

Publisher: Institute for Operations Research and the Management Sciences (INFORMS)  
INFORMS is located in Maryland, USA



## Management Science

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

### Microgeography and the Direction of Inventive Activity

Christian Catalini

To cite this article:

Christian Catalini (2018) Microgeography and the Direction of Inventive Activity. *Management Science* 64(9):4348-4364.  
<https://doi.org/10.1287/mnsc.2017.2798>

Full terms and conditions of use: <https://pubsonline.informs.org/page/terms-and-conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact [permissions@informs.org](mailto:permissions@informs.org).

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2017, The Author(s)

Please scroll down for article—it is on subsequent pages

INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

# Microgeography and the Direction of Inventive Activity

Christian Catalini<sup>a</sup>

<sup>a</sup>MIT Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142

Contact: [christian@catalini.com](mailto:christian@catalini.com),  <http://orcid.org/0000-0003-1312-6705>

Received: December 2, 2016

Accepted: March 15, 2017


Published Online in Articles in Advance:  
July 19, 2017

<https://doi.org/10.1287/mnsc.2017.2798>

Copyright: © 2017 The Author(s)

**Abstract.** I provide novel empirical evidence grounded in an original theoretical framework to explain why colocation matters for the rate, direction, and quality of scientific collaboration. To address endogeneity concerns due to selection into colocation and matching, I exploit the constraints imposed on the spatial allocation of labs on the Jussieu campus of Paris by the removal of asbestos from its buildings. Consistent with search costs constituting a major friction to collaboration, colocation increases the likelihood of joint research by 3.5 times, an effect that is mostly driven by lab pairs that face higher search costs *ex ante*. Furthermore, separation does not negatively affect collaboration between previously colocated labs. However, while colocated labs grow increasingly similar in topics and literature cited, separated ones embark on less correlated research trajectories. Research outcomes, instead, seem to be mostly influenced by how distance affects execution costs: after colocation, labs are more likely to pursue both lower-quality projects (a selection effect) and high-quality projects (an effort effect). Opposite effects on quality are observed after separation. Whereas search costs affect which scientists are likely to collaborate together, execution costs shape the quality of their output.

**History:** Accepted by Ashish Arora, entrepreneurship and innovation.

 **Open Access Statement:** This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. You are free to download this work and share with others for any purpose, except commercially, if you distribute your contributions under the same license as the original, and you must attribute this work as “*Management Science*. Copyright © 2017 The Author(s). <https://doi.org/10.1287/mnsc.2017.2798>, used under a Creative Commons Attribution License: <https://creativecommons.org/licenses/by-nc-sa/4.0/>.”

**Funding:** This research was funded by the MIT Sloan School of Management, the Centre for Innovation and Entrepreneurship at the Rotman School of Management, and the Martin Prosperity Institute.

**Supplemental Material:** The online appendix is available at <https://doi.org/10.1287/mnsc.2017.2798>.

**Keywords:** colocation • idea recombination • collaboration • search costs • microgeography • low-opportunity cost time

“The truth is, it may have changed my whole life in a respect. I guess if I had any long-run thoughts then, it was to make a career doing statistics, econometrics, probability models, and things like that. But when I started talking on a regular basis with Paul, and he was so full of ideas and thoughts, it was impossible not to find my interests moving toward more straight economics. In a way the location of that office and the fact that we liked each other so much had a major influence on the direction my career took.”

—Nobel laureate Robert Solow describing how his office location and resulting friendship with Nobel laureate Paul Samuelson affected his career

(Dizikes 2011)

## 1. Introduction

When we select a location, we are committing to spending a disproportionate share of our most scarce resource, time, in that particular place. Individuals and organizations, anticipating the role geographic distance will play in their allocation of time, and in defining the opportunities and talent they will have access

to, pay particular attention to location decisions. As a result, location choices are highly endogenous to economic outcomes.

Geographic proximity not only increases the chance of a serendipitous interaction, but also lowers the cost of a scheduled one. Both types of interactions make it substantially easier to search for new collaborators within our local environment. When colocated, joint execution costs are also lower, as coordination, monitoring, and the transfer of complex information can all rely on more frequent, face-to-face interactions.

Thus, organizations spend considerable time and resources optimizing the spatial allocation of their teams and invest in infrastructure to allow for interdisciplinary work to flourish when new opportunities are identified. This endogenously shapes the trajectory of a university, R&D lab, or start-up, as the spatial allocation often results from priors about which teams and individuals will benefit the most from proximity. When a misalignment between the objectives of an organization and its current layout emerges, efforts are made to compensate for geographic distance, for example

by scheduling temporary collocation (e.g., joint meetings, conferences) and remote interactions to recreate the benefits of collocation. A recent, large-scale example is Microsoft's ambitious relocation of 1,200 engineers, which cut travel time between buildings for employees by 46% with the explicit objective of encouraging face-to-face conversations and serendipitous interactions instead of email and Skype meetings (Nielsen 2016).

Empirically, this makes it extremely challenging to understand *why* collocation matters and under which conditions different mechanisms are responsible for the benefits we attribute to proximity. The objective of this paper is to focus on how *search costs* and *joint execution costs* shape inventive outcomes, and to provide novel empirical evidence and a theoretical framework that can help us separate between these two competing, but not necessarily exclusive, mechanisms.

If search costs are a key friction to collaboration, then collocation should have a positive effect on the probability of collaboration, but separation may not necessarily have a negative effect, as teams that are aware of each other may be able to compensate for distance through temporary collocation and remote interactions. If instead joint execution costs are driving collaboration decisions, then the effects observed after collocation and separation should be exact opposites of each other (as execution costs would increase almost immediately with distance). Furthermore, whereas search costs do not have a strong implication for the observed value of joint projects, if lower execution costs allow teams to apply more effort toward advancing their ideas, then the effect of collocation on quality will be ambiguous. On the one hand, lower execution costs may induce teams to select lower-quality projects (a selection effect). On the other hand, if lower execution costs allow teams to endogenously apply more effort—and effort improves the quality of the underlying idea—we may also observe an increase in right-tail outcomes after collocation (an effort effect).<sup>1</sup>

I explore the relative role of these mechanisms in a setting where the spatial allocation is constrained by reasons that are orthogonal to the outcomes of interest—i.e., where exogenous variation is injected in the process of deciding where different teams are placed. Combined with a difference-in-differences approach, this mitigates the endogeneity concerns typically linked to selection into collocation.

The setting is the university campus of Paris Jussieu, the leading scientific and medical complex in France. Following a research report by INSERM on the carcinogenic effects of asbestos (June 21, 1996), the French government introduced a full ban of the fire retardant from all public buildings, including the Jussieu ones. Starting in 1997, a separate entity (Etablissement Public d'Aménagement Universitaire de la Région Ile-de-France (EPAURIF)) was put in charge of the asbestos

removal process on campus, which led to five massive waves of lab relocations over 17 years. Because of the complexity and urgency<sup>2</sup> of the cleaning process, labs were forced to move often on short notice and with little influence over their new location. As a consequence, many labs found themselves next to new neighbors, and the overall spatial allocation was severely more constrained than before. I exploit this variation, combined with the reconstruction of fine-grained, longitudinal information on inventive outcomes, to try to understand why collocation matters for the rate, direction, and quality of inventive activity.

Consistent with the search costs mechanism, results show that collocation increases the likelihood that two labs will collaborate by 3.5 times, an effect that is mostly driven by pairs that faced higher search costs *ex ante*. Moreover, separation has a nonsignificant effect on the probability and rate of collaboration, suggesting that labs that were previously exposed to each other are later able to sustain collaboration also over distance. At the same time, while colocated labs become increasingly similar in the topics they work on and references they cite,<sup>3</sup> separated labs embark on less correlated research trajectories. This is consistent with search costs increasing considerably even at relatively low levels of geographic distance, a result that supports past literature that has linked proximity to better information diffusion (Allen 1977, Cowgill et al. 2009), formation of social ties (Stuart and Liu 2010, Liu 2013), and knowledge flows (Jaffe et al. 1993, Thompson and Fox-Kean 2005, Thompson 2006).

Furthermore, conditional on collaboration, the quality of a lab pair's output also changes following a change in distance. Colocated labs are 1.36 times more likely to produce a paper that will end up in the highest quartile of the citation distribution, and their variance in outcomes increases. Opposite results are observed for lab pairs that are separated. This is consistent with collocation affecting joint execution costs, and with both a selection and an effort effect playing a role in this context. Interestingly, the collaborations resulting from interactions between labs that faced higher search costs *ex ante* are more likely to be of high impact, suggesting that arbitrage opportunities may exist in encouraging interactions between communities of scientists that do not overlap through other channels (e.g., joint conferences and journals).

Taken together, the findings highlight that whereas search costs mostly affect which scientists are likely to collaborate together, execution costs shape the quality of their output. By allocating space, organizations profoundly shape the evolution of scientific trajectories and the types of opportunities that are explored by different teams. This involves a trade-off between the perhaps more efficient exploitation of established research paradigms and the more costly exploration

of new ones. Colocation is an expensive way to lower search costs, as supported by the overall changes in the collaboration portfolios of the labs observed in the data: during the moves, in aggregate, labs focused more inwards, increasing within-lab research at the expense of across-lab collaborations. While this may be a result of the suboptimal set of local peers the moves offered the labs relative to their ideal choices, it is also a reminder of how pursuing research across disciplines is a more costly endeavor than incremental, within-discipline work both for an institution and the scientists involved. In the absence of complementary changes in incentives (e.g., to favor cross-disciplinary research), scientists will focus where search and joint execution costs are lower first. Although temporary forms of colocation may not be as effective as the longer periods studied here in leading to actual knowledge flows and collaboration, they may still allow for cross-pollination between research trajectories and breakthrough discoveries at a lower cost.

The layout of the remainder of the paper is as follows. In the next section, I further develop the basic hypotheses of the theoretical framework to guide empirical predictions and interpretations of the findings. Section 3 describes the empirical setting, data, and empirical strategy. Section 4 reports the main results. Section 5 concludes.

## 2. Theoretical Framework

The ability of an economy to generate, diffuse, and recombine ideas has a profound influence on its ability to sustain growth (Lucas 1988, Romer 1990, Weitzman 1998). Our understanding of a key economic phenomenon behind agglomeration and growth—localization economies—relies on, among other factors, basic assumptions about how knowledge is recombined locally versus over distance.

