important decisions, such as promotion, tenure, awards and recognition (Dahlander & McFarland, 2013;

Stephan, 1996). The knowledge similarity requirements that improve the likelihood that elemental collaborations persist into copublications may differ from the processes that shape long-term collaborations on peer-reviewed research publications. In addition to being at the open end of research inquiry, grant applications tend to have finite submission deadlines and these can impose resource, attention, coordination and other constraints on potential partners (Dahlander, O'Mahony & Mann, 2016). Hence, potential partners may turn to other individuals with whom they share significant knowledge overlap, due to fewer costs and frictions associated with idea generation and developing a viable core concept. Therefore, the effects of common knowledge may be even more critical to shaping elemental collaborations. Due to these reasons, we expect that elemental collaborations are more likely to persist into a final product of knowledge creation when knowledge sharing partners have significant knowledge overlap in terms of their field and intellectual interests.

Hypothesis 2c (H2c). An elemental collaboration is more likely to result in knowledge creation as the field similarity between two individuals increases.

Hypothesis 2d (H2d). An elemental collaboration is more likely to result in knowledge creation as the intellectual similarity between two individuals increases.

2.5. Knowledge Diffusion and Knowledge Similarity

Among research scientists, forward citations to others' publications are a primary means for diffusing knowledge. Beyond knowledge diffusion, forward citations constitute a critical means of social recognition for acknowledging the contributions of predecessors (Merton [1942], 1973) and tracing the path of scientific discovery and diffusion (Stephan, 1996). Some scholars have argued that the number of citations an article or body of work has received is perhaps the most common way to measure the importance of an individual's contribution to science (Stephan, 1996). According to this view, citations have been deemed a critical currency in the cycles of scientific credit (Latour & Woolgar, 1979), both driving research impact (e.g., H-index, i-10 index, impact factor) and constituting the basis of reward systems in science, including promotion, status, funding, peer esteem, honors and awards (Boudreau & Lakhani, 2015).

There are generally two views of how knowledge is diffused in science: openness and secrecy. According to the Mertonian norms of communalism or "openness," publication enables scientists to establish priority of discovery and allows them to be the first to communicate an advance in knowledge (Merton [1942], 1973). However, publication also marks an important transition point in the diffusion process, because it means that a scientist has agreed to relinquish his or her property rights over a discovery and allow others to freely use this knowledge (Merton [1942], 1973). Thus, publication promotes the open diffusion of scientific knowledge, as long as scientists' own internal agents (i.e., other scientists) appropriately recognize and diffuse their work (Boudreau & Lakhani, 2015; Stephan, 1996). We would therefore expect partners with greater knowledge similarity to be more likely to absorb and diffuse each other's knowledge.

On the other hand, social recognition is a discretionary act among scientists, and strong evidence points to the existence of counter-norms promoting secrecy, competition, and information withholding (Haas & Park 2010; Haeussler, Jiang, Thursby, & Thursby, 2014). Scientists often compete for similar resources, funding, and recognition. Given limited resources, scientists need assurance that they will be remunerated for openly diffusing others' knowledge (Hagstrom, 1974; Murray & O'Mahony, 2007; Reschke, Azoulay, & Stuart, 2018).

On balance, the evidence suggests that scientists do not unequivocally diffuse each other's publications. Among scientists with greater knowledge overlap, knowledge sharing may heighten the competition effect, particularly as interaction may emphasize the commonalities between them.

Because scientists compete for priority (Merton [1942], 1973), recognition (Campbell et al., 2002), funding and resources (Stephan, 1996), and promotion criteria (Blumenthal et al., 1997), social recognition of peers would not only highlight similarities but may also detract attention and resources away from the focal scientist's own work (Campanario & Acedo, 2007). We expect that partners sharing high field or intellectual similarity may refrain from citing each other's research to outperform their peers.

Hypothesis 3a (H3a). Knowledge sharing is less likely to lead to knowledge diffusion as the field similarity between individuals increases.

Hypothesis 3b (H3b). Knowledge sharing is less likely to lead to knowledge diffusion as the intellectual similarity between individuals increases.

3. EXPERIMENTAL METHODS

In an ideal setting, all of the prior interactions and efforts that go into knowledge production would be fully observable to scholars to theorize and validate through empirical observations. The reality, however, is that the vast majority of prior scholarly work concerned with how knowledge is produced has primarily focused on *observed* outcomes—e.g., papers, funding, patents, citations, team structure, etc.—to draw inferences about the mechanisms underlying the knowledge production process (Dahlander & McFarland, 2013; Fleming et al., 2007; Staw & Epstein, 2000). For example, research on scientists will include the papers they publish, the collaborators they have worked with, and the knowledge that they have developed and have diffused through citations. However, a concern with relying on published trace data is that it masks all of the work and activities that occurred prior to knowledge production, such as the entire risk set of alters an individual may have interacted with prior to settling on a particular team. The empirical shortcomings of not being able to directly observe knowledge sharing and the drivers of knowledge production may insert biases in our inferences, such as self-selection and survivor bias. Some work aims to address these concerns by extending the risk set to unsuccessful collaborations (e.g., non-awarded grant applications) to examine quality-adjusted research output of published (i.e., visible) trace data (Arora & Gambardella, 2005; Ayoubi, Pezzoni, & Visentin, 2019; Ganguli, 2017). That said, unobserved heterogeneity between the two groups remains a persistent challenge.

A feasible alternative is to design a field experiment that enables the capture of data around interactions between scientists and overcomes concerns around endogeneity of affiliation, team formation, and knowledge exposure by randomizing the encounters that the scientists have with each other. The benefit of this approach can provide causal explanations about the factors that impact knowledge production, weighed against the challenges of drawing inferences with smaller sample sizes in an experiment as opposed to relying on all observed data of many more scientists.

3.1. Setting and Research Design

We carried out our study in the context of a medical symposium for research on advanced imaging, which is used to detect diseases and other health conditions early, to allow health care practitioners to direct patients to the health care services that they need. We collaborated with the administrators of a large U.S. medical school to layer the medical symposium onto the University's Clinical and Translational Science Center pilot grant program, which provides seed funding in the form of pilot grants to support nascent research efforts that are awarded competitively to faculty within the University.[1] The purpose of the grant opportunity was to solicit proposals to improve methods for using advanced imaging technologies to address unmet clinical needs, and offered $50,000 per award for up to 15 pilot grants. A major challenge in the field of advanced imaging is that

[1]The same field experiment was used in an earlier paper to investigate the effect of search costs on collaborations in the immediate aftermath of the experiment. It did not include long-run knowledge production outcomes or leverage the sociometric badge data, described in Section 3.3.

furthering the knowledge frontier requires expertise in the latest imaging tools and technologies and a deep understanding of the health problems to which they could be applied. These different types of knowledge are typically held by people from distinct disciplinary backgrounds, which makes advanced imaging an ideal setting for our research.

In November 2011, we invited all life sciences faculty and researchers at the university to a medical research symposium for the unique grant funding opportunity. Our field experiment involved faculty and researchers at the University and its affiliated hospitals and institutions, which are independently owned and managed, with each appearing as a separate entity in hospital rankings and lists of National Institutes of Health (NIH) grant recipients. We communicated to applicants that eligibility to submit a grant application was conditional on attending the symposium. In the first stage, investigators interested in applying for the grants submitted a statement of interest in which they briefly described a specific medical problem that advanced imaging techniques could potentially address. We collected basic biographical information (e.g., degree, institution, department appointment) at this stage.

3.2. Participants and Randomization of Knowledge-Sharing Partners

The symposia were held on January 31, February 1, and February 2, 2012, at one of the university's innovation labs. Figure 1 summarizes the key details of the randomization: 402 unique participants were randomly assigned to one of three nights of the symposium, one of four breakout rooms, one of two groups, as well as a poster location to stand next to around the perimeter of the room.[2] Participants were provided an electronic device, called a "sociometric badge" to automatically

[2]Compared to the entire Harvard Medical School (HMS) population (N = 22,625), participants were more likely to be PhDs, had more prior publications on average and were more likely to be instructors, assistant or associate professors relative to the overall distribution at HMS. These differences are consistent with the academic research setting of the symposium and the pilot grant funding opportunity. See Table A1 for summary statistics between participants and the HMS population.

record their face-to-face interactions during the symposium (Kim et al. 2012). Each of the three nights featured a 30-minute welcome address, followed by two 45-minute poster sessions in breakout rooms, with a 15-minute social break in between the poster sessions. Within each breakout room, scientists were randomly assigned to a poster location and either the first or second poster group. The scientists presented and exchanged ideas with one another during the poster sessions. The grant administrators prepared the posters to be a standard size and in a standard format, based on each participant's statements of interest. Thus, treatment pairs assigned to the same breakout rooms had more opportunities for knowledge sharing than control pairs.

Shortly after the symposia, all participants received an e-mail invitation to submit applications for the pilot grants or concept awards by the deadline of March 8, 2012. At this time, they also received PDF booklets with the contact information and posters of all participants so that all researchers had identical information apart from the knowledge acquired in the breakout rooms.

----INSERT FIGURE 1 HERE----

3.3. Data Collection and Variables

We tested our hypotheses using data from a variety of resources from the advanced imaging symposium and the six years of publication records from 2013 to 2018 on the attendees. In the analyses, we did not include the year 2012 to remove potential research topics or ideas that were in progress prior to the symposium. We used data from the scientists' registration form for the symposium, which contained information about their institution, department, academic position, self-identity as an imager or clinician, and statements of interest. The sociometric badges automatically recorded their face-to-face encounters when two badges were facing each other with a direct line of sight within a 30-degree cone of 1m. We verified that all recorded interactions were within 10m using Bluetooth proximity data (Kim et al., 2012) and required that two badges be in contact with each other over a span of at least one minute (Ingram & Morris, 2007). We collected face-to-face interaction data for 306 (74%) scientists who attended the symposium, and the subsequent analyses are based on these scientists.[3] After the symposium, we collected information on the coapplicants and the awardees of the advanced imaging grants and used the Scopus database to collect scientists' publication records.