Despite frequent references in the literature on knowledge flows and localization to the role of colocation in knowledge transmission and recombination (Breschi and Lissoni 2004, Mairesse and Turner 2005, Singh 2005, Agrawal et al. 2006, Fleming et al. 2007, Belenzon and Schankerman 2013), there is more limited empirical evidence on the underlying mechanisms invoked to explain its effects. In part, this is due to the difficulty of finding plausibly exogenous variation in location choice. Most of the existing literature relies on observational data (Olson and Olson 2000, Van den Bulte and Moenaert 1998, Kabo et al. 2014, Fayard and Weeks 2007, Kabo et al. 2015, Crescenzi et al. 2017), which is subject to selection bias from individuals, teams, and organizations choosing where to locate. This makes it challenging to isolate the effect of colocation from confounders and other forms of proximity, such as proximity in social space and in knowledge

space. Knowledge production, moreover, is increasingly a collaborative process (Wuchty et al. 2007, Jones et al. 2008) between colocated and geographically dispersed teams of scientist (Adams et al. 2005, Katz 1994, Freeman et al. 2015), making it difficult to trace knowledge flows through these alternative channels.

Scientists are evaluated based on the quality and quantity of their output (e.g., in terms of scientific impact measured through citations, publications, outlets). Two key costs shape how scientists (and teams of scientists) make collaboration decisions: search costs and execution costs. Search costs, by defining the choice set of possible collaborators, influence the likelihood that any two individuals will explore a joint research project to begin with. Execution costs determine if, given an idea of a certain quality, it makes sense for a specific team to invest time, resources, and effort in developing it.

Colocation has a profound effect on both costs: (1) by increasing the chance of an interaction (both serendipitous and planned), colocation drastically lowers search costs for new collaborators; (2) by lowering the cost of face-to-face meetings, coordination costs, monitoring costs, and the cost of transferring complex information, colocation also reduces joint execution costs. If by applying effort toward a project scientists are able to improve its ultimate value, then lower execution costs, locally, endogenously change the optimal level of effort scientists may want to apply to local versus distant projects.

In the next sections, I focus on how geographic proximity influences search and execution costs to build predictions about how microgeography affects the rate and type of scientific collaborations between teams of scientists (labs).

### 2.1. Search Costs

In this context, search costs are defined as the frictions scientists incur in finding new collaborators and collaboration opportunities. A reduction in search costs should therefore induce an increase in the probability of collaboration between two labs and possibly shift over time the collaboration portfolios of the scientists affected. Recent experimental evidence has shown that even within the same university, search frictions can be substantial: after randomly collocating scientists in the same room for a 90-minute information-sharing session, Boudreau et al. (2017) observe a 75% increase in the probability that they will co-apply for a grant.

An opposite increase in search costs, instead, may have a more ambiguous effect on the probability of collaboration, especially if we believe that once two scientists are aware of each other's research agenda, they can keep communicating new ideas cheaply over distance—i.e., when it comes to search, if social proximity (e.g., past collaboration) can partially substitute

for geographic proximity, then separating groups of scientists that were previously colocated may have a smaller effect (or no effect at all) on collaboration.<sup>4</sup>

Search also takes place through temporary collocation (e.g., conferences),<sup>5</sup> and through codified information (e.g., published research). As a result, in cases where search costs are likely to be already low ex ante because scientists attend the same conferences or read the same journals, changes in geographic proximity may have a less pronounced effect on the likelihood of collaboration.

This can be summarized in the following hypothesis:

**Hypothesis 1.** *After collocation, scientists should be more likely to collaborate, and the effect will be stronger the higher search costs are ex ante. Separation, instead, should have no effect on the probability of collaboration.*

## 2.2. Execution Costs

Joint execution costs include the cost of face-to-face meetings, coordination, monitoring, and transfer of complex information between teams of scientists. The first effect of a reduction in joint execution costs is, as in the case of search costs, an increase in the probability of collaboration (this time possibly skewed toward preexisting pairs). After collocation, scientists should collaborate together more.<sup>6</sup> At the same time, while search costs may not increase immediately after separation because of the long-term benefits from past exposure, execution costs should increase relatively quickly, negatively affecting joint projects. To summarize:

**Hypothesis 2.** *After collocation, scientists should be more likely to collaborate with each other (and increase their rate of collaboration) because of lower execution costs. After separation, the opposite should be observed.*

Whereas the effect of a decrease in execution costs on the rate of collaboration is straightforward, the effect on the realized value of the resulting projects is ambiguous. If the value of a project depends both on the original idea's intrinsic quality as well as on the amount of effort dedicated to advancing it, then collocation may induce: (1) the development of lower-quality ideas (a selection effect); (2) the application of effort for any given idea quality (an effort effect). In other words, if we assume that research projects improve when more effort is allocated to their development, and that it is cheaper to do so when colocated (lower execution costs), then the ultimate impact of collocation on outcomes is the composition of the selection and effort effect pushing in opposite directions.

On the one hand, lower execution costs locally may induce scientists to engage in marginal, lower-value projects. Intuitively, if projects below a certain value cannot be published (i.e., if there is a minimum threshold for publication) and effort improves the payoff of an idea, then the threshold for developing a research

project under collocation will be lower. Conditional on an idea being developed, this should lead to an increase on the left tail of the outcome distribution.

On the other hand, if scientists can improve a project more efficiently when colocated (e.g., through additional face-to-face meetings, better knowledge transfer, team coordination, etc.), this leads to the expected value of colocated projects being higher for any underlying idea quality. Intuitively, scientists can achieve the same outcome with ideas of lower starting quality because they can execute on them more efficiently. Additionally, it can be shown<sup>7</sup> that if the distribution of idea quality is skewed toward the low end (i.e., if most scientific ideas are of low impact and a few are of very high impact), then the difference in effort is a (weakly) increasing function of idea quality (i.e., the difference in effort between colocated and distant projects increases with project quality). This leads to the following hypothesis:

**Hypothesis 3.** *After collocation, conditional on projects being developed, we should observe more projects at the low end of the distribution (selection effect) and at the high end of the distribution (effort effect). Separation should lead to opposite results.*

The above prediction relies on the skewed nature of inventive outcomes, on the idea that collocation offers lower execution costs, and on the assumption that more effort improves the value of a project.

Empirically, a key challenge is due to the fact that whenever scientists have unrealistic expectations about the value of a project (e.g., when they believe a project is of publishable quality when it is not), or whenever projects are abandoned as scientists update their priors, part of the left tail of the outcome distribution will not be observed (as we only see published research), leading to underestimating the effects of collocation and separation on lower-quality ideas. This limits what we can learn from bibliometrics data regarding the full distribution of research outcomes. The analysis on quality in Section 4.3 will be therefore more informative about right-tail outcomes and only provide suggestive evidence about the left tail.

Moreover, because of lower search and execution costs, collocation is a strategic choice, and observational data will overestimate the impact of these effects on inventive outcomes. The relocations of labs on the Jussieu campus, because of the external constraints imposed on space allocation by the asbestos removal process, can help us understand if the mechanisms suggested in the theoretical framework are correct, and improve our understanding of the conditions under which collocation is more likely to matter for collaborative inventive outcomes. To address additional concerns about selection into collocation and matching between labs, the key part of the empirical analysis

relies on a difference-in-differences approach at the lab-pair level, as described in the next section. Lab-level portfolio choices are also presented to understand the aggregate effects of the moves on the trade-offs labs make in deciding who to collaborate with.

### 3. Empirical Setting

The setting is the university campus of Paris Jussieu, which hosts the Université Pierre et Marie Curie (UPMC). According to the 2016 U.S. News and World Report rankings, UPMC is the top institution in France (10th in Europe, 49th on a global scale). It houses the faculty of sciences of Sorbonne Universités and approximately 31,000 students (5,900 master students and 3,000 doctoral students), as well as 3,750 research-active professors (80% of its staff work in research centers).<sup>8</sup> Three recent Nobel laureates are from UPMC: Pierre-Gilles de Gennes (Physics, 1991), Claude Cohen-Tannoudji (Physics, 1997), and Françoise Barré Sinoussi (Medicine, 2008). Strong areas of specialization include mathematics (fifth worldwide), physics (15th), space science (20th), geosciences (21st), neuroscience (43rd), environment and ecology (49th), and biology and biochemistry (51st).

The campus went through five massive waves of labs relocations over 17 years (1997–2014) due to the removal of asbestos from its buildings. The moves started when the French government, reacting to a research report by INSERM on the carcinogenic effects of asbestos (June 21, 1996),<sup>9</sup> introduced a full ban of the fire retardant material from public buildings. The Jussieu campus was built extensively using asbestos.

Interviews with scientists from the labs confirmed that given the nature and urgency<sup>10</sup> of the cleaning process, labs were forced to move often under short notice and with minimal influence over their new location. A separate entity (EPAURIF) was put in charge of the cleaning process, which started with labs that were relatively easy to move (e.g., theoretical labs in mathematics, computer science) and only later reached labs with sophisticated instrumentation and machinery (e.g., applied physics). During the relocation period, entire sections of the campus were progressively isolated and renovated. Because of the complexity and costs of the operation, lab requirements were often not a priority, resulting in many labs complaining about the moves. Whereas for some scientists the actual moving process was a cause of delays (one scientist estimated a one-year delay in productivity over a 10-year period), others did not find it disruptive at all (a different scientist, who does theoretical work, estimated just one week without lab access).

As in other research-intensive institutions, interactions across labs are common, and the relocations ended up separating labs that were colocated and interacted before. The same moves also placed labs next to

new neighbors, in some cases with positive effects on collaboration. In the data set, the aggregate amount of colocation between lab pairs slightly decreases during the moves, from 7.3% to 6.9% of pairs.<sup>11</sup> Whereas the mean distance between lab pairs across broad fields of science (e.g., natural versus life sciences) and within subfields (e.g., within chemistry) is broadly stable over time, the distance across subfields (e.g., chemistry and physics) decreases, providing plausibly exogenous variation in the composition of a lab's neighbors. The data also provides substantial variation in terms of which types of lab pairs are affected by the moves at any point in time.<sup>12</sup> The mean number of collaborations per year for colocated lab pairs increases from 0.034 during the pre-period to 0.042 during the moves. The change is substantially larger for pairs that are within a broad field, from 0.039 to 0.056.