3.3.1. Dependent Variables

Knowledge transfer. We base our knowledge transfer measure on the MeSH lexicon. In the life sciences, the U.S. National Library of Medicine (NLM) uses a controlled MeSH taxonomy of keywords to index biomedical and health-related information for articles appearing in MEDLINE/PubMed, the NLM Catalog, and other NLM databases. Each article is associated with a set of MeSH keywords that describe the content of the citation. MeSH keywords are assigned by professional science librarians and computer algorithms to ensure global and consistent assignment of keywords across the life sciences (Coletti & Bleich, 2001). We extracted the unique MeSH keywords associated with each scientist's publications to create two vectors of MeSH keywords: the first with all MeSH keywords prior to the advanced imaging symposium (i.e., pre-2012), and the second with all MeSH keywords after the symposium (i.e., 2013–2018). For each scientist-pair $\{i,j\}$, we then counted the number of MeSH terms that j transferred to i and the number of MeSH terms that i transferred to j by taking the intersection of MeSH keywords between i's pre-2012 vector and j's post-2012 vector and between j's pre-2012 vector and i's post-2012 vector,

[3]Badges were randomly assigned to all symposium attendees. 26% of badges malfunctioned and did not record face-to-face interaction data; participants were not aware of whether they were assigned a working or faulty badge. For each symposium night, the sociometric badges recorded interactions for 64.4%, 68.2%, and 90.4% of the participants (Table A2). We conducted balance checks on all available scientist covariates between the complete and observed sample and found no differences between the samples (Table A3). As robustness, we also perform our main OLS regression analyses for each knowledge production outcome on the full sample of participants to verify that the results are directionally consistent across the full sample and the observed badge sample.

excluding any MeSH terms that were common to both *i* and *j* in their pre-2012 MeSH vectors. Our resulting knowledge transfer measure, *MeSH keyword transfer* is the count of “transferred” MeSH keywords between pair {*i*,*j*}, normalized by the total number of MeSH keywords in scientists *i* and *j*’s post-2012 MeSH vectors, expressed as a percentage. The resulting measure ranges from 0% to 100% to and is interpreted as the percentage of post-symposium MeSH keywords that were transferred between scientist-pair {*i*,*j*}, with 0% representing no transferred MeSH keywords and 100% representing complete transfer.

Knowledge creation. We measure knowledge creation as the post-symposium (2013-2018) count of *Copublications* between scientist-pair {*i*,*j*} in peer-reviewed journals, conference proceedings, and book chapters.

Knowledge diffusion. We measure knowledge diffusion as the post-symposium (2013-2018) count of non-coauthored *Forward citations* between scientist-pair {*i*,*j*} in peer-reviewed journals, conference proceedings, and book chapters.

3.3.2. Independent Variables

Knowledge sharing. We use two alternative variables to capture knowledge sharing. First, we use the dummy variable, *Same room,* to measure whether scientist-pair {*i*, *j*} was randomly assigned to the same room (treatment pairs) or different breakout rooms (control pairs). Second, we use the dummy variable, *F2F (face-to-face) communication* to measure whether scientist-pair {*i*, *j*} engaged in least one minute of interaction, recorded using sociometric badges.

Knowledge similarity. Knowledge similarity is comprised of field and intellectual similarity. We measure *Field similarity* using clinical areas (third-party coded from the scientists’ statements of interest), which pertain to the primary area of responsibility for “bedside” patient care. There were a total of 24 unique clinical areas (e.g., oncology, neurology, immunology), and some statements of interests (4.24%) spanned two clinical areas, such as neurology/endocrinology. We then use a categorical variable to indicate whether scientist-pair $\{i, j\}$ shared *Low*, *Moderate* or *High field similarity* depending on whether their clinical areas shared no (85.5%), partial (1.9%) or complete (12.6%) overlap, respectively.

We measure *Intellectual similarity* using the count of common MeSH keywords shared by scientist-pair $\{i, j\}$ prior to 2012 and dichotomize the distribution into three equal-sized groups, to indicate whether scientist-pair $\{i, j\}$ had low, moderate, or a high number of common MeSH keywords prior to the 2012 symposium. *Low intellectual similarity* corresponds to 0–2 common keywords, *Moderate intellectual similarity* corresponds to 3–11 common keywords, and *High intellectual similarity* corresponds to more than 11 common keywords.

Elemental collaboration. We use *Grant coapplicant* to measure an elemental collaboration, which captured whether scientist-pair $\{i, j\}$ coapplied on the grant following the symposium.

3.3.3. Other Variables

The analysis strategy relies most critically on the research design's randomization. We use dummy variables for each symposium night and room (i.e., fixed effects) to control for unobserved night and room characteristics. To test H2c and H2d, we use the dummy variable, *Grant awardee* to capture whether scientist-pair $\{i,j\}$ included a grant recipient. Among the 306 scientists, 13 pilot grant proposals were awarded funding, comprising 6.54 percent of pairs with grant awardees.

3.4. Estimation Approach

We wish to estimate the effects of field and intellectual similarity between knowledge sharing partners on knowledge production outcomes. The unit of analysis is the scientist-pair$\{i,j\}$, and pairs are considered to be "at risk" if they attended the same night of the symposium—a total of 15,817 pairs. First, we analyze the effect of being in the same (treatment) versus different (control) breakout rooms on knowledge production outcomes using ordinary least squares (OLS) regression. We then interact *Same room* with *Field similarity*, and *Same room* with *Intellectual similarity* to examine knowledge similarity effects. Second, we analyze the effect of face-to-face communication on knowledge production outcomes using instrumental variable (IV) regression. We use an IV approach to account for the common endogeneity issue in pairwise interaction data. We exploit exogenous variation in the likelihood of interaction between scientists who are assigned to the same versus different breakout rooms by using *Same room* as an instrument for *F2F communication* in the first stage, and the estimates for *F2F communication* in the second stage. We then interact *F2F communication* with *Field similarity*, and *F2F communication* with *Intellectual similarity* to examine knowledge similarity effects.

We address the non-independence, common-person problem of dyadic regressions by estimating robust standard errors that are simultaneously clustered on both members of the dyad, using multi-way clustering, developed theoretically by Cameron, Gelbach and Miller (2011) and implemented for Stata in clus_nway.ado (Kleinbaum et al., 2013).

4. RESULTS

In this section, we begin by presenting descriptives of the main variables. Table 2 presents the means, standard deviations, and correlations of the main variables. We also note that the summary statistics of the covariates for same room vs. different room pairs indicate that the randomization achieved balance across covariates (see Table A4).

----INSERT TABLE 2 HERE----

Figure 2 shows the yearly trend in copublication (left) and forward citation (right) rates between different room (N = 11,611) and same room (N = 4,206) pairs by year, in the six years to and since the symposium, with 95% CIs. The plots show that there were no differences between the same room and different room pairs before the knowledge sharing intervention. After the intervention, we observe an increase in copublication and citation rates for the same room pairs.

----INSERT FIGURE 2 HERE----

Next, we present our main regression results in subsections, beginning with knowledge transfer, then creation, and lastly, diffusion. We note that the F-statistics for the IV regressions are all above the threshold of 10 for strong instruments (Table A5), and that our results are robust to the inclusion of all scientist-pair covariates (Table A6) and the reduced form (OLS) models for the full sample of participants (Tables A7-A9).

4.1. Knowledge Transfer Results

We present the OLS and IV regression knowledge transfer results in Table 3. The dependent variable is the percentage of scientist-pair {*i*,*j*}'s MeSH post-symposium keywords that were transferred between *i* and *j*, with 0% being no transfer, and 100% being complete transfer. Models 1–2 present the baseline models with main effects only; Models 3–4 add the field similarity interaction terms; Models 5–6 add the intellectual similarity interaction terms; and Models 7–8 show the full results with both interaction terms.

----INSERT TABLE 3 HERE----

H1a and H1b suggest that field and intellectual similarity have an inverted U-shape effect on the relationship between knowledge sharing and knowledge transfer, respectively. Models 3 and 4 indicate that there is a positive but not statistically significant effect of moderate field similarity on knowledge transfer among same room pairs (0.712, $p = 0.132$) and communicating pairs (4.088, $p = 0.197$). Models 5 and 6 show that compared to pairs with high intellectual overlap, scientists with moderate intellectual similarity transferred a higher percentage of MeSH keywords for both same room pairs (0.362, p = 0.042) and communicating pairs (3.348, p = 0.032), respectively. These results are consistent in Models 7 and 8, which include both interaction terms.

Figure 3(a) shows the change in knowledge transfer among communicating vs. non-communicating pairs and intellectual similarity from Model 4, with 95% CIs, and illustrates the inverted U-shaped relationship. While communication did not meaningfully impact knowledge transfer among pairs with low field overlap, it led to an increase of 3.581% among pairs with moderate intellectual overlap from 4.085% [3.795%, 4.376%] to 7.666% [4.995%, 10.338%] percent, a nearly two-fold increase, compared to a 0.233% [-1.239%, 1.705%] increase for high intellectual overlap pairs. The percentage increase among moderately overlapping pairs corresponds to about 6 new MeSH keywords per scientist, roughly equivalent to the average number of MeSH keywords in one publication.[4]

----INSERT FIGURE 3 HERE----

Across the models, we find significant cooperative effects of knowledge transfer for pairs with moderate intellectual similarity. Thus, we confirm H1b but not H1a.

4.2. Knowledge Creation

We present the OLS and IV knowledge creation regression results in Table 4. The dependent variable is the number of copublications between scientist-pair $\{i,j\}$. Models 1–2 present the baseline models with main effects only; Models 3–4 add the field similarity interaction terms; Models 5–6 add the intellectual similarity interaction terms; Models 7–8 show the full results with both interaction terms; Model 9–11 adds *Grant coapplicant*, followed by the field and intellectual similarity interaction terms, as well as both interaction terms, and controls for *Grant awardee*.

----INSERT TABLE 4 HERE----

[4]In the post-symposium period, each scientist has M = 172.57 (S.D. = 142.61) MeSH keywords and M = 6.85 (S.D. = 4.28) MeSH keywords per publication.

H2a and H2b predict that field and intellectual similarity have an inverted U-shape effect on the relationship between knowledge sharing and knowledge creation, respectively, while H2c and H2d predict that knowledge sharing partners who start an elemental collaboration are more likely to copublish when they share greater field and intellectual overlap, respectively. Models 3 and 4 show that compared to same field pairs, there is a positive and statistically significant effect of moderate field similarity on knowledge creation among same room pairs (0.329, $p = 0.047$) and communicating pairs (2.159, $p = 0.036$). There is also a positive effect of low field similarity on knowledge creation among same room pairs (0.133, $p = 0.092$) and communicating pairs (1.040, $p = 0.101$) but the relationship is weak and not significant. Models 5 and 6 show there is a positive but not statistically significant effect of moderate intellectual similarity on knowledge creation among same room pairs (0.0409, $p = 0.257$) and communicating pairs (0.333, $p = 0.234$). The results are consistent in Models 7 and 8, which includes both interaction terms.

Figure 3(b) plots the change in copublications between communicating and non-communicating pairs and field similarity, with 95% CIs from Model 3, and illustrates the estimated inverted U-shape relationship. We observe that while communication did not meaningfully benefit pairs with low field overlap, communication led to an increase of 1.174 [-0.207, 2.556] copublications among pairs with moderate field overlap, and -0.984 [-1.731, -0.237] fewer copublications among pairs with high field overlap. The patterns suggest that communication reduced the search costs of identifying synergistic collaborations with scientists sharing some field interests, resulting in a reallocation of resources away from pairs with high field overlap.

We also examined the extent that the knowledge sharing intervention impacted the scientists' overall research portfolios. Turning to the scientists' post-symposium publications, each scientist had an average of M = 32.394 (S.D. = 35.412) peer-reviewed research articles; the increase of 1.174 publications among pairs with moderate field overlap corresponds to about 7% of a scientist's publications, which is of considerable magnitude from a 90 minute intervention. Turning to changes in research direction, we also note that same room pairs were more likely to publish in advanced imaging journals (e.g., *Magnetic Resonance in Medicine*, *NMR in Biomedicine*): 21% for same room vs. 6% for different room pairs ($t = 2.302$, $p = 0.0249$).[5] This suggests that the knowledge sharing treatment not only reduced search costs of finding collaborators but also potentially reshaped the pairs' research trajectories.

Turning to H2c, Model 9 shows that controlling for *Grant awardee*, grant coapplicants are more likely to copublish when they are from the same field (low: -2.427, $p = 0.189$; moderate: -3.303, $p = 0.042$). Lastly, turning to H2d, Model 10 indicates a negative but not significant effect of decreasing intellectual similarity on the tendency for grant coapplications to persist into copublications (low: -0.639, $p = 0.759$; moderate: -1.083, $p = 0.465$). The results remain consistent in Model 11, which includes both field and intellectual similarity interaction terms. This suggests that the catalyzing force that initiates joint work to create a grant application may differ from the more distal factors that result in copublications. In the near-term, the structural organization of universities into fields and disciplines may incentivize elemental collaborations prioritizing field similarity, but over the long run, synergistic collaborations can emerge among pairs with less overlapping knowledge. Thus, our results show support for H2a and H2c, but not H2b or H2d.

4.3. Knowledge Diffusion

We present the OLS and IV knowledge diffusion regression results in Table 5. The dependent variable is the number of forward citations between scientist-pair $\{i,j\}$. Models 1–2 present the

[5] A journal was coded as an advanced imaging if the journal title included any of the three technologies featured at the symposium (physiological magnetic resonance, positron emission tomography, and optical imaging). There were 48 unique journals for same room vs. 169 unique journals for the different room pairs.

baseline models with main effects only; Models 3–4 add the field similarity interaction terms; Models 5–6 add the intellectual similarity interaction terms; Models 7–8 show the full results with both interaction terms.

----INSERT TABLE 5 HERE----

H3a and H3b predict that knowledge sharing partners are less likely to diffuse knowledge as their field and intellectual similarity increase, respectively. Models 3 and 4 show that increasing field similarity significantly reduced knowledge diffusion among same room pairs (low: 0.366, $p = 0.024$; moderate: 1.003, $p = 0.098$) and communicating pairs (low: 2.823, $p = 0.033$; moderate: 6.603, $p = 0.050$). Models 5 and 6 show that there is no evidence that high intellectual similarity reduced knowledge diffusion among same room (low: -0.0487, $p = 0.739$; moderate: -0.003, $p = 0.986$) or communicating pairs (low: -0.426, $p = 0.712$; moderate: -0.091, $p = 0.936$). The results are consistent in Models 7 and 8, which include both knowledge similarity interaction terms.

Figure 3(c) plots the change in knowledge diffusion (forward citations) between communicating and non-communicating pairs from Model 4, with 95% CIs. We observe that while communication did not lead to a significant change in forward citations for pairs with low field overlap, and led to a marginal increase of 3.521 [-1.465,8.508] citations for pairs with moderate field overlap, we observe a significant decrease from 0.874 [0.347, 1.402] to -2.207 [-4.326, -0.088] for same field pairs, corresponding to a reduction of about 3.082 citations per pair.

To generate greater insight into the types of publications being cited among the low and moderate field overlap pairs, we compared the publication titles of cited papers for pairs assigned to the same vs. different rooms. We observe that the same room pairs were more likely to cite papers using advanced imaging technologies (e.g., sample paper titles included: “Massively parallel MRI detector arrays”; “An fMRI study of facial emotion processing patients with schizophrenia"). Overall, there was a greater use of advanced imaging technologies in the publication titles of same room pairs: 42.6% for same room pairs vs. 31.3% for different room pairs (t = 2.406, p = 0.017).[6] This suggests that the knowledge sharing intervention enabled pairs from different fields to learn about "outside" work, which they subsequently integrated into their own future research. In contrast, the knowledge sharing intervention may have led to a crowding out effect among more similar pairs. In summary, we find significant competitve effects of knowledge diffusion on field similarity, confirming H3a but not H3b.

5. DISCUSSION

The ability to produce and manage knowledge is a critical source of competitive advantage for organizations of all types (Argote, 1999, Hargadon & Sutton, 1997). Yet the knowledge production process is both time-intensive and uncertain (Maggitti, Smith, & Katila, 2013). To address this issue, the premise of this paper is based on the notion that "engineering serendipity" can promote greater knowledge sharing and more efficient knowledge production. We systematically examine how the degree of field and intellectual similarity between knowledge sharing partners affects the likelihood that knowledge sharing affects three knowledge production outcomes: knowledge transfer, creation, and diffusion.

We find both cooperative and competitive effects of serendipitous knowledge sharing on knowledge production. On one hand, knowledge sharing leads to greater knowledge transfer and creation when people share some overlapping research interests: knowledge sharing partners are more likely to transfer and acquire MeSH keywords from one another when they already have some intellectual overlap and more likely to copublish when they share field overlap, with field

[6] There were a total of 149 (same room) and 415 (diff. room) forward citations for pairs with low/moderate field overlap, respectively. A publication title was coded as advanced imaging if it included any of the three technologies featured at the symposium (physiological magnetic resonance, positron emission tomography, and optical imaging).

similarity being more critical to collaborations in the near-term than over longer horizons. On the other hand, knowledge sharing appears to have little benefit for pairs with low knowledge similarity and reduces knowledge diffusion between people from the same field: we observe that pairs from the same field are less likely to cite each other after interacting.

These findings make a number of contributions to the literature. First, we contribute to the literature on search costs and information frictions associated with knowledge production (Agrawal & Goldfarb, 2008; Boudreau et al., 2017; Catalini, 2018; Hansen, 1999; Szulanski, 1996), by focusing on the extent that “engineered serendipity” can mitigate these some of these barriers. Our work suggests that knowledge dissimilarity remains a barrier to knowledge production even in the long-run, as pairs are more likely to be productive when they share sufficient common knowledge. Consistent with this observation, knowledge overlap is a greater constraint on near-term than long-term knowledge creation. Knowledge similarity can be a catalyzing force for near-term research activities because there are fewer start-up costs and uncertainties associated with initiating collaborations with similar others, such as people working in the same field. In [blinded earlier paper], we found that the knowledge sharing intervention’s effect on grant coapplications was most beneficial for pairs from the same field. We expand on these findings to show that while field overlap remains a difficult hurdle to overcome in the near-term, over time, scientists place smaller premiums on knowledge similarity, instead emphasizing synergistic collaborations where partners have both common knowledge and divergent expertise. This shift may be due to the time and repeated interactions needed to develop an initial encounter into a multiplex relationship (Ingram and Morris, 2007; Uzzi, 1997), or the evolving expectations from collaborators as projects move from exploratory research to joint execution of research ideas. They also suggest that homophilous collaborations (Dahlander & McFarland, 2013) can be attributed in part to search costs and information costs. By differentiating between field and intellectual similarity, we show that the formal organizational structure associated with field specialties—due to forces of departmental or disciplinary attraction (Blau, 1973; Rawlings et al., 2015) tend to exert a stronger influence on shaping *visible* collaborative behaviors than those of intellectual specialties, which are less likely to benefit from formal organizational incentives.