An academic campus is an appealing environment for studying the role of geographic proximity on collaboration: knowledge production is one of its defining activities, it is possible to measure collaborations and their long-run impact, knowledge flows can be partially captured by looking at cited references, and research agendas can be mapped into knowledge space using keywords and cited references. Moreover, labs increasingly rely on a mix of proximate and distant collaborators (both in geographic and knowledge space) for advancing their research agendas.

#### 3.1. Data

The data set combines information on 39,527 publications from the labs at Jussieu (1980–2010) with fine-grained location data over time. Publications are retrieved from SciVerse Scopus<sup>13</sup> and parsed to extract affiliation data. Forty-two thousand, four hundred ninety-four unique affiliation strings are cleaned and harmonized with a series of algorithmic and manual steps to match them to a specific lab.<sup>14</sup> Whenever location data is available in the papers<sup>15</sup> it is extracted to complement information retrieved from the UPMC archives and the EPAURIF website<sup>16</sup> to reconstruct the spatial allocation of labs over time. Paper affiliations that are not matched to the campus are geocoded using a combination of three different services (Google Maps API, Bing Maps API, and the Data Science Toolkit) to identify them as either French or international affiliations.

The core of the campus resembles a chessboard (see Online Appendix Figure A-1) and is composed of a series of towers connected by corridor buildings. Distances are obtained by manually geocoding the location of each tower and connecting building on Google Earth.<sup>17</sup>

Out of 39,527 publications, 6% are collaborations across different labs at Jussieu. In the final data set (see Table 1), the average minimum distance between any

**Table 1.** Summary Statistics for the Main Sample

Lab-pair-year level	Mean	Std. dev.	Min	Max	N
<b>Location</b>					
Minimum distance (km)	0.17	0.106	0	0.449	183,359
Colocated (same building)	0.072	0.258	0	1	295,435
Treatment year (colocation)	2,003.99	3.722	1,997	2,010	7,338
Treatment year (separation)	2,003.63	4.089	1,997	2,010	5,877
<b>Collaboration</b>					
Probability of collaboration	0.004	0.064	0	1	295,435
Number of collaborations	0.006	0.126	0	15	295,35
<b>Quality and type of research</b>					
Citation weighted collaborations	0.204	7.328	0	2,058	295,435
Maximum number of citations	0.164	6.053	0	2,058	295,435
Standard deviation of citations <sup>a</sup>	25.666	38.964	0	271.442	321
Keyword similarity	0.076	0.108	0	0.881	116,851
Cited references similarity	0.174	0.115	0	0.677	84,865

<sup>a</sup>Conditional on collaborating at least twice in focal year.

two lab pairs is 0.17 km (approximately 550 feet) and 7.2% of them are colocated at any point in time (same tower or corridor building).<sup>18</sup> Over time, 383 lab pairs switch from not colocated to being in the same building during the asbestos cleaning process (at a rate of 16–52 each year), with 37% of the changes happening in the first five years. Two hundred forty-eight lab pairs experience the opposite change (i.e., are separated by the moves), with 34% of the events taking place within five years.

Collaboration across labs is a rare event, and only 0.4% of all lab pairs collaborate in any given year, receiving on average 0.20 citations per lab-pair-year. As often observed, the distribution of citations is very skewed, with most papers receiving no citations, and the most productive lab pair totaling 2,058 citations from the papers published in a single year.

Citation data is obtained from Scopus in 2016, and quartiles for the citation distribution are built by year using a large sample of articles in the relevant fields of science. Scopus data is also used to retrieve, whenever available, author and index keywords,<sup>19</sup> as well as the full set of references cited by each paper.

Cited references and keyword data are then used to create measures of proximity in knowledge space between the labs. In particular, the cosine similarity between vectors of cited references (or keywords) used by each lab is calculated using the Scikit-Learn python module.<sup>20</sup> To ensure that these proximity measures are not inflated by direct collaborations (which would share all cited references and keywords by design), coauthored papers between the focal labs are dropped—i.e., the vectors of cited references (or keywords) are based on independent publications. Moreover, when the sample is split by lab pairs with above versus below the median distance in knowledge space (e.g., Table 7), the measure is defined before the moves

start to avoid it being influenced by the effect of colocation (or separation). In the data set, labs that do not share any keyword have a cosine similarity of zero, the pair with the most overlap has a score of 0.88,<sup>21</sup> and the mean keyword similarity is 0.08. Similarity in cited references is on average higher (0.17), but the highest observed score is lower (0.68).

### 3.2. Empirical Strategy

Estimating the effect of colocation on inventive outcomes using observational data is likely to return biased results, the main endogeneity concern being a selection effect. If labs value proximity to other labs they want to collaborate with, then a basic ordinary least-squares (OLS) regression of the number of collaborations on geographic distance will positively bias the effect of proximity on collaboration. Whereas colocated lab pairs were 2.75 times more likely to collaborate than not-colocated pairs before the moves started,<sup>22</sup> the premium is only 14% (and not significant) during the relocations, which is consistent with the new spatial allocation being suboptimal in the post-period.

In a perfect experiment, we would randomize the location of labs and observe how collaborations between colocated pairs differ from those between distant ones. During the asbestos removal, qualitative evidence confirmed that needs that were orthogonal to the research agendas of the scientists involved (e.g., ease of relocation of a type of lab, cost minimization for the removal operation) constrained space assignment and shaped *when* and *where* different labs were moved. This introduces exogenous variation in the set of neighbors a lab is offered.<sup>23</sup> To account for idiosyncratic reasons a lab pair may be more or less likely to collaborate in the first place, the analysis uses lab-pair fixed effects. Importantly, the pair fixed effects also capture the degree of influence a particular pair of labs may have on campus resource allocation decisions (funding, personnel, space allocation).

The econometric analysis uses a difference-in-differences approach<sup>24</sup> at the lab-pair-year level which exploits the variation in colocation (and separation) generated by the moves that take place during the asbestos-removal period. I estimate variations of

$$Y_{ijt} = \gamma_{ij} + \delta_t + \beta \text{AfterColocation}_{ijt} + \epsilon_{ijt}, \quad (1)$$

where  $Y_{ijt}$  is a dummy for collaboration in focal year<sup>25</sup> between lab  $i$  and lab  $j$  in year  $t$ ;  $\beta$  is the coefficient of interest and the lab-pair fixed effects,  $\gamma_{ij}$ , mean that  $\beta$  is identified as the within-pair effect on collaboration after labs become colocated because of the asbestos removal, relative to their before period; and  $\delta_t$  is a year effect. The *AfterColocation* dummy is equal to 1 if the lab pair becomes colocated because of the moves, and is set back to 0 when the labs are separated again. Similarly, I estimate the effect of separation in the same regression using variations of

$$Y_{ijt} = \gamma_{ij} + \delta_t + \beta_1 \text{AfterColocation}_{ijt} + \beta_2 \text{AfterSeparation}_{ijt} + \epsilon_{ijt}. \quad (2)$$

Both equations are also estimated with nonlinear models and for different outcomes: number of collaborations in focal year, maximum number of citations received (conditional on publication), number of papers published in different quartiles of the citation distribution, similarity in cited references space and in keyword space, probability of publishing in a journal that is new for at least one of the labs, etc.

Before moving to the difference-in-differences framework, the paper uses regressions at the lab-year level for the same set of outcomes to explore how the relocation period influenced collaboration and the composition of the labs' portfolios (in terms of share of affiliations classified as same lab, different lab, other French lab or international lab, etc.). All lab-year-level regressions include lab fixed effects and year fixed effects.

## 4. Results

The first section of the results focuses on assessing the overall effect of the moves on the labs involved. Empirically, it is important to determine if the relocation period had a negative effect on the labs' productivity, and how labs adjusted their collaboration portfolios because of the changes in their neighbors. Since this part of the analysis does not rely on the difference-in-differences empirical strategy described above, results will only be informative but not conclusive with respect to the mechanisms proposed in the theoretical framework.

The second section will directly test Hypothesis 1 on the effects of a change in proximity on the probability of collaboration. This is done by looking at pairs of labs that find themselves in the same building because of

the asbestos removal, as well as pairs that are separated because of it. In particular, separated pairs will be informative about the ability of other forms of proximity (e.g., social proximity) to compensate for geographic distance. Robustness is shown to highlight the absence of a pre-trend in collaboration among pairs that are going to be colocated, which is consistent with the moves not being driven by the labs' research agendas.

Knowledge distance between labs is then introduced as a way to compare lab pairs with different ex ante search costs, and to see if the changes in the probability of collaboration are consistent with the search cost mechanism. Since Hypotheses 1 and 2 do not differ in terms of what they predict we should observe after colocation, but differ in terms of what we should see after separation, this section will also compare the relative role of search versus execution costs in defining who collaborates. Lastly, the section explores the effect of search costs over longer periods of time by using information on the references cited in the affected papers.

The third section introduces quality (as proxied by citations) to test Hypothesis 3, and to see if outcomes are consistent with lower execution costs under colocation leading to both more marginal ideas being developed (selection effect), as well as higher-quality ideas (effort effect).

### 4.1. Lab-Level Results and Collaboration Portfolios

One may worry that the relocations on the Jussieu campus had a negative effect on the productivity of the labs. This would influence how we would interpret the results from the difference-in-differences analysis in the next sections of the paper. Furthermore, since labs are likely to reallocate their resources toward different types of projects as their local environment changes, before moving to the lab-pair-level regressions, it is useful to descriptively explore how aggregate collaboration portfolios shifted during the study period because of the moves.

The unit of analysis in Table 2 is a lab-year, lab fixed effects are included to account for unobservable differences between labs that are constant over time (e.g., field of science, relative scale and resources of the lab within the institution, etc.), and year fixed effects are introduced to control for changes in productivity over time (e.g., increased resources available to the campus, changes in the national science policy, etc.).<sup>26</sup> The key explanatory variable in the table is a dummy equal to 1 during the *Relocation Period*—i.e., during all of the years during which a specific lab is moved from its original location because of the asbestos removal (and 0 otherwise). As can be seen both in terms of raw publication counts (column (1)), and in terms of quality-adjusted output (column (2)), the moves are not correlated with a decrease in output for the affected labs relative to



**Table 2.** Lab-Level Outcomes

Variables	(1) Publications	(2) Citation weighted pubs	(3) Mean cites	(4) Max cites	(5) Std. dev. cites
<i>Relocation Period</i>	10.8003*** (2.3118)	246.6649*** (82.7443)	−0.0157 (1.6609)	28.0384* (14.6638)	6.4865** (2.6019)
Lab fixed effects	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	3,951	3,951	3,951	3,951	3,951
R-squared	0.161	0.070	0.015	0.015	0.011
Number of labs	328	328	328	328	328

Notes. *Relocation Period* is a dummy equal to 1 for all of the years during which a specific lab is moved from its original location because of the asbestos removal. Citation data is collected from Scopus in 2016. Robust standard errors clustered at the lab level in parentheses.

\* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

the controls (i.e., labs that have not been moved yet or that have returned to their original location). Interestingly, results on quality are suggestive of an increase on both tails of the outcome distribution: the effect on mean citations is negative although nonsignificant (column (3)), max citations increase (column (4)), and the moves are correlated with a positive increase in the standard deviation of citations (column (5)).

It is important to highlight that multiple mechanisms, including changes in neighboring labs, infrastructure, and research type, can explain these results. For example, labs could have changed the composition of their projects in response to shifts in their local environment in a way that favored higher-variance projects—e.g., if the moves increased the likelihood that a lab had access to peers from a different discipline, the polarization in outcomes could be a result of an increase in cross-disciplinary work. This would be consistent with past research that has shown that recombination of ideas that span more distant areas of the knowledge space exhibit higher variance (e.g., because the ideas involved are recombined less often and therefore represent a less explored area of the research landscape; see Fleming and Sorenson 2004, Singh and Fleming 2010). In this context, while keeping the expected value of their projects constant (i.e., while staying on their original risk–return indifference curve), labs could have undertaken some projects that offered higher risk but also higher reward. Alternatively, the moves could have induced a temporary, one-time shift in research agendas because of the reshuffling of scientists and labs (a “novelty” effect).

For these reasons, the next section will move away from the lab-level analysis and control for many of these confounders by looking at variation within lab pairs. With the introduction of lab-pair fixed effects, the difference-in-differences approach allows to account for the idiosyncratic reasons two labs may (or may not) collaborate with each other, and controls for unobserved heterogeneity in the type of research a particular pair may be conducting. Within a lab pair, the constraints imposed by the moves will deliver plausibly

exogenous variation in distance, allowing us to look at outcomes while keeping issues related to the matching between labs constant—i.e., within a lab pair, changes in collaboration and outcomes will be predominantly driven by shifts in proximity. Additionally, the analysis of pre-trends in collaboration within the difference-in-differences framework will allow us to visually investigate the presence of selection into colocation during the moves—i.e., if labs are endogenously paired in an effort to improve their collaborations, this should be observable in the data.

Before moving to the lab-pair analysis, it is useful to check how the relocation period shifted the collaboration portfolios of the labs at the aggregate level. Table 3 uses the same specification of the previous table to look at how the shares of affiliations on the labs’ papers changed during the moves. The dependent variables here are, respectively, the share of affiliations that are international (column (1)), within France but not from the institution (column (2)), and within the institution (column (3)). Interestingly, the moves saw a decrease in the share of international collaborators; a

**Table 3.** Lab Level—Overall Collaboration Portfolio

Variables	(1) Share international	(2) Share French	(3) Share within institution
<i>Relocation Period</i>	−0.0237* (0.0121)	0.0018 (0.0137)	0.0218 (0.0146)
Lab fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Observations	3,951	3,951	3,951
R-squared	0.144	0.072	0.207
Number of labs	328	328	328

Notes. *Relocation Period* is a dummy equal to 1 for all of the years during which a specific lab is moved from its original location because of the asbestos removal. Shares are calculated based on the Scopus author affiliations data available for each paper and then aggregated at the lab level on a yearly basis. “Share French” does not include the Jussieu campus and affiliated labs. Robust standard errors clustered at the lab level in parentheses.

\* $p < 0.1$ .

**Table 4.** Lab Level—Collaboration Portfolio Within Institution

Variables	(1) Share across labs	(2) Colocated	(3) Not colocated	(4) Intensive margin	(5) Extensive margin	(6) Colocated intensive margin	(7) Colocated extensive margin	(8) Not colocated intensive margin	(9) Not colocated extensive margin
<i>Relocation Period</i>	−0.0210** (0.0093)	0.0155*** (0.0048)	−0.0365*** (0.0082)	−0.0035 (0.0066)	−0.0175** (0.0075)	0.0030 (0.0037)	0.0125*** (0.0031)	−0.0065 (0.0051)	−0.0300*** (0.0075)
Lab fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,951	3,951	3,951	3,951	3,951	3,951	3,951	3,951	3,951
R-squared	0.022	0.023	0.034	0.065	0.130	0.036	0.035	0.041	0.119
Number of labs	328	328	328	328	328	328	328	328	328

*Notes.* *Relocation Period* is a dummy equal to 1 for all of the years during which a specific lab is moved from its original location because of the asbestos removal. All shares are calculated based on the Scopus author affiliations data available for each paper and then aggregated at the lab level on a yearly basis. The table only includes affiliations that can be assigned to the Jussieu campus and affiliated labs. Robust standard errors clustered at the lab level in parentheses.

\*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

nonsignificant, small positive change in collaborators from other French institutions; and an increase in within-institution research. While the regression controls for lab fixed effects and year fixed effects, as in the previous table multiple mechanisms could explain these changes. One interpretation is that the campus lost competitiveness because of the moves, becoming more inward focused. A different one is that the reshuffling improved local opportunities by offering labs new neighbors, making outside collaboration relatively less appealing.

To further test how internal collaboration was affected, Table 4 uses the same approach of Table 3 but only focuses on internal collaborations (i.e., all of the shares are now defined over the total number of affiliations assigned to the institution). Column (1) highlights how the increase in within-institution research observed in column (3) of Table 3 is entirely driven by collaborations within labs: the share of across-lab collaborations actually decreases, suggesting that the moves may have worsened the local environment, forcing labs to increase collaboration within their unit. This is consistent with the relocations leading to a suboptimal space allocation relative to what the labs would have selected if they were in charge of it. The next two pairs of columns in Table 4, take the share from column (1) and further decompose it across different dimensions. In particular, columns (2) and (3) split the collaborations across labs (i.e., column (1)) between labs that are colocated versus labs that are distant: the negative result from the first column comes from a drastic decrease in the share of collaborators from not colocated labs (column (3)) that is not compensated by an equal increase in collaborations with colocated labs (column (2)). When column (1) is instead decomposed into intensive margin versus extensive margin collaborations (i.e., between labs that had collaborated

before the relocations versus not), the data show that the decay in across-lab collaborations is mostly driven by extensive margin pairs—i.e., during the relocations, labs were substantially less likely to explore research with labs they had not collaborated with before.

The aggregate result of columns (4) and (5), however, hide different heterogeneous effects by microgeography: in columns (6)–(9), the share of column (1) is divided into colocated (intensive versus extensive) and not colocated (intensive versus extensive) collaborations. Whereas the results are consistent with geography only slightly facilitating (or obstructing) collaborative work on the intensive margin (the coefficients are respectively positive in column (6) and negative in column (8), but in both cases not significant), columns (7) and (8) suggest that microgeography plays a substantially more important role on the extensive margin of collaboration. This seems consistent with search costs having a disproportionate effect on defining who collaborates with whom in the absence of past exposure or social proximity. This hypothesis will be further tested in the next section.

Overall, at least in the Jussieu case, the increase in collaboration with newly colocated pairs (columns (6) and (7)) seems to be more than offset by the decrease with not colocated labs (columns (8) and (9)). Furthermore, the result in column (1) (combined with the effect of column (3) in Table 3) is consistent with labs being more inward focused during the moves, potentially because they faced a less optimal set of local peers around them relative to what they would have selected in an ideal scenario. Together with the observed increase in best outcomes and variance in Table 2, and with the rise in colocated experimentation on the extensive margin (column (7) in Table 4), this raises the question of how collocation and separation directly affected research outcomes once we keep lab pairs constant.

## 4.2. Probability of Collaboration and the Role of Search Costs

In the theoretical framework, both a reduction in search costs (Hypothesis 1) and in joint execution costs (Hypothesis 2) predict an increase in the probability of collaboration after colocation. If joint execution costs are particularly important, then we should also observe a tangible rise in the rate of collaboration. The two mechanisms have different implications for the reverse move though—i.e., for what we should observe after two labs that were previously colocated are suddenly separated. If search costs do not increase immediately after separation because of past exposure, then separation should have little to no effect on the probability of collaboration. If joint execution costs are instead disproportionately driving collaboration decisions, then separation should be accompanied with a drop in collaboration. This section tests these hypotheses by first looking at within-lab-pair changes in collaboration following changes in distance, and then by exploring heterogeneous effects by pairs that faced *ex ante* higher versus lower search costs to compare the two mechanisms more directly.

The first part of Hypotheses 1 and 2 are tested in Table 5: after colocation, lab pairs are 3.5 times more likely to collaborate (column (1)) and collaborate on average 2.5–3.3 times more (columns (3) and (2), respectively). Results are robust and qualitatively similar across different functional forms (OLS, Poisson, logit, rare event logit). Robust standard errors are clustered at the lab-pair level, and all regressions include lab-pair fixed effects as well as year fixed effects. The estimated effects do not change if lab-year trends are included in the regression. The findings are consistent with lower search and execution costs increasing the

**Table 5.** Increase in Collaboration and Colocation

Variables	(1) OLS 1/0	(2) OLS #Collabs	(3) Poisson #Collabs
<i>After Colocation</i>	0.0146*** (0.0054)	0.0210** (0.0094)	0.8981*** (0.2814)
Lab-pair fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Observations	295,435	295,435	10,383
R-squared	0.005	0.005	
Number of lab pairs	35,805	35,805	587
Log likelihood			-2,929

*Notes.* The dependent variable in column (1) is a dummy equal to 1 if the lab pair collaborates in the focal year. The dependent variable in columns (2) and (3) is the number of collaborations. *After Colocation* is equal to 1 when a lab pair becomes colocated because of the moves and 0 otherwise. The Poisson specification with fixed effects drops all lab pairs where collaboration is never observed, hence the smaller number of observations. Robust standard errors clustered at the lab-pair level in parentheses.