Second, this study advances understanding of how opportunities for temporary colocation, offered by conferences, symposia and similar events potentially impact the direction and quality of scientific knowledge production. Although prior research suggests that temporary colocation is critical to knowledge production (Catalini, 2018; Catalini et al., 2020; Campos et al., 2018; Chai & Freeman, 2019), it has typically inferred face-to-face interaction from observational data, rather than actual acts of knowledge sharing. We suggest that one attractive aspect of conferences over informal serendipitous encounters (e.g., in hallways, watercoolers) is that they have a focusing effect on the topic of knowledge sharing and idea exchange. This may be why even brief, highly structured, information-rich interventions can have long-term consequences on knowledge production, most notably by promoting research activities on the conference's topics of inquiry. It is important to note that the effects of these events go beyond monetary incentives, which may be an early catalyst to attract participants but the effects are more widespread and longer lasting.

Third, we make empirical and methodological contributions by designing a prospective study that amalgamates multiple sources and uses of data, namely a natural field experiment, direct observations of face-to-face communications and archival publication data to study long-run knowledge production outcomes. This careful combination of data sources enables us to investigate the causal relationships between knowledge sharing and knowledge production over different time horizons, which are not possible with archival data (Fleming et al., 2007). Given the rise in field experiments in innovation research (Chatterji, Delecourt, Hasan & Koning; Hasan & Koning, 2019), research designs that at the outset plan for short- and long-term time horizons may yield deeper insights that can help to offset the costs associated with multiple forms of data collection. In addition, our research design can be used as a template for future experiments with multi-term horizons, in terms of learning from its features and its potential shortcomings.

In this research, we have made a thorough effort to analyze how knowledge sharing leads to knowledge production in science. However, our study has some empirical limitations. First, our study focused on an interdisciplinary setting in a large, highly selective university. Therefore, it may be interpreted as a best-case scenario in fostering knowledge production, as we must draw a boundary around the network and incentives under consideration. The symposium was highly structured to facilitate knowledge sharing, while knowledge production is a core activity for academics. Second, the geographic proximity of scientists may have facilitated scheduled or serendipitous encounters with greater ease (Catalini, 2018). Third, our focus on the life sciences also draws upon a specific population, where intragroup competition is normative (Haas & Park, 2010), in part due to the "winner-takes-all" model of rewards and recognition (Stephan, 1996). We recognize that other settings may feature greater geographic dispersion or promote more openness and cooperation; a promising future direction would be to extend this research to different settings.

We also focused most heavily on archival data (e.g., public trace data from publications) to examine long-term knowledge production outcomes, which provided greater insights into the scientists' behaviors rather than the reasons behind the observed patterns. Although we made efforts to qualify these findings by examining the content of the knowledge transfer, publications, and citations, future research can seek to supplement archival data with alternative methods (e.g., surveys, interviews) that continue to track the scientists' interaction patterns, as well as the invisible track of "failed" knowledge production. Such complementary data sources will ultimately provide deeper insights on how engineered serendipitous encounters between people can be cultivated into productive relationships over time. This future work would facilitate a deeper performative assessment of how knowledge sharing can be "serendipitously engineered" to shape the quality of knowledge outcomes.

Overall, this study takes a critical step towards identifying the processes that explain when serendipitous encounters shape knowledge production outcomes among innovating individuals. We show that brief, information-rich interactions between people with some overlapping knowledge interests can have a productive effect on knowledge transfer, creation and diffusion.

6. REFERENCES

Abbott, A. (2001). *Time matters: On theory and method.* Chicago, IL: University of Chicago Press.

Adams, J. D., Black, G. C., Clemmons, J. R., & Stephan, P. E. (2005). Scientific teams and institutional collaborations: Evidence from US universities, 1981–1999. *Research Policy*, *34*(3), 259-285.

Agrawal, A., & Goldfarb, A. (2008). Restructuring research: Communication costs and the democratization of university innovation. *American Economic Review*, *98*(4), 1578-90.

Allen, T. J. (1977). *Managing the flow of technology: Technology transfer and the dissemination of technological information within the R&D organization.* Cambridge, MA: MIT Press.

Argote, L. (1999). *Organizational learning: Creating, retaining and transferring.* Boston, MA: Kluwer.

Arora, A., & Gambardella, A. (2005). The globalization of the software industry: perspectives and opportunities for developed and developing countries. *Innovation Policy and the Economy*, *5*, 1-32.

Arrow, K. (1962). *Economic welfare and the location of resources for invention*. *The rate and direction of inventive activity.* Princeton, NJ: Princeton University Press.

Ayoubi, C., Pezzoni, M. and Visentin, F., 2019. The important thing is not to win, it is to take part: What if scientists benefit from participating in research grant competitions?. *Research Policy*, 48(1), pp.84-97.

Azoulay, P., Ding, W., & Stuart, T. (2009). The impact of academic patenting on the rate, quality and direction of (public) research output. *The Journal of Industrial Economics*, *57*(4), 637-676.

Bechky, B. (2011). Making Organizational Theory Work: Institutions, Occupations, and Negotiated Orders. *Organization Science 22*(5), 1157–1167.

Biancani, S., Dahlander, L., McFarland, D. A., & Smith, S. (2018). Superstars in the making? The broad effects of interdisciplinary centers. *Research Policy*, *47*(3), 543-557.

Biancani, S., McFarland, D. A., & Dahlander, L. (2014). The semiformal organization. *Organization Science*, *25*(5), 1306-1324.

Blau, P. M. (1973)). *The organization of academic work*. New York: Wiley.

Blumenthal, D., Campbell, E. G., Anderson, M. S., Causino, N., & Louis, K. S. (1997). Withholding research results in academic life science: evidence from a national survey of faculty. *Jama*, *277*(15), 1224-1228.

Bourdieu, P. (1984). *Distinction: A social critique of the judgment of taste*. Cambridge, MA: Harvard University Press.

Boudreau, K. J., Brady, T., Ganguli, I., Gaule, P., Guinan, E., Hollenberg, A., & Lakhani, K. R. (2017). A field experiment on search costs and the formation of scientific collaborations. *Review of Economics and Statistics*, *99*(4), 565-576.

Boudreau, K. J., Guinan, E. C., Lakhani, K. R., & Riedl, C. (2016). Looking across and looking beyond the knowledge frontier: Intellectual distance, novelty, and resource allocation in science. *Management Science, 62*(10), 2765-2783.

Boudreau, K. J., & Lakhani, K. R. (2015). "Open" disclosure of innovations, incentives and follow-on reuse: Theory on processes of cumulative innovation and a field experiment in computational biology. *Research Policy*, *44*(1), 4-19.

Bower, G. H., Hilgard, E. R. (1981). *Theories of learning*. Englewood Cliffs, NJ: Prentice-Hall.

Campanario, J. M., & Acedo, E. (2007). Rejecting highly cited papers: The views of scientists who encounter resistance to their discoveries from other scientists. *Journal of the American Society for Information Science and Technology, 58*(5), 734-743.

Campbell, E. G., Clarridge, B. R., Gokhale, M., Birenbaum, L., Hilgartner, S., Holtzman, N. A., & Blumenthal, D. (2002). Data withholding in academic genetics: Evidence from a national survey. *Jama*, *287*(4), 473-480.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, *29*(2), 238-249.

Campos, R., Leon, F., & McQuillin, B. (2018). Lost in the storm: the academic collaborations that went missing in Hurricane ISSAC. *The Economic Journal*, *128*(610), 995-1018.

Carlile, P. R. (2004). Transferring, translating, and transforming: An integrative framework for managing knowledge across boundaries. *Organization Science*, *15*(5), 555-568.

Catalini, C. (2018). Microgeography and the direction of inventive activity. *Management Science*, *64*(9), 4348-4364.

Catalini, C., Fons-Rosen, C., & Gaulé, P. (2020). How Do Travel Costs Shape Collaboration?. *Management Science*.

Caza, B. B., Moss, S., & Vough, H. (2018). From synchronizing to harmonizing: The process of authenticating multiple work identities. *Administrative Science Quarterly, 63*(4), 703-745.

Chai, S., & Freeman, R. B. (2019). Temporary colocation and collaborative discovery: Who confers at conferences. *Strategic Management Journal*, *40*(13), 2138-2164.

Chatterji, A., Delecourt, S., Hasan, S., & Koning, R. (2019). When does advice impact startup performance?. *Strategic Management Journal, 40*(3), 331-356.

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, 128-152.

Coletti, M. H., & Bleich, H. L. (2001). Medical subject headings used to search the biomedical literature. *Journal of the American Medical Informatics Association, 8*(4), 317-323.

Dahlander, L., & McFarland, D. A. (2013). Ties that last: Tie formation and persistence in research collaborations over time. *Administrative Science Quarterly*, *58*(1), 69-110.

Dahlander, L., O'Mahony, S., & Gann, D. M. (2016). One foot in, one foot out: how does individuals' external search breadth affect innovation outcomes?. *Strategic Management Journal*, *37*(2), 280-302.

Fang, C., Lee, J., & Schilling, M. A. (2010). Balancing exploration and exploitation through structural design: The isolation of subgroups and organizational learning. *Organization Science*, *21*(3), 625-642.