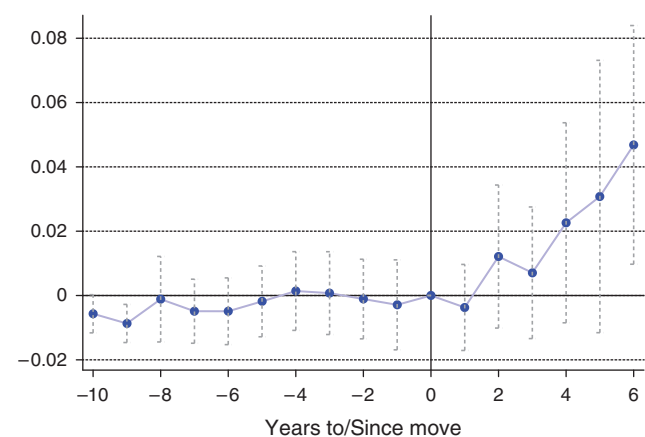
\*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

likelihood (column (1)) and rate (columns (2) and (3)) of collaboration.

It is important to stress that the presence of the lab-pair fixed effects takes care of the idiosyncratic influence a pair of labs may have on campus (e.g., in terms of funding, ability to bargain for more or better resources and infrastructure). Nevertheless, one may still worry that influential labs might have exerted pressure on EPAURIF to assign them a particular temporary location, or to change the timing of their move. From an identification perspective, the main worry is that the moves were driven by preexisting collaboration patterns and therefore endogenous to the outcomes of interest. If that were the case, and labs were able to obtain a spot next to the labs with which they had an interest in collaborating more, then we should observe a rise in collaboration that predates the relocation. Figure 1 reassures us this is not the case, as the rise in collaboration follows colocation (no pre-trend) and builds progressively over the years in the post-period. In fact, there is no activity until two years or more after the move. The figure plots the estimated coefficient of an OLS regression with year and lab-pair fixed effects for all of the years before and after the move.<sup>27</sup> The dependent variable is the probability of collaboration (dummy), and the error bars represent 95% confidence intervals based on robust standard errors (clustered at the lab-pair level).<sup>28</sup>

Past exposure may allow labs to contact a now-distant lab when the right opportunity emerges or, if opportunities have already been identified, labs can make commitments to keep their research active through joint seminars, temporary colocation, and distant interactions. Therefore, if search costs are a key friction to

**Figure 1.** (Color online) Probability of Collaboration and Colocation



*Notes.* Estimated coefficient for years before and after the move. The dependent variable is collaboration (1/0). Regression includes lab-pair fixed effects and year fixed effects. Error bars represent 95% confidence intervals based on robust standard errors clustered at the lab-pair level.

**Table 6.** Colocation and Separation

Variables	(1) OLS 1/0	(2) OLS #Collabs	(3) Poisson #Collabs
<i>After Colocation</i>	0.0148*** (0.0054)	0.0217** (0.0093)	0.8871*** (0.2814)
<i>After Separation</i>	0.0101 (0.0089)	0.0388 (0.0278)	-0.4520 (0.3171)
Lab-pair fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Observations	295,435	295,435	10,383
R-squared	0.005	0.005	
Number of lab pairs	35,805	35,805	587
Log likelihood			-2,926

*Notes.* The dependent variable in column (1) is a dummy equal to 1 if the lab pair collaborates in the focal year. The dependent variable in columns (2) and (3) is the number of collaborations. *After Colocation* is equal to 1 when a lab pair becomes collocated because of the moves and 0 otherwise. *After Separation* is equal to 1 when a previously collocated pair is separated because of the moves and 0 otherwise. The Poisson specification with fixed effects drops all lab pairs where collaboration is never observed, hence the smaller number of observations. Robust standard errors clustered at the lab-pair level in parentheses.

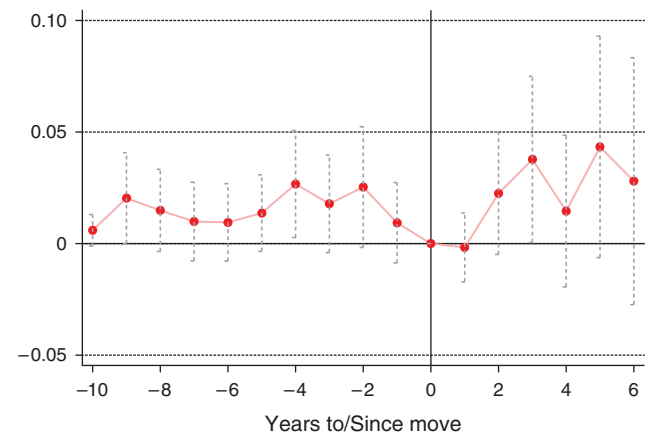
\*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

collaboration, separation should not have a strong, negative effect on the probability of collaboration.<sup>29</sup> If joint execution costs constitute a barrier to collaboration instead, separated pairs should reduce their rate of collaboration.

Consistent with the search costs mechanism, in Table 6 separation has a noisy, nonsignificant effect on the probability and intensity of collaboration. The coefficient is positive and nonsignificant in the OLS regressions (columns (1) and (2)) and negative and nonsignificant in the Poisson specification (column (3)).<sup>30</sup> Meanwhile, controlling for separation leaves the results on collocation unchanged.

Figure 2, which plots the estimated regression coefficients<sup>31</sup> for the years before and after separation, highlights three facts: (1) lab pairs that are going to be separated have a higher propensity to collaborate before the move (all of the coefficients in the pre-period are positive and many are statistically different than zero); (2) the relocation process seems to generate a slight, temporary decay in collaboration (years -1 to +1); (3) lab pairs revert to their mean collaboration propensity in the later periods (although results are noisy, potentially because some pairs recover and others do not).

Overall, results from Tables 5 and 6 are consistent with the idea that search costs may be a more important friction than joint execution costs in defining if two labs will collaborate or not. If the key mechanism through which collocation facilitates collaboration is by helping scientists discover new potential collaborators, then the effect of proximity should be strongest where

**Figure 2.** (Color online) Probability of Collaboration and Separation

*Notes.* Estimated coefficient for years before and after the move. The dependent variable is collaboration (1/0). Regression includes lab-pair fixed effects and year fixed effects. Error bars represent 95% confidence intervals based on robust standard errors clustered at the lab-pair level.

search costs are ex ante high and less pronounced where search costs are lower because alternative channels for knowledge diffusion are likely to exist (e.g., joint conferences, journals, etc.).

To proxy for the fact that two labs may be working in related areas before the moves (i.e., within more proximate communities of science), Table 7 uses two sources of data: keywords listed on publications and the full set of references cited in the papers. To calculate a measure of proximity in knowledge space between two labs, the cosine distance between the vectors of keywords (or of references cited) used by each lab is calculated excluding all direct collaborations (which would exhibit perfect overlap in keywords and references by design). Lab pairs that have an above-the-median similarity in keywords and references are classified as facing lower search costs ex ante (columns (1) and (3)), since they are more likely to overlap in topics and the literature they cite (and potentially conferences, etc.). Lab pairs that have below-the-median similarity (columns (2) and (4)) are classified as experiencing higher search costs ex ante, since they do not seem aware of each other's research topics and body of knowledge.

Interestingly, whereas the two classifications assign slightly different sets of labs to each bin, the interpretation of the main effect is consistent between them. When search costs are low ex ante (columns (1) and (3)), collocation has a positive but nonsignificant effect on the probability of collaboration. This is consistent with these lab pairs already overlapping through other channels and being aware of each other's work (as evidenced by the similar keywords and literature used). As additional evidence that search costs may

**Table 7.** Probability of Collaboration and Search Costs

Variables	(1) OLS 1/0 low search costs (keywords)	(2) OLS 1/0 high search costs (keywords)	(3) OLS 1/0 low search costs (cited ref.)	(4) OLS 1/0 high search costs (cited ref.)
<i>After Colocation</i>	0.0056 (0.0126)	0.0179* (0.0093)	0.0006 (0.0268)	0.0145* (0.0077)
<i>After Separation</i>	-0.0197* (0.0112)	0.0244 (0.0185)	-0.0080 (0.0180)	0.0059 (0.0136)
Lab-pair fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Observations	48,710	60,837	20,698	77,951
R-squared	0.008	0.016	0.008	0.012
Number of lab pairs	3,864	6,102	1,366	6,485

*Notes.* The dependent variable in all columns is a dummy equal to 1 if the lab pair collaborates in the focal year. In columns (1) and (3), only pairs with above the median cosine similarity (low search costs) are included. In columns (2) and (4), only pairs with below the median similarity (high search costs) are included. The cosine similarity measures are based on the vectors of keywords used by each lab in columns (1) and (2), and on the vectors of cited references in columns (3) and (4). In all cases, the measures are calculated before the moves start and do not include direct collaborations between the focal labs (which would share mechanically all keywords and cited references). *After Colocation* is equal to 1 when a lab pair becomes colocated because of the moves and 0 otherwise. *After Separation* is equal to 1 when a previously colocated pair is separated because of the moves and 0 otherwise. Robust standard errors clustered at the lab-pair level in parentheses.

\* $p < 0.1$ .

not be the key constraint for collaboration between these pairs, separation does seem to induce a decay in their likelihood of collaboration, possibly because some of these marginal, within-field collaborations are only sustainable when joint execution costs are low. Moreover, when search costs are high, not only do we observe a strong and positive effect of colocation on collaboration, but separation has only a noisy effect on outcomes, supporting the idea that for these pairs, the key constraint is being aware of each other’s agenda.

A consequence of this result is that over time, labs may become more similar because of lower search costs and repeated exposure, exhausting some of the best arbitrage opportunities in idea space. After separation, because of higher execution costs, some labs that are very similar (columns (1) and (3)) may reduce collaboration with each other in favor of new opportunities with newly colocated labs that bring more novelty to their approaches (columns (2) and (4)).