Fleming, L. (2001). Recombinant uncertainty in technological search. *Management science*, *47*(1), 117-132.

Fleming, L., Mingo, S., & Chen, D. (2007). Collaborative brokerage, generative creativity, and creative success. *Administrative Science Quarterly*, *52*(3), 443-475.

Ganguli, I., 2017. Saving Soviet science: The impact of grants when government R&D funding disappears. *American Economic Journal: Applied Economics*, 9(2), pp.165-201.

Haas, M. R., & Park, S. (2010). To share or not to share? Professional norms, reference groups, and information withholding among life scientists. *Organization Science*, *21*(4), 873-891.

Haeussler, C., Jiang, L., Thursby, J., & Thursby, M. (2014). Specific and general knowledge sharing among competing academic researchers. *Research Policy*, *43*(3), 465-475.

Hansen, M. T. (1999). The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits. *Administrative Science Quarterly*, *44*(1), 82-111.

Hagstrom, W. O. (1974). Competition in science. *American Sociological Review*, *39*(1), 1-18.

Hargadon, A., & Sutton, R. I. (1997). Technology brokering and innovation in a product development firm. *Administrative Science Quarterly*, 716-749.

Hasan, S., & Koning, R. (2019). Prior ties and the limits of peer effects on startup team performance. *Strategic Management Journal*, *40*(9), 1394-1416.

Ingram, P., & Morris, M. W. (2007). Do people mix at mixers? Structure, homophily, and the “life of the party”. *Administrative Science Quarterly*, *52*(4), 558-585.

Jeppesen, L. B., & Lakhani, K. R. (2010). Marginality and problem-solving effectiveness in broadcast search. *Organization Science*, *21*(5), 1016-1033.

Jones, B. F. (2009). The burden of knowledge and the “death of the renaissance man”: Is innovation getting harder?. *The Review of Economic Studies*, *76*(1), 283-317.

Kim, T., McFee, E., Olguin, D. O., Waber, B., & Pentland, A. S. (2012). Sociometric badges: Using sensor technology to capture new forms of collaboration. *Journal of Organizational Behavior*, *33*(3), 412-427.

Kleinbaum, A. M., Stuart, T. E., & Tushman, M. L. (2013). Discretion within constraint: Homophily and structure in a formal organization. *Organization Science*, *24*(5), 1316-1336.

Kogut, B., & Zander, U. (1992). Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science*, *3*(3), 383-397.

Lakhani, K., von Hippel, E. (2003). How open source software works: “free” user-to-user assistance. *Research Policy 32*(6), 925–943.

Latour B, Woolgar S (1979) *Laboratory life.* Beverly Hills, CA: Sage.

Laursen, K., Masciarelli, F., & Prencipe, A. (2012). Regions matter: How localized social capital affects innovation and external knowledge acquisition. *Organization Science*, *23*(1), 177-193.

Leahey E, Beckman CM, Stanko TL (2017) Prominent but less productive: The impact of interdisciplinarity on scientists' research. *Administrative Science Quarterly* 62(1):105–139.
Lee, S. (2019). Learning-by-moving: can reconfiguring spatial proximity between organizational members promote individual-level exploration?. *Organization Science*, *30*(3), 467-488.
Levinthal, D. A., & March, J. G. (1993). The myopia of learning. *Strategic Management Journal*, *14*(S2), 95-112.
Maggitti PG, Smith KG, Katila R. (2013) The complex search process of invention. *Research Policy* 42(1):90–100.
Maurer, I., & Ebers, M. (2006). Dynamics of social capital and their performance implications: Lessons from biotechnology start-ups. *Administrative Science Quarterly*, *51*(2), 262-292.
McFadyen MA, Cannella Jr, A. A. (2004). Social capital and knowledge creation: Diminishing returns of the number and strength of exchange relationships. *Academy of Management Journal 47*(5), 735–746.
McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*(1), 415-444.
Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations* Chicago, IL: University of Chicago Press.
Murray, F., & O'Mahony, S. (2007). Exploring the foundations of cumulative innovation: Implications for organization science. *Organization Science*, *18*(6), 1006-1021.
Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization Science*, *5*(1), 14-37.
Pain, E. (2018, February 28). Collaborating for the win. *Science.* Retrieved from https://www.sciencemag.org/careers/2018/02/collaborating-win
Phelps, C., Heidl, R., & Wadhwa, A. (2012). Knowledge, networks, and knowledge networks: A review and research agenda. *Journal of Management*, *38*(4), 1115-1166.
Rawlings, C. M., McFarland, D. A., Dahlander, L., & Wang, D. (2015). Streams of thought: Knowledge flows and intellectual cohesion in a multidisciplinary era. *Social Forces*, *93*(4), 1687-1722.
Reschke, B. P., Azoulay, P., & Stuart, T. E. (2018). Status spillovers: The effect of status-conferring prizes on the allocation of attention. *Administrative Science Quarterly*, *63*(4), 819-847.
Rosenkopf, L., & Almeida, P. (2003). Overcoming local search through alliances and mobility. *Management Science*, *49*(6), 751-766.
Simon, H. A. (1991). Bounded rationality and organizational learning. *Organization Science*, *2*(1), 125-134.
Singh, J. (2005). Collaborative networks as determinants of knowledge diffusion patterns. *Management science*, *51*(5), 756-770.
Singh, J., & Fleming, L. (2010). Lone inventors as sources of breakthroughs: Myth or reality?. *Management Science*, *56*(1), 41-56.
Song, J., Almeida, P., & Wu, G. (2003). Learning–by–hiring: When is mobility more likely to facilitate interfirm knowledge transfer?. *Management Science*, *49*(4), 351-365.
Spender, J. C. (1996). Making knowledge the basis of a dynamic theory of the firm. *Strategic Management Journal*, *17*(S2), 45-62.

Staw, B. M., & Epstein, L. D. (2000). What bandwagons bring: Effects of popular management techniques on corporate performance, reputation, and CEO pay. *Administrative Science Quarterly*, *45*(3), 523-556.

Stephan, P. E. (1996). The economics of science. *Journal of Economic literature*, *34*(3), 1199-1235.

Szulanski, G. (1996). Exploring internal stickiness: Impediments to the transfer of best practice within the firm. *Strategic Management Journal*, *17*(S2), 27-43.

Uzzi, B. (1997). Social structure and competition in interfirm networks: The paradox of embeddedness. *Administrative Science Quarterly*, 42(1), 35-67.

Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical combinations and scientific impact. *Science*, *342*(6157), 468-472.

Zhang, L., Guler, I. (2019). How to Join the Club: Patterns of Embeddedness and the Addition of New Members to Interorganizational Collaborations, *Administrative Science Quarterly*.

TABLE 1 Summary of Knowledge Constructs

Knowledge Construct	Definition	Citations
Knowledge Production	A multi-component process that involves knowledge transfer, creation and diffusion.	Grant, 1996; Kogut & Zander, 1992; Nonaka, 1994; Spender, 1996
Knowledge Transfer	The movement of facts, relationships, and insights from one setting to another	Argote, 1999; Hansen, 1999; Szulanski, 1996
Knowledge Creation	The generation of facts, relationships, and insights to solve problems that are new to the knowledge frontier	Arrow 1962; Boudreau & Lakhani, 2015; Nonaka 1994
Knowledge Diffusion	The dissemination of facts, relationships, and insights that are subsequently used by others	Fleming & Singh, 2010; Rogers, 1983; Singh, 2005
Knowledge Similarity	The degree of common knowledge between two individuals' field of study and intellectual interests	Cohen & Levinthal, 1990; Carlile, 2004; Dahlander & McFarland, 2013; Leahey et al., 2017
Field similarity	The dimension of knowledge similarity that emerges from the overlap between two individual's educational background, skills and training	Bechky, 2011; Carlile, 2004; Bourdieu, 1984; Haas & Park, 2010
Intellectual similarity	The dimension of knowledge similarity that emerges from the overlap between two individual's personal interests and passions	Caza et al., 2018; Dahlander & McFarland, 2013

TABLE 2 Correlation Between Main Variables (N = 15,817)

	Variable	Mean	Std. dev.	1	2	3	4	5	6	7
1	MeSH keyword transfer	3.693	3.709							
2	Copublications	0.031	1.004	0.010						
3	Forward citations	0.265	2.497	0.019	0.479					
4	Grant coapplicant	0.002	0.044	0.004	0.097	0.027				
5	Same room	0.266	0.442	0.024	-0.004	-0.001	0.006			
6	F2F communication	0.078	0.268	0.003	0.047	0.039	0.067	0.198		
7	Intellectual similarity	1.076	0.810	-0.005	0.037	0.043	0.037	-0.002	0.047	
8	Field similarity	0.272	0.673	0.437	0.029	0.074	0.017	-0.008	0.021	0.026

Note: Intellectual and field similarity are categorical variables with the following distributions: intellectual (low: 32.4%, moderate: 32.7%, high: 35.0%) and field (low: 85.5%, 1.9%,12.6%)

TABLE 3 Regression Models of Knowledge Transfer – % of MeSH Keywords Transferred between scientist-pair $\{i,j\}$; N = 15,817

VARIABLES	Model 1 OLS	Model 2 IV	Model 3 OLS	Model 4 IV	Model 5 OLS	Model 6 IV	Model 7 OLS	Model 8 IV
Same room	0.107		0.0344		0.0336		-0.0379	
	(0.0979)		(0.144)		(0.138)		(0.172)	
F2F communication		0.838		0.214		0.233		-0.512
		(0.763)		(1.108)		(1.025)		(1.352)
Same room x Low field similarity			0.0707				0.0686	
			(0.150)				(0.150)	
Same room x Moderate field similarity			0.712				0.681	
			(0.471)				(0.468)	
Same room x Low intellectual similarity					-0.175		-0.171	
					(0.218)		(0.218)	
Same room x Moderate intellectual similarity					0.362		0.361	
					(0.177)		(0.177)	
F2F x Low field similarity				0.621				0.804
				(1.204)				(1.239)
F2F x Moderate field similarity				4.088				4.059
				(3.169)				(3.379)
F2F x Low intellectual similarity						-1.459		-1.531
						(1.753)		(1.869)
F2F x Moderate intellectual similarity						3.348		3.305
						(1.564)		(1.644)
Low field similarity	0.140	0.171	0.121	0.104	0.141	0.176	0.123	0.0900
	(0.135)	(0.135)	(0.143)	(0.183)	(0.135)	(0.137)	(0.143)	(0.189)
Moderate field similarity	-0.130	-0.136	-0.316	-0.606	-0.136	-0.174	-0.315	-0.638
	(0.274)	(0.264)	(0.269)	(0.381)	(0.272)	(0.274)	(0.268)	(0.398)
Low intellectual similarity	-4.130	-4.120	-4.130	-4.123	-4.081	-4.015	-4.082	-4.013
	(0.238)	(0.237)	(0.238)	(0.238)	(0.246)	(0.282)	(0.246)	(0.289)
Moderate intellectual similarity	-0.880	-0.862	-0.880	-0.866	-0.974	-1.095	-0.974	-1.096
	(0.153)	(0.152)	(0.153)	(0.152)	(0.163)	(0.196)	(0.163)	(0.204)
Night FE	Y	Y	Y	Y	Y	Y	Y	Y
Room FE	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.224	0.219	0.224	0.218	0.224	0.201	0.224	0.216

Multiway, robust standard errors in parentheses. Significance stars (*) are omitted.

TABLE 4 Regression Models of Knowledge Creation – # of copublications between scientist-pair {*i*,*j*}; N = 15,817

VARIABLES	Model 1 OLS	Model 2 IV	Model 3 OLS	Model 4 IV	Model 5 OLS	Model 6 IV	Model 7 OLS	Model 8 IV	Model 9 OLS	Model 10 OLS	Model 11 OLS
Same room	-0.00683		-0.127		-0.0294		-0.148		-0.00784	-0.00822	-0.00799
	(0.0144)		(0.0769)		(0.0360)		(0.0931)		(0.0148)	(0.0148)	(0.0147)
F2F communication		-0.0535		-0.984		-0.221		-1.147			
		(0.111)		(0.616)		(0.268)		(0.746)			
Same room x Low field sim.			0.133				0.132				
			(0.0788)				(0.0783)				
Same room x Mod. field sim.			0.329				0.330				
			(0.165)				(0.164)				
Same room x Low intellectual sim.					0.0306		0.0307				
					(0.0378)		(0.0372)				
Same room x Mod. intellectual sim.					0.0409		0.0397				
					(0.0361)		(0.0355)				
F2F x Low field sim.				1.040				1.048			
				(0.634)				(0.637)			
F2F x Mod. field sim.				2.159				2.110			
				(1.029)				(1.022)			
F2F x Low intellectual sim.						0.240		0.205			
						(0.297)		(0.306)			
F2F x Mod. intellectual sim.						0.333		0.332			
						(0.280)		(0.290)			
Grant coapplicant									3.708	2.482	3.810
									(1.583)	(1.211)	(1.614)
Coapplicant x Low field sim.									-2.427		-2.299
									(1.845)		(2.109)
Coapplicant x Mod. field sim.									-3.303		-3.405
									(1.621)		(1.659)
Coapplicant x Low intellectual sim.										-0.639	-0.324
										(2.085)	(2.236)
Coapplicant x Mod. intellectual sim.										-1.083	-0.571
										(1.482)	(1.632)
Awardee									0.000295	-0.00167	-2.47e-05
									(0.0140)	(0.0137)	(0.0140)
Low field sim.	-0.109	-0.111	-0.14	-0.220	-0.109	-0.112	-0.144	-0.222	0.0552	0.0405	0.0552
	(0.0540)	(0.0566)	(0.0735)	(0.120)	(0.0539)	(0.0579)	(0.0733)	(0.121)	(0.0422)	(0.0441)	(0.0422)
Moderate field sim.	-0.0529	-0.0525	-0.139	-0.297	-0.0534	-0.0567	-0.140	-0.296	0.0887	0.0981	0.0887
	(0.0659)	(0.0655)	(0.0701)	(0.132)	(0.0660)	(0.0675)	(0.0706)	(0.130)	(0.0536)	(0.0539)	(0.0536)
Low intellectual sim.	-0.0696	-0.0702	-0.0695	-0.0738	-0.0777	-0.0908	-0.0777	-0.0916	-0.0651	-0.0643	-0.0645
	(0.0222)	(0.0226)	(0.0221)	(0.0240)	(0.0297)	(0.0437)	(0.0294)	(0.0455)	(0.0222)	(0.0224)	(0.0224)
Moderate intellectual sim.	-0.0657	-0.0669	-0.0656	-0.0699	-0.0765	-0.0928	-0.0760	-0.0954	-0.0602	-0.0591	-0.0593
	(0.0209)	(0.0219)	(0.0207)	(0.0231)	(0.0272)	(0.0399)	(0.0268)	(0.0418)	(0.0209)	(0.0209)	(0.0210)
Night FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Room FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.003	0.002	0.004	--	0.003	--	0.004	0.012	0.015	0.012	0.015

Multi-way robust standard errors in parentheses. Significance stars (*) are omitted.

TABLE 5 Regression Models of Knowledge Diffusion – # of forward citations between scientist-pair {*i*,*j*}; N = 15,817

VARIABLES	Model 1 OLS	Model 2 IV	Model 3 OLS	Model 4 IV	Model 5 OLS	Model 6 IV	Model 7 OLS	Model 8 IV
Same room	-0.0687		-0.399		-0.0540		-0.384	
	(0.0710)		(0.165)		(0.150)		(0.211)	
F2F communication		-0.537		-3.082		-0.391		-2.910
		(0.551)		(1.331)		(1.115)		(1.656)
Same room x Low field similarity			0.366				0.367	
			(0.162)				(0.162)	
Same room x Moderate field similarity			1.003				0.999	
			(0.605)				(0.605)	
Same room x Low intellectual similarity					-0.0487		-0.0476	
					(0.146)		(0.146)	
Same room x Low intellectual similarity					-0.00251		-0.00592	
					(0.145)		(0.145)	
F2F x Low field similarity				2.823				2.835
				(1.324)				(1.332)
F2F x Moderate field similarity				6.603				6.679
				(3.364)				(3.334)
F2F x Low intellectual similarity						-0.426		-0.537
						(1.154)		(1.162)
F2F x Moderate intellectual similarity						-0.0908		-0.112
						(1.138)		(1.135)
Low field similarity	-0.316	-0.336	-0.413	-0.633	-0.316	-0.335	-0.413	-0.632
	(0.129)	(0.137)	(0.165)	(0.256)	(0.129)	(0.140)	(0.165)	(0.258)
Moderate field similarity	-0.209	-0.205	-0.471	-0.955	-0.209	-0.204	-0.471	-0.962
	(0.208)	(0.214)	(0.185)	(0.360)	(0.208)	(0.215)	(0.186)	(0.358)
Low intellectual similarity	-0.446	-0.452	-0.446	-0.462	-0.433	-0.417	-0.433	-0.418
	(0.0920)	(0.0926)	(0.0918)	(0.0950)	(0.102)	(0.141)	(0.102)	(0.143)
Moderate intellectual similarity	-0.353	-0.364	-0.352	-0.373	-0.352	-0.354	-0.351	-0.361
	(0.0762)	(0.0791)	(0.0760)	(0.0810)	(0.0866)	(0.128)	(0.0862)	(0.130)
Night FE	Y	Y	Y	Y	Y	Y	Y	Y
Room FE	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.011	0.004	0.011	--	0.011	0.004	0.011	--

Multi-way, robust standard errors in parentheses. Significance stars (*) are omitted.

FIGURE 1 Randomization of Participants By Night, Room, Group and Poster Location

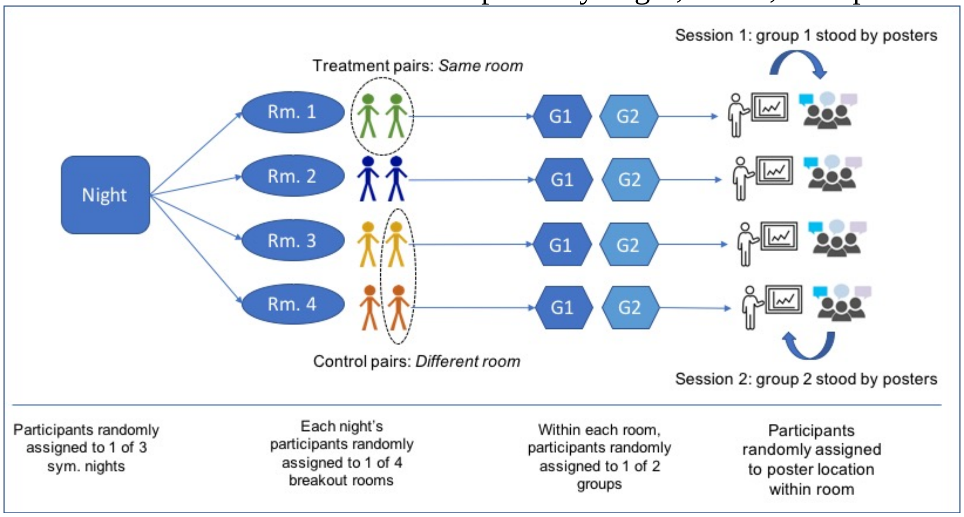


FIGURE 2 Yearly Trend in Copublication (Left) and Forward Citation (Right) Rates Between Same Room (Treatment) and Different Room (Control) Pairs with 95% CIs