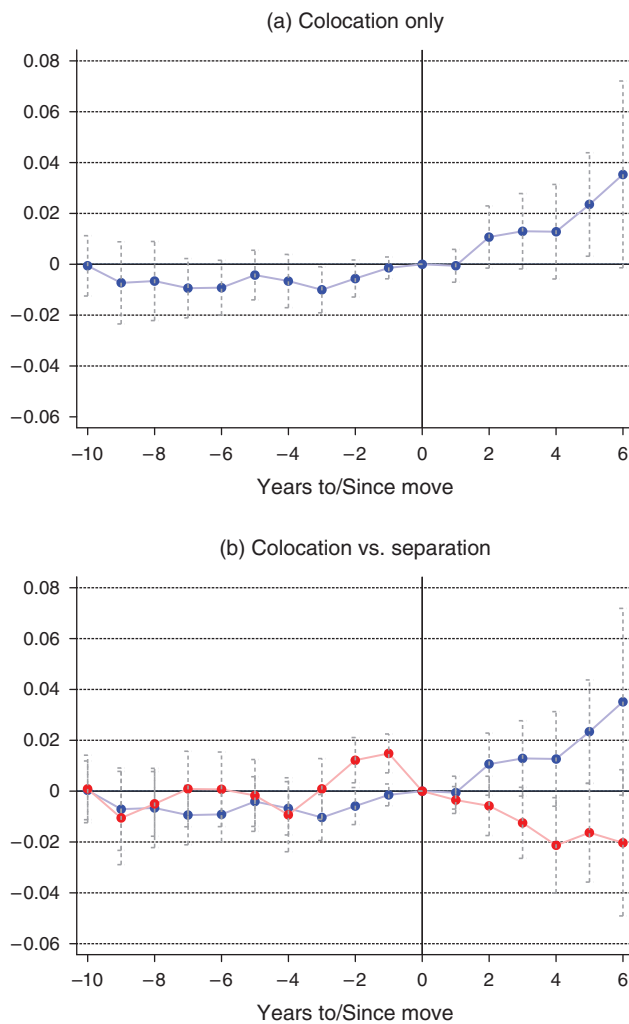
To test if repeated exposure through colocation leads to an increase in similarity and awareness of each other’s research, Figure 3(a) plots the estimated regression coefficients<sup>32</sup> for the years before and after colocation using the cosine similarity in cited references as the dependent variable. The measure captures the extent of overlap in backward cites for a specific lab pair in a given year (excluding joint collaborations). Overlap does not increase until two or more years after colocation, which is consistent with changes taking place slowly over time through knowledge flows and direct collaboration.<sup>33</sup>

Adding separation to the regression of similarity in cited references (Figure 3(b)) shows that the reverse

process takes place when distance increases: after separation, previously colocated labs progressively grow apart, a result that is striking because in the two years before separation, they were actually more likely to cite the same references than the controls.<sup>34</sup> The neighbors of a lab, by exposing it to new knowledge, ideas, and collaboration opportunities, seem to profoundly shape its research trajectory. In this particular setting, whereas collaboration seems to recover from an increase in geographic distance (Figure 2)—possibly also because labs can make purposeful investments in temporary colocation and remote interactions to compensate for it—knowledge flows seem to be negatively affected by separation. Ironically, as separated labs embark on less correlated research trajectories, they may also set the stage for more novel recombination of ideas in the future.

### 4.3. Quality of Collaborations and Execution Costs

This last section explores the effect of proximity on the quality of collaborations. Hypothesis 3 predicts that if geographic proximity has a tangible effect on joint execution costs, then after colocation, we should observe more projects both on the left tail of the quality distribution (a selection effect) as well as on the right tail (an effort effect). Separation instead should generate opposite effects. It is important to highlight that the prediction refers to the observed outcomes conditional on a project being developed, but empirically an additional issue arises from the fact that not all projects are observed: e.g., if scientists are overly optimistic about their chances of publishing their results, or if part of the projects that are initially developed are dropped before submission, then we will never see projects of

**Figure 3.** (Color online) Similarity in Cited References Space

*Notes.* Estimated coefficient for years before and after the move. The dependent variable is the cosine similarity (based on cited references) between the two labs. The higher line in the post-period is for colocation, and the lower line is for separation. Regression includes lab-pair fixed effects and year fixed effects. Error bars represent 95% confidence intervals based on robust standard errors clustered at the lab-pair level.

the lowest quality. As a result, any effect on the left tail of the outcome distribution should be considered a lower estimate of the true effect of colocation (or separation) on inventive outcomes.

Table 8 explores how the citation distribution<sup>35</sup> changes after the moves, conditional on the lab pairs collaborating. Overall, colocation seems to increase the number of collaborations that will end on both tails of the distribution (first and fourth quartiles), and to decrease activity in the third quartile (negative and significant) and potentially in the second quartile (negative but not significant). A comparison between the estimated coefficients for colocation shows that the fourth quartile (highest quartile, column (4)) is statistically different at 1% from the one in the third quartile

(column (3)), and at 10% from the one in the second quartile (column (2)). It is not statistically different from the coefficient in the bottom quartile (column (1)), and second and third quartiles are not statistically different from each other. The noisier results in the first quartile could be due to the fact that collaborations that do not generate a publication are not observed, truncating the left tail. Separation exhibits an almost symmetric pattern, with a positive (although nonsignificant) increase in the third quartile and a significant decay in the fourth.

In Table 9, the right tail of the outcome distribution is further explored by separating lab pairs that faced high versus low search costs before the moves (based on the same approach used in Table 7, columns (1) and (2)). Whereas for pairs with low search costs, both colocation and separation generate only noisy and insignificant results, when search costs are high, colocation leads to high impact research, and separation generates the opposite result. In the online appendix, robustness is shown using a different proxy for quality (citation weighted publications) and by looking at the standard deviation of citations as a way to capture effects on both tails of the outcome distribution (Online Appendix Table A-2).

While only suggestive, the results in Table 9 support the idea that pairs that usually face high search costs (potentially also because space allocation is generally optimized by discipline) are not only more responsive to changes in distance (as we have seen in Table 7), but are also able to produce higher-impact research when given the opportunity to take advantage of lower execution costs.

One interpretation of this result is that these collaborations constitute arbitrage opportunities across domains of science, and that if we were able to observe these pairs for substantially longer periods of time, the positive effect on the right tail may revert to the mean. Another interpretation is that once faced with a more heterogeneous set of local peers, labs were encouraged to embark on higher-risk, higher-reward projects (which would be consistent with the increase in variance). The result is also consistent with the view that multidisciplinary research tends to be of higher variance and, when successful, of higher impact. The effects are also a reminder of how search frictions can be a tangible obstacle to impactful research outside of a community of science. In these cases, colocation (whether temporary or not) can be used to remove some of these frictions and introduce novelty within current research agendas.

## 5. Limitations

This study has a number of limitations. First, only collaborations that end up in a peer-reviewed publication are observed, meaning that the left tail of the outcome

**Table 8.** Colocation, Separation, and the Citation Distribution

Variables	(1) # Collabs 1st quartile (lowest)	(2) # Collabs 2nd quartile	(3) # Collabs 3rd quartile	(4) # Collabs 4th quartile (highest)
<i>After Colocation</i>	0.2135 (0.2241)	-0.0719 (0.2082)	-0.3937* (0.2007)	0.4250** (0.1675)
<i>After Separation</i>	-0.1236 (0.2638)	-0.2239 (0.2190)	0.3725 (0.3538)	-0.3403** (0.1656)
Lab-pair fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Observations	1,221	1,221	1,221	1,221
R-squared	0.061	0.081	0.070	0.199
Number of lab pairs	627	627	627	627

*Notes.* The dependent variable in all columns is the number of publications by the lab pair in the focal year within the given quartile of the citation distribution. Citation data is obtained from Scopus in 2016, and quartiles for the citation distribution are built by year using a large sample of articles in the relevant fields of science. The first quartile represents papers with the lowest level of citations, the fourth the highest. Results are conditional on collaboration in the focal year. *After Colocation* is equal to 1 when a lab pair becomes colocated because of the moves and 0 otherwise. *After Separation* is equal to 1 when a previously colocated pair is separated because of the moves and 0 otherwise. Robust standard errors clustered at the lab-pair level in parentheses.

\* $p < 0.1$ , \*\* $p < 0.05$ .

distribution is not observed. Whereas the analysis on cited references and keywords may be able to capture the outcome of some of the interactions (planned or serendipitous) that do not translate into a paper, the analyses on the rate and quality of research misses them. The measured effects on the left tail of the outcome distribution are therefore likely to represent a lower estimate of the true effects. Furthermore, collaborations across labs are rare relative to collaborations within labs, which limits the set of tests that can be conducted on this sample (e.g., further splits of the sample across more fine-grained dimensions).

**Table 9.** Colocation, Separation and Max Citations

Variables	(1) Max cites (low search costs)	(2) Max cites (high search costs)
<i>After Colocation</i>	27.9531 (31.6185)	40.0822*** (13.3165)
<i>After Separation</i>	6.6615 (11.3936)	-36.4662*** (6.9199)
Lab-pair fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	639	529
R-squared	0.142	0.118
Number of lab pairs	293	294

*Notes.* The dependent variable in all columns is the maximum number of citations received by the publications of the lab pair in the focal year. Citation data is obtained from Scopus in 2016. In column (1), only pairs with above the median cosine similarity in keywords (low search costs) are included. In column (2), only pairs with below the median similarity in keywords (high search costs) are included. Results are conditional on collaboration in focal year. *After Colocation* is equal to 1 when a lab pair becomes colocated because of the moves and 0 otherwise. *After Separation* is equal to 1 when a previously colocated pair is separated because of the moves and 0 otherwise. Robust standard errors clustered at the lab-pair level in parentheses.

\*\*\* $p < 0.01$ .

Second, location information is imprecise,<sup>36</sup> generating noise in the exact timing of a move and potentially biasing the estimates downward. The same applies to the ability to correctly capture and match every publication of the entities involved, as affiliation data is difficult to clean and harmonize at scale at the lab level. Hopefully, as bibliometrics and algorithms for disambiguation improve, better data will become available.