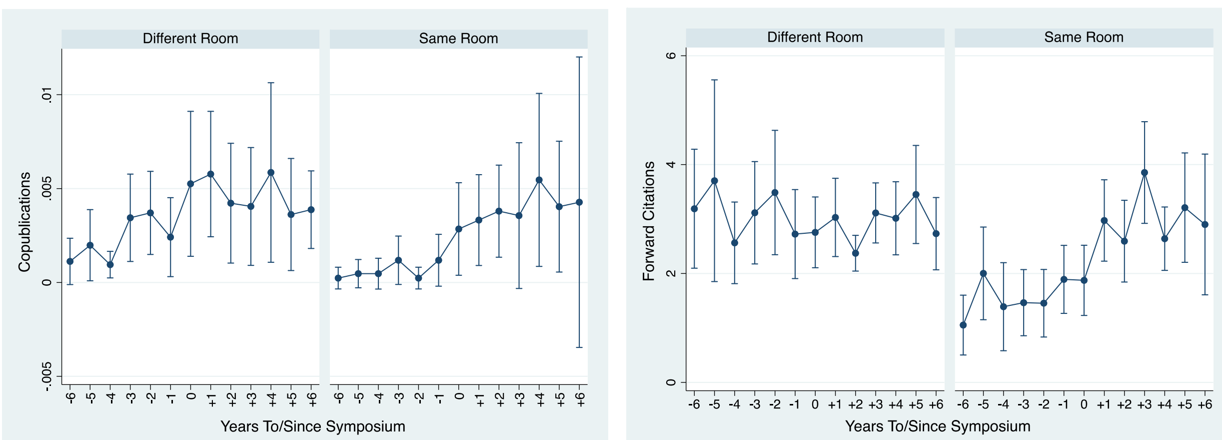
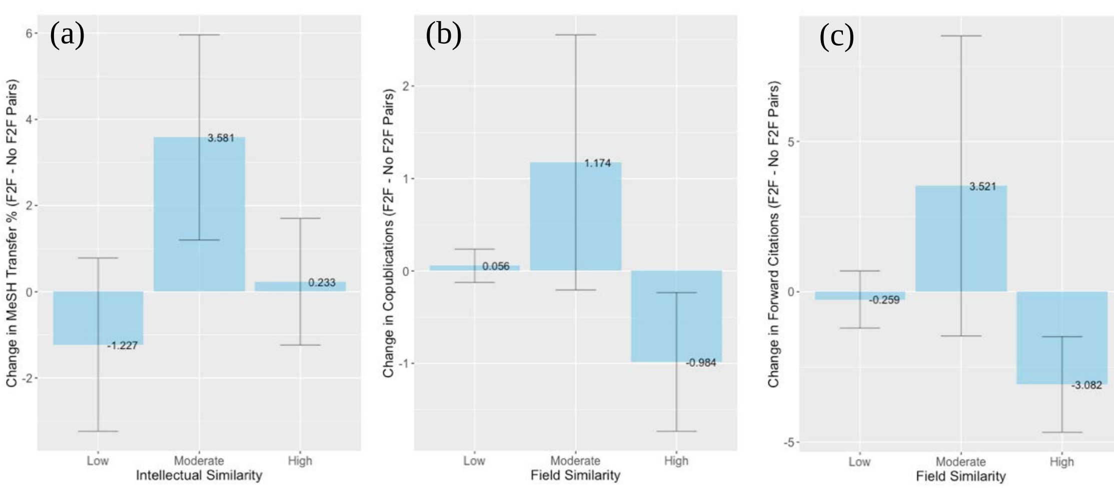


FIGURE 3 Change in (a) Knowledge Transfer, (b) Knowledge Creation, (c) Knowledge Diffusion and Knowledge Similarity for Communicating and Non-communicating Pairs



APPENDIX for

Engineering Serendipity: When Does Knowledge Sharing Lead to Knowledge Production?

TABLE A1 Summary Statistics of HMS Population vs. Symposium Participants

Sample Means	*HMS Profiles*	*Participants*	*Difference*
Degree			
MD	0.604	0.572	0.031
PhD	0.382	0.493	-0.111
Publications	17.837	22.169	-4.332
Rank			
Professor	0.061	0.037	0.024
Associate Professor	0.066	0.157	-0.091
Assistant Professor	0.108	0.204	-0.096
Instructor	0.278	0.331	-0.052
Postdoc/Fellow	0.401	0.219	0.182
Other	0.086	0.052	0.034
Longwood	0.619	0.510	0.109
Hospital			
Beth Israel Deaconess Medical Center	0.128	0.139	-0.011
Massachusetts General	0.245	0.371	-0.125
Brigham and Women's	0.201	0.184	0.017
Children's Hospital	0.120	0.129	-0.009
Radiology Department	0.054	0.266	-0.213
Observations	22,625	402	

TABLE A2 Comparison of Observed and Full Sample of Participants by Night, Room and Group

Symposium Night	Count (%) of Total
1	92 (63.4%)
2	90 (68.2%)
3	124 (90.4%)
Total	306 (73.9%)
Breakout Room	Count (%) of Total
1	57 (60.6%)
2	114 (85.7%)
3	63 (67.0%)
4	72 (77.2%)
Total	306 (73.8%)
Poster Group	Count (%) of Total
1	150 (73.9%)
2	156 (73.8%)
Total	306 (73.8%)

Note: N = 413 observations for the full sample, comprised of 402 participants (392 scientists that attended one night of the symposium and 10 “super experts” with deep expertise in advanced imaging technologies who were invited to attend between one and three symposium nights, for a total of 21 observations).

TABLE A3 Covariate Balance Checks Between Observed and Full Sample of Participants

Covariate	χ^2 test statistic and p value
Female	χ^2 (1) = 0.1159, p = 0.734
Faculty rank	χ^2 (7) = 5.6566, p = 0.580
Institution	χ^2 (23) = 24.076, p = 0.400
Department	χ^2 (55) = 47.579, p = 0.751
Clinical area	χ^2 (38) = 45.223, p = 0.196
Imager	χ^2 (1) = 0.4540, p = 0.500
Imaging technology	χ^2 (19) = 20.167, p = 0.385
Super expert	χ^2 (1) = 5.281, p = 0.022

Note: N = 413 observations for the full sample, comprised of 402 participants (392 scientists attending one night of the symposium and 10 “super experts” with deep expertise in advanced imaging technologies who were invited to attend between one and three symposium nights, for a total of 21 observations).

TABLE A4 Summary Statistics For Same Room versus Different Room Pairs (N = 15,817)

Covariate	Treatment: Same Room	Control: Different Room	Difference
Previous coauthor	0.001	0.002	-0.001
Same institution	0.179	0.189	-0.010
Same department	0.102	0.097	0.005
Same imaging tech.	0.175	0.171	0.003
Both imagers	0.170	0.175	-0.005
Both clinicians	0.343	0.335	0.009
Same rank	0.202	0.218	-0.016
Both female	0.084	0.084	-0.000
Both male	0.506	0.496	0.001

TABLE A5 First Stage OLS Regression Models of F2F Communication on Same Room, Field Similarity and Intellectual Similarity (N = 15,817)

VARIABLES	Model 1	Model 2	Model 3	Model 4
Same room	0.128	0.137	0.136	0.145
	(0.00888)	(0.0211)	(0.0129)	(0.0230)
Same room x Low field sim.		-0.0111		-0.0108
		(0.0225)		(0.0226)
Same room x Mod. field sim.		0.0420		0.0424
		(0.0522)		(0.0525)
Same room x Low intellectual sim.			-0.00673	-0.00619
			(0.0170)	(0.0171)
Same room x Mod. intellectual sim.			-0.0196	-0.0194
			(0.0166)	(0.0166)
Low field sim.	-0.0374	-0.0344	-0.0375	-0.0346
	(0.00923)	(0.00841)	(0.00923)	(0.00841)
Moderate field sim.	0.00704	-0.00392	0.00729	-0.00376
	(0.0293)	(0.0246)	(0.0293)	(0.0247)
Low intellectual sim.	-0.0114	-0.0114	-0.00961	-0.00979
	(0.00706)	(0.00705)	(0.00591)	(0.00589)
Moderate intellectual sim.	-0.0214	-0.0215	-0.0163	-0.0164
	(0.00627)	(0.00627)	(0.00566)	(0.00565)
Night FE	Y	Y	Y	Y
Room FE	Y	Y	Y	Y
First stage F-statistic	361.031	173.583	184.852	103.413
R-squared	0.045	0.045	0.045	0.046

Multi-way, robust standard errors in parentheses.

TABLE A6 Regression Models of Knowledge Outcomes With Pairwise Covariates; N = 15,817

VARIABLES	Model 1 Transfer OLS	Model 2 Transfer IV	Model 3 Creation OLS	Model 4 Creation IV	Model 5 Creation OLS	Model 6 Diffusion OLS	Model 7 Diffusion IV
Same room	-0.0694		-0.0901		0.00193	-0.262	
	(0.166)		(0.0698)		(0.0133)	(0.173)	
F2F communication		-0.808		-0.698			-1.959
		(1.284)		(0.544)			(1.301)
Same room x Low field sim.	0.0961		0.0883			0.276	
	(0.146)		(0.0596)			(0.128)	
Same room x Mod. field sim.	0.670		0.273			0.860	
	(0.482)		(0.149)			(0.569)	
Same room x Low intellectual sim.	-0.149		0.0132			-0.0795	
	(0.211)		(0.0311)			(0.136)	
Same room x Mod. intellectual sim.	0.395		0.0246			-0.0356	
	(0.173)		(0.0316)			(0.137)	
F2F x Low field similarity		1.064		0.700			2.110
		(1.204)		(0.475)			(1.036)
F2F x Mod. field sim.		3.999		1.715			5.697
		(3.463)		(0.893)			(3.046)
F2F x Low intellectual sim.		-1.330		0.0704			-0.778
		(1.795)		(0.247)			(1.066)
F2F x Mod. intellectual sim.		3.629		0.210			-0.353
		(1.628)		(0.251)			(1.056)
Grant coapplicant					2.989		
					(1.274)		
Coapplicant x Low field sim.					-2.913		
					(1.705)		
Coapplicant x Mod. field sim.					-2.635		
					(1.311)		
Coapplicant x Low intellectual sim.					0.901		
					(2.116)		
Coapplicant x Mod. field sim.					0.696		
					(1.415)		
Low field sim.	-0.403	-0.668	-0.00190	-0.0782	0.0455	-0.0671	-0.329
	(0.222)	(0.375)	(0.0113)	(0.0623)	(0.0406)	(0.0751)	(0.223)
Moderate field sim.	-0.109	-0.0510	0.0932	0.142	0.0529	0.293	0.447
	(0.142)	(0.189)	(0.0530)	(0.0860)	(0.0389)	(0.134)	(0.198)
Low intellectual sim.	-4.044	-3.989	-0.0331	-0.0378	-0.0290	-0.316	-0.282
	(0.244)	(0.283)	(0.0128)	(0.0237)	(0.00960)	(0.0752)	(0.106)
Moderate intellectual sim.	-0.973	-1.113	-0.0372	-0.0478	-0.0284	-0.254	-0.244
	(0.162)	(0.203)	(0.0130)	(0.0235)	(0.0111)	(0.0639)	(0.0997)
Previous coauthor	-0.302	-0.526	6.274	6.311	6.181	12.72	12.84
	(0.903)	(1.001)	(3.511)	(3.550)	(3.581)	(6.666)	(6.722)
Same institution	-0.0145	-0.0338	0.0691	0.0704	0.0657	0.127	0.141
	(0.107)	(0.111)	(0.0258)	(0.0275)	(0.0257)	(0.0588)	(0.0657)
Same department	-0.226	-0.271	0.0962	0.104	0.0924	0.316	0.358
	(0.171)	(0.186)	(0.0495)	(0.0598)	(0.0501)	(0.116)	(0.132)
Same imaging technology	-0.146	-0.171	0.0430	0.0442	0.0408	0.0461	0.0552
	(0.0923)	(0.0972)	(0.0321)	(0.0338)	(0.0315)	(0.0671)	(0.0712)
Both clinicians	-0.512	-0.521	0.0255	0.0254	0.0281	0.0117	0.0154
	(0.112)	(0.115)	(0.0172)	(0.0179)	(0.0174)	(0.0591)	(0.0607)
Both imagers	0.123	0.142	0.0204	0.0176	0.0208	0.136	0.124
	(0.172)	(0.177)	(0.0320)	(0.0336)	(0.0308)	(0.0929)	(0.0920)
Same rank	-0.457	-0.459	-0.0122	-0.0112	-0.0100	-0.0215	-0.0150
	(0.105)	(0.106)	(0.0148)	(0.0143)	(0.0147)	(0.0482)	(0.0467)
Both female	-0.367	-0.393	0.00837	0.00867	0.0109	-0.0332	-0.0252
	(0.142)	(0.148)	(0.0223)	(0.0241)	(0.0229)	(0.0541)	(0.0684)
Both male	0.229	0.235	0.0120	0.0126	0.0123	0.0613	0.0691
	(0.127)	(0.132)	(0.0143)	(0.0140)	(0.0137)	(0.0569)	(0.0565)
R-squared	0.235	0.227	0.076	0.060	0.084	0.062	0.049

All regression models include night and room FE. Multi-way, robust standard errors in parentheses.

TABLE A7 OLS Regression Models of Knowledge Transfer – % of MeSH Keywords Transferred between scientist-pair {*i,j*}; N = 28,258 (Full Sample)

VARIABLES	Model 1	Model 2	Model 3	Model 4
Same room	0.0882	0.0707	0.0537	0.0426
	(0.0807)	(0.147)	(0.106)	(0.159)
Same room x Low field sim.		0.0120		0.00516
		(0.143)		(0.143)
Same room x Mod. field sim.		0.356		0.323
		(0.325)		(0.328)
Same room x Low intellectual sim.			-0.0993	-0.0975
			(0.143)	(0.144)
Same room x Mod. intellectual sim.			0.199	0.197
			(0.135)	(0.135)
Low field sim.	0.166	0.163	0.165	0.163
	(0.112)	(0.116)	(0.112)	(0.115)
Moderate field sim.	-0.0567	-0.147	-0.0608	-0.142
	(0.225)	(0.212)	(0.224)	(0.212)
Low intellectual sim.	-4.586	-4.586	-4.561	-4.561
	(0.200)	(0.200)	(0.203)	(0.203)
Moderate intellectual sim.	-1.026	-1.026	-1.075	-1.075
	(0.146)	(0.146)	(0.153)	(0.153)
Night FE	Y	Y	Y	Y
Room FE	Y	Y	Y	Y
R-squared	0.263	0.263	0.264	0.264

Multi-way, robust standard errors in parentheses.

TABLE A8 OLS Regression Models of Knowledge Creation – # of copublications between scientist-pair {*i*,*j*}; N = 28,258 (Full Sample)

VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Same room	-0.0202	-0.0856	-0.0508	-0.112	-0.0235	-0.0240	-0.0247
	(0.0148)	(0.0557)	(0.0300)	(0.0670)	(0.0145)	(0.0144)	(0.0143)
Same room x Low field sim.		0.0727		0.0699			
		(0.0546)		(0.0537)			
Same room x Mod. field sim.		0.166		0.162			
		(0.100)		(0.0991)			
Same room x Low intellectual sim.			0.0388	0.0377			
			(0.0267)	(0.0259)			
Same room x Mod. intellectual sim.			0.0558	0.0530			
			(0.0276)	(0.0264)			
Gant coapplicant					4.030	4.030	5.206
					(1.248)	(1.248)	(1.472)
Coapplicant x Low field sim.					-1.388		-0.471
					(1.758)		(1.976)
Coapplicant x Mod. field sim.					-3.604		-4.775
					(1.302)		(1.531)
Coapplicant x Low intellectual sim.						-3.679	-3.891
						(1.673)	(1.924)
Coapplicant x Mod. intellectual sim.						-3.789	-3.961
						(1.481)	(1.803)
Low field sim.	-0.103	-0.122	-0.104	-0.121	-0.0844	-0.0875	-0.0848
	(0.0343)	(0.0460)	(0.0344)	(0.0458)	(0.0333)	(0.0342)	(0.0333)
Moderate field sim.	-0.0775	-0.120	-0.0781	-0.119	-0.0568	-0.0755	-0.0574
	(0.0387)	(0.0463)	(0.0387)	(0.0462)	(0.0390)	(0.0402)	(0.0390)
Low intellectual sim.	-0.0858	-0.0858	-0.0955	-0.0953	-0.0811	-0.0733	-0.0732
	(0.0179)	(0.0179)	(0.0227)	(0.0225)	(0.0175)	(0.0171)	(0.0171)
Moderate intellectual sim.	-0.0783	-0.0785	-0.0922	-0.0917	-0.0721	-0.0653	-0.0654
	(0.0166)	(0.0167)	(0.0211)	(0.0209)	(0.0161)	(0.0158)	(0.0157)
Night FE	Y	Y	Y	Y	Y	Y	Y
Room FE	Y	Y	Y	Y	Y	Y	Y
R-squared	0.004	0.004	0.004	0.004	0.027	0.027	0.034

Multi-way robust standard errors in parentheses.

TABLE A9 OLS Regression Models of Knowledge Diffusion – # of forward citations between scientist-pair {*i*,*j*}; N = 28,258 (Full Sample)

VARIABLES	Model 1	Model 2	Model 3	Model 4
Same room	-0.0408	-0.216	-0.0201	-0.193
	(0.0445)	(0.105)	(0.0951)	(0.134)
Same room x Low field sim.		0.191		0.193
		(0.102)		(0.102)
Same room x Mod. field sim.		0.577		0.575
		(0.328)		(0.326)
Same room x Low intellectual sim.			-0.0563	-0.0585
			(0.0916)	(0.0915)
Same room x Mod. intellectual sim.			-0.00884	-0.0171
			(0.0915)	(0.0909)
Low field sim.	-0.178	-0.227	-0.178	-0.227
	(0.0752)	(0.0954)	(0.0753)	(0.0952)
Moderate field sim.	-0.134	-0.281	-0.135	-0.280
	(0.114)	(0.107)	(0.114)	(0.106)
Low intellectual sim.	-0.288	-0.288	-0.274	-0.273
	(0.0582)	(0.0582)	(0.0636)	(0.0633)
Moderate intellectual sim.	-0.224	-0.224	-0.221	-0.220
	(0.0483)	(0.0484)	(0.0538)	(0.0535)
Night FE	Y	Y	Y	Y
Room FE	Y	Y	Y	Y
R-squared	0.008	0.008	0.008	0.008

Multi-way, robust standard errors in parentheses.