Third, while the relocations on the Jussieu campus are substantially more constrained than typical observational data on campus moves, they do not constitute random assignment. It is reassuring to see that the difference-in-differences estimates, and in particular the pre-trends, do not hint at strong selection into colocation during the study period.

Fourth, the data does not allow perfect separation and direct measurement of some of the mechanisms discussed in the theoretical framework (e.g., face-to-face interactions, serendipitous conversations, etc.).

## 6. Conclusions

The paper provides novel empirical evidence grounded in an original theoretical framework to explain why colocation matters for the rate, direction, and quality of scientific collaboration. To address endogeneity concerns due to selection into colocation and matching, I exploit the constraints imposed on the spatial allocation of labs on the Jussieu campus of Paris by the removal of asbestos from its buildings.

The analyses highlight under which conditions search and joint execution costs are more likely to be responsible for the patterns observed in the data. Consistent with recent experimental (Boudreau et al. 2017) and observational (Kabo et al. 2014) evidence, search frictions can be a substantial obstacle to collocation even within the boundaries of the same

institution: labs that become colocated because of the moves are 3.5 times more likely to collaborate with each other. Supporting the search costs hypothesis, effects are driven by pairs that are more likely to face higher search costs *ex ante*, such as labs that do not work on similar topics and that do not cite the same literature. Furthermore, separating previously colocated labs does not lead to a decay in collaboration, which is consistent with past exposure allowing scientists to compensate for distance through temporary collocation and remote interactions, as well as through inventive and organizational networks (Crescenzi et al. 2016, Monge et al. 1985, Breschi and Lissoni 2004, Singh 2005, Azoulay et al. 2011).

At the same time, over long periods of time, search costs seem to affect the research trajectories of separated scientists: while colocated labs grow increasingly similar in topics and literature cited, separated ones undertake less correlated research trajectories. Ironically, by embarking on different paths, labs may be setting the stage for future, high-impact idea recombinations.

Lower execution costs under collocation, by endogenously allowing scientists to improve ideas more efficiently when proximate, also have an effect on the quality of inventive outcomes. Whereas search costs profoundly influence who a scientist is more likely to collaborate with, joint execution costs appear to shape the distribution of outcomes, conditional on collaboration. Consistent with the theoretical framework, after collocation, more collaborations are observed on both tails of the outcome distribution. The high-impact collaborations that emerge are predominantly coming from lab pairs that faced higher search costs *ex ante*: this points to potential gains from connecting labs within an institution that may not overlap directly through other channels, as these labs may bring novel ideas to domain-specific research agendas. Results are also consistent with past research that has shown that diverse inventive teams are correlated with novel idea recombination (Singh and Fleming 2010); that breakthrough, Nobel Prize contributions are correlated with scientists being embedded in different communities of research at the time of their development (Ham and Weinberg 2016); and that winning solutions to innovation problems tend to come from solvers from a different field of technical expertise (Jeppesen and Lakhani 2010).

The findings also point out some of the strategic trade-offs that the spatial allocation of teams entails. By optimizing space based on current beliefs of where opportunities are, organizations are making a strategic choice that will profoundly shape their R&D trajectory. Space acts as a powerful layer of incentives, which can be used to define not only the intensity of interactions, but also their quality. Organizations that either need to move away from a declining trajectory or want to

explore radically novel opportunities can colocate previously separated teams to encourage serendipitous (and planned) conversations between individuals with different priors, ideas, and knowledge. When collocation is not an option, other forms of temporary collocation could be strategically used to compensate for the lower chance of an interaction and higher search costs. Interestingly, once individuals are aware of each other, proximity plays a lesser role, and collaboration can also be sustained over distance. At the same time, this seems to come at the cost of right-tail outcomes, a result that future research may be able to unpack further, and that advances in communication technology and virtual reality may be able to undo (e.g., by recreating the benefits of in-person, face-to-face interactions and serendipitous conversations).

The moves on the Jussieu campus, by injecting exogenous variation into a process otherwise optimized by scientific fields, also highlight how scientific communities, also because of endogenous space allocation, can become an obstacle to breakthrough research. While it may not be optimal for an institution to relocate scientists to overcome these barriers,<sup>37</sup> forms of temporary collocation (e.g., joint conferences) could be strategically used to encourage cross-pollination across disciplines.

Since Marshall's seminal work (Marshall 1890) on localization economies, scholars have been interested in why collocation matters for the generation of new ideas. While we do know that the spatial allocation of inventors and scientists has an impact on the diffusion of information and ultimately on innovation, we still know surprisingly little about the micro-foundations of knowledge recombination. This study is a first step toward helping us understand the mechanisms at work at different levels of distance.

### Acknowledgments

The author thanks Ajay Agrawal, Ashish Arora, Pierre Azoulay, Alessandro Bonatti, Kevin Boudreau, Jeff Furman, Alberto Galasso, Alfonso Gambardella, Patrick Gaule, Avi Goldfarb, Matt Grennan, Nicola Lacetera, Karim Lakhani, Mara Lederman, Chris Liu, Hong Luo, Liz Lyons, Alex Oettl, Peter Thompson, Carlos Serrano, Tim Simcoe, Olav Sorenson, Paula Stephan, Scott Stern, Jane Wu, the associate editor and referee team, seminar participants at MIT, the University of Toronto, Bocconi University, Boston University, the Fox School of Business, Harvard Business School, IESE Business School, London Business School, the London School of Economics, the Krannert School of Management, the Scheller College of Business, SMU Cox School of Business, Universitat Pompeu Fabra, the Yale School of Management, Wilfrid Laurier University, the Academy of Management, the Association of American Geographers 2012 Annual Meeting, the the 2011 Economics of the Creative Economy conference, and Collegio Carlo Alberto (Turin) for comments. Mia Rozenbaum, Taranjit Singh, and Consuela-Emilia Tataru provided excellent research assistance. All errors remain the author's own.



## Endnotes

<sup>1</sup> This assumes that inventive outcomes are skewed (i.e., most ideas are of very low quality), that ideas need to be of sufficient quality to be published (minimum effort level), and that applying more effort improves the quality of a research project.

<sup>2</sup> Additional, more strict European regulation was published in 2003 (directive 2003/18/EC), 2004 (directive 2004/37/EC), 2007 (2007/30/EC), and 2009 (2009/148/EC).

<sup>3</sup> Five or more years after becoming neighbors, labs are 20% more similar in terms of the references they cite.

<sup>4</sup> This is consistent with past research that has shown that social proximity can compensate for geographic distance (Agrawal et al. 2006, Sorenson et al. 2006).

<sup>5</sup> For example, Chai (2014) finds that participation in the same conference is positively correlated with future collaboration.

<sup>6</sup> Additionally, if the arrival rate of ideas is higher under colocation because of lower search costs (e.g., serendipitous interactions during low-opportunity cost time), this effect would be amplified.

<sup>7</sup> See the Agrawal et al. (2016) model of slack time and innovation for a detailed description. In the paper, the authors assume that idea quality follows an exponential distribution, that the cost function is convex in effort, and that there is a minimum effort requirement for an idea to be developed.

<sup>8</sup> <http://www.upmc.fr/en/university.html> (accessed April 2, 2016).

<sup>9</sup> See <http://www.ipubli.inserm.fr/handle/10608/20> (accessed April 2, 2016).

<sup>10</sup> Additional, more strict European regulation was published in 2003 (directive 2003/18/EC), 2004 (directive 2004/37/EC), and 2007 (2007/30/EC) and 2009 (2009/148/EC).

<sup>11</sup> Some of the labs are placed in temporary sites.

<sup>12</sup> For example, in 1997, 38.9% of pairs that experience a change in colocation are within subfields, and 61.1% across subfields; in 1998 the opposite is true, with 69.7% within and 30.3% across; in 1999, 89.5% is across, 10.5% is within; in 2000 all pairs are across; in 2001, 66.6% are across and 33.4% are within; in 2002 the opposite is true, with 68.8% within and 31.2% across, etc.

<sup>13</sup> <https://www.elsevier.com/solutions/scopus/content> (accessed July 6, 2017).

<sup>14</sup> The initial set of affiliations is first run through a script that divides the strings into their main components (i.e., department name, institute name, lab name, building name, floor, address, zip code, city, and country). Normalized lists are then created for the key labs, institutes, and departments, so that each string can be mapped to a unified name. Eighty-seven percent of affiliation strings are matched to a UPMC lab or entity, resulting in a final set of 36,822 affiliation strings.

<sup>15</sup> Five thousand eight hundred forty-seven unique affiliation instances. Location information is extracted by searching for building numbers, names, and their potential abbreviations, and then manually checked to confirm accuracy.

<sup>16</sup> See <http://www.epaurif.fr/documentation/transferts> (accessed May 1, 2015).

<sup>17</sup> Geodesic distances are calculated using a method developed by Thaddeus Vincenty and implemented in Stata by Nichols (2003).

<sup>18</sup> The choice of this level of analysis is driven by data constraints, as floor-level data is not always available, particularly for early years. Same-floor-level estimates for colocation (within the subsample with floor information) are typically larger; hence, the same tower and corridor building effects probably constitute a lower bound of the true effects of microgeography.

<sup>19</sup> For index keywords, according to Scopus: "A team of professional indexers assigns index terms to records according to the following

controlled vocabularies (in addition to keywords supplied by authors themselves): GEOBASE Subject Index (geology, geography, earth and environmental sciences), EMTREE (life sciences, health), MeSH (life sciences, health), FLX terms, WTA terms (fluid sciences, textile sciences), Regional Index (geology, geography, earth and environmental science), Species Index (biology, life sciences), Ei Thesaurus (controlled and uncontrolled terms) (engineering, technology, physical sciences)." Source: <http://info.sciencedirect.com/scopus/scopus-in-detail/content-coverage-guide/metadata> (accessed May 1, 2015).

<sup>20</sup> The Scikit-Learn python module (<http://scikit-learn.org/stable/modules/metrics.html>) is used to calculate the L2-normalized dot product of the two vectors of keywords  $x$  and  $y$  as  $\text{cosinesimilarity}(x, y) = x \cdot y^T / \|x\| \|y\|$ . See also <http://nlp.stanford.edu/IR-book/html/htmledition/dot-products-1.html> (accessed May 1, 2015).

<sup>21</sup> If the two vectors were identical, then the cosine similarity would be equal to 1.

<sup>22</sup> In an OLS regression that includes lab-pair fixed effects and year fixed effects. Results are robust to excluding the fixed effects.

<sup>23</sup> As a comparison, the expansion of a campus would not offer the same degree of exogenous variation as labs would have more influence over the allocation of the new space, and the expansion itself could be the result of promising research being conducted on campus.

<sup>24</sup> In other words, it does not simply rely on a single shock, but on multiple moves staggered in time.

<sup>25</sup> But can also be the number of collaborations or key statistics linked to the type and quality of papers that emerge (e.g., citations, similarity in keyword or cited references space).

<sup>26</sup> Descriptive statistics at this level of analysis are reported in Online Appendix Table A-1.

<sup>27</sup> The baseline is any year more than 10 years away from the move.

<sup>28</sup> In the appendix, robustness is provided by building the same figure with the number of collaborations as the dependent variable (Online Appendix Figure A-2), and by ignoring the five most recent years of moves in the data set. Whereas moves on the campus continued until 2014 (and a few are still taking place in 2016), the data used in this paper cover 1980–2010. When estimating the effects, Online Appendix Figure A-3 limits the sample to moves between 1997 and 2005.

<sup>29</sup> Empirically, it is important to highlight that separated lab pairs were more likely to be related both in knowledge space and research agendas (since the campus was optimized to minimize distance by field before the moves started)—i.e., these pairs are also more likely to overlap in other circumstances (e.g., teaching, conferences) and have a higher baseline risk of collaboration.

<sup>30</sup> Part of the noise in the result could be due to crowding out taking place for some of these pairs as labs shift part of their collaboration portfolios toward their new neighbors, as described in the lab-level regressions.

<sup>31</sup> OLS regression with lab-pair fixed effects and year fixed effects. Error bars represent 95% confidence intervals based on robust standard errors clustered at the lab-pair level.

<sup>32</sup> OLS regression with lab-pair fixed effects and year fixed effects. Error bars represent 95% confidence intervals based on robust standard errors clustered at the lab-pair level.

<sup>33</sup> The effect is close in timing to the rise in collaboration observed in Figure 1, which suggests that labs become more similar mostly through purposeful interactions. At the same time, it becomes positive and significant substantially earlier, which is consistent with at least some interactions predating collaboration or taking place in the absence of collaboration. In the online appendix, similar graphs show that colocated labs are also more likely to start publishing in a journal that is new to them but not to their collaborator (Online

Appendix Figure A-4) and exhibit increasing overlap in the keywords they use outside of direct collaborations (Online Appendix Figure A-5).

<sup>34</sup>The estimated coefficient for the separation line at  $t - 2$  and  $t - 1$  is positive and significant, which is consistent with the moves separating lab pairs that were actually benefiting from colocation in the pre-period.

<sup>35</sup>Citation data were obtained from Scopus in 2016, and quartiles for the citation distribution are built by year using a large sample of articles in the relevant fields of science.

<sup>36</sup>An analysis on the subsample of data for which floor-level information is available shows effects roughly twice as large as those measured in Table 5 for labs that not only share the same building, but also the same floor.

<sup>37</sup>Results on collaboration portfolios highlight how labs actually became more inward focused during the moves, potentially because of the worsened set of neighboring labs relative to their preferences.

## References

- Adams J, Clemmons R, Black C, Stephan P (2005) Scientific teams and institutional collaborations: Evidence from U.S. universities. *Res. Policy* 34:259–285.
- Agrawal A, Cockburn I, McHale J (2006) Gone but not forgotten: Labor flows, knowledge spillovers and enduring social capital. *J. Econom. Geography* 6(5):571–591.
- Agrawal A, Catalini C, Goldfarb A, Luo H (2016) Slack time and innovation. Rotman School of Management Working Paper 2599004, University of Toronto, Toronto.
- Allen TJ (1977) *Managing the Flow of Technology* (MIT Press, Cambridge, MA).
- Azoulay P, Zivin JSG, Sampat BN (2011) The diffusion of scientific knowledge across time and space: Evidence from professional transitions for the superstars of medicine. Technical report, National Bureau of Economic Research.
- Belenzon S, Schankerman M (2013) Spreading the word: Geography, policy, and knowledge spillovers. *Rev. Econom. Statist.* 95(3): 884–903.
- Boudreau KJ, Brady T, Ganguli I, Gaule P, Guinan E, Hollenberg A, Lakhani KR (2017) A field experiment on search costs and the formation of scientific collaborations. *Rev. Econom. Statist.* 99(4):565–576.
- Breschi S, Lissoni F (2004) Knowledge networks from patent data: Methodological issues and research targets. Moed HF, Glänzel W, Schmoch U, eds. *Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S&T Systems* (Springer, Berlin), 613–643.
- Chai S (2014) Temporary colocation and collaborative discovery. Working paper, Harvard University, Cambridge, MA.
- Cowgill B, Wolfers J, Zitzewitz E (2009) Using prediction markets to track information flows: Evidence from Google. Das S, Ostrovsky M, Pennock D, Szymanski B, eds. *Auctions, Market Mechanisms and Their Applications*, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Vol. 14 (Springer, Berlin).
- Crescenzi R, Filippetti A, Iammarino S (2017) Academic inventors: Collaboration and proximity with industry. *J. Tech. Transfer* 42(4):730–762.
- Crescenzi R, Nathan M, Rodríguez-Pose A (2016) Do inventors talk to strangers? Proximity and collaborative knowledge creation. *Res. Policy* 45(1):177–194.
- Dizikes P (2011) The office next door. *MIT Tech. Rev.* (October 25), <https://www.technologyreview.com/s/425881/the-office-next-door>.
- Fayard A-L, Weeks J (2007) Photocopiers and water-coolers: The affordances of informal interaction. *Organ. Stud.* 28(5):605–634.
- Fleming L, Sorenson O (2004) Science as a map in technological search. *Strategic Management J.* 25(8–9):909–928.
- Fleming L, King I Charles, Juda AI (2007) Small worlds and regional innovation. *Organ. Sci.* 18(6):938–954.
- Freeman R, Ganguli I, Murciano-Goroff R (2015) Why and wherefore of increased scientific collaboration. Jaffe AB, Jones BF, eds. *The Changing Frontier: Rethinking Science and Innovation Policy* (University of Chicago Press, Chicago), 17–48.
- Ham JC, Weinberg BA (2016) Novelty, knowledge spillovers and innovation: Evidence from Nobel laureates. ZBW GLO Discussion Paper 30, Global Labor Organization.
- Jaffe A, Trajtenberg M, Henderson R (1993) Geographic localization of knowledge spillovers as evidenced by patent citations. *Quart. J. Econom.* 108(3):577–598.
- Jeppesen LB, Lakhani KR (2010) Marginality and problem-solving effectiveness in broadcast search. *Organ. Sci.* 21(5):1016–1033.
- Jones BF, Uzzi B, Wuchty S (2008) Multi-university research teams: Shifting impact, geography and social stratification in science. *Science* 322(5905):1259–1262.
- Kabo F, Hwang Y, Levenstein M, Owen-Smith J (2015) Shared paths to the lab a sociospatial network analysis of collaboration. *Environ. Behav.* 47(1):57–84.
- Kabo FW, Cotton-Nessler N, Hwang Y, Levenstein MC, Owen-Smith J (2014) Proximity effects on the dynamics and outcomes of scientific collaborations. *Res. Policy* 43(9):1469–1485.
- Katz J (1994) Geographical proximity and scientific collaboration. *Scientometrics* 31(1):31–43.
- Liu C (2013) Brokerage by design: Formal structure, geography, and crosscutting ties. Working paper, Rotman School of Management, University of Toronto, Ontario.
- Lucas REJ (1988) On the mechanics of economic development. *J. Monetary Econom.* 22(1):3–42.
- Mairesse J, Turner L (2005) Measurement and explanation of the intensity of co-publication in science: An analysis at the laboratory level. NBER Working Paper 11172, National Bureau of Economic Research, Cambridge, MA.
- Marshall A (1890) *Principles of Economics* (Macmillan, London).
- Monge PR, Rothman LW, Eisenberg EM, Miller KI, Kirste KK (1985) The dynamics of organizational proximity. *Management Sci.* 31(9):1129–1141.
- Nichols A (2003) VINCENTY: Stata module to calculate distances on the Earth's surface. Statistical Software Components, Boston College.
- Nielsen C (2016) How Microsoft used an office move to boost collaboration. *Harvard Bus. Rev.* (October 11), <https://hbr.org/2016/10/how-microsoft-used-an-office-move-to-boost-collaboration>.
- Olson GM, Olson JS (2000) Distance matters. *Human-Computer Interaction* 15(2):139–178.
- Romer PM (1990) Endogenous technological change. *J. Political Econom.* 98(5):71–102.
- Singh J (2005) Collaborative networks as determinants of knowledge diffusion patterns. *Management Sci.* 51(5):756–770.
- Singh J, Fleming L (2010) Lone inventors as sources of breakthroughs: Myth or reality? *Management Sci.* 56(1):41–56.
- Sorenson O, Rivkin JW, Fleming L (2006) Complexity, networks and knowledge flow. *Res. Policy* 35(7):994–1017.
- Stuart T, Liu C (2010) Boundary spanning in a for-profit research lab: An exploration of the interface between commerce and academe. Working paper, Harvard University, Cambridge, MA.
- Thompson P (2006) Patent citations and the geography of knowledge spillovers: Evidence from inventor- and examiner-added citations. *Rev. Econom. Statist.* 88(2):383–388.
- Thompson P, Fox-Kean M (2005) Patent citations and the geography of knowledge spillovers: A reassessment. *Amer. Econom. Rev.* 95(1):450–460.
- Van den Bulte C, Moenaert RK (1998) The effects of R&D team collocation on communication patterns among R&D, marketing, and manufacturing. *Management Sci.* 44(11, part 2):S1–S18.
- Weitzman ML (1998) Recombinant growth. *Quart. J. Econom.* 113(2):331–360.
- Wuchty S, Jones B, Uzzi B (2007) The increasing dominance of teams in production of knowledge. *Science* 316(5827):1036–1039.