University of Bath

Alternative formats
If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

Management Science

How Do Travel Costs Shape Collaboration?

Christian Catalini, Christian Fons-Rosen, Patrick Gaulé

How Do Travel Costs Shape Collaboration?

Christian Catalini,[a,b,c] Christian Fons-Rosen,[d,e] Patrick Gaulé[f,g]

[a] MIT Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142; [b] National Bureau of Economic Research, Cambridge, Massachusetts 02138; [c] Max Planck Institute for Innovation and Competition, 80539 München, Germany; [d] University of California, Merced, Merced, California 95343; [e] Center for Economic and Policy Research, Washington, District of Columbia 20009; [f] University of Bath, Bath BA2 7AY, United Kingdom; [g] IZA Institute of Labor Economics, 53113 Bonn, Germany
Contact: catalini@mit.edu, https://orcid.org/0000-0003-1312-6705 (CC); cfons-rosen@ucmerced.edu, https://orcid.org/0000-0003-2620-7205 (CF-R); p.gaule@bath.ac.uk (PG)

Abstract. We develop a simple theoretical framework for thinking about how geographic frictions, and in particular travel costs, shape scientists' collaboration decisions and the types of projects that are developed locally versus over distance. We then take advantage of a quasi-experiment—the introduction of new routes by a low-cost airline—to test the predictions of the theory. Results show that travel costs constitute an important friction to collaboration: after a low-cost airline enters, the number of collaborations increases between 0.3 and 1.1 times, a result that is robust to multiple falsification tests and causal in nature. The reduction in geographic frictions is particularly beneficial for high-quality scientists that are otherwise embedded in worse local environments. Consistent with the theory, lower travel costs also endogenously change the types of projects scientists engage in at different levels of distance. After the shock, we observe an increase in higher-quality and novel projects, as well as projects that take advantage of complementary knowledge and skills between subfields, and that rely on specialized equipment. We test the generalizability of our findings from chemistry to a broader data set of scientific publications and to a different field where specialized equipment is less likely to be relevant, mathematics. Last, we discuss implications for the formation of collaborative research and development teams over distance.

1. Introduction

The drastic reduction in communication costs brought by the diffusion of the internet initially led to claims about a future in which technology could overcome geographic frictions and facilitate the rapid exchange of ideas, goods, and services independent of distance (Cairncross 1997, Friedman 2005). From an empirical standpoint, this "death of distance" hypothesis has found limited support, as most evidence points to agglomeration mattering more, not less, than before across a variety of settings (Leamer and Levinsohn 1995, Forman et al. 2005, Blum and Goldfarb 2006, Agrawal et al. 2015). Instead of substituting for colocation, digital interactions often complement it (Agrawal and Goldfarb 2008),[1] resulting in nonobvious changes in

how teams and organizations structure collaborations and develop new ideas when communication costs are low, but teamwork and research and development (R&D) require specialized expertise and resources that are geographically dispersed (Adams et al. 2005, Wuchty et al. 2007, Jones et al. 2008).[2]

Moreover, not all types of interactions have benefited in the same way from improvements in communication technology. Colocation plays a disproportionate role in the serendipitous discovery of new collaborators and ideas (Catalini 2018), and in the absence of offline opportunities for interaction, search frictions can prevent individuals from finding ideal collaborators even within the boundaries of the same institution (Boudreau et al. 2017). Similarly, exchanges that require the

1

transfer of complex information and tacit knowledge (Polanyi 1958, Von Hippel 1994) still heavily rely on face-to-face interactions (Gaspar and Glaeser 1998, Rosenthal and Strange 2001, Storper and Venables 2004). As a result, firms, communities of experts and teams invest substantial amounts of time, effort and resources to ensure that the right individuals can be colocated—even if only temporarily—to discuss ideas, make progress on projects, and develop the relationships that can later support more effective interactions over distance. Such temporary forms of colocation have been shown to foster both idea diffusion and the formation of new collaborations (Chai and Freeman 2018).[3]

If face-to-face interactions are instrumental in finding and evaluating new collaborators, establishing trust, and advancing joint work, then as communication costs drop, if they are a complement and not a substitute to remote interactions, they should become more valuable. Furthermore, their absence would likely constitute the key remaining friction in the formation and operation of geographically distributed teams. Ironically, by making online communication extremely efficient, the internet may have enhanced the role that travel technology plays in the economy.

The objective of this paper is to develop and test a simple theoretical framework for thinking about how geographic frictions, and in particular travel costs, shape collaboration decisions and the types of projects that are developed locally versus over distance. The model highlights a key trade-off individuals face when deciding if they should work with a local versus a distant collaborator: whereas the global pool of potential collaborators is often deeper and may therefore offer an ideal match, collaboration over distance incurs additional communication and travel costs. We build on this basic tension in a context where individuals endogenously allocate effort to projects based on their potential, and where a project's variance in outcomes or the need for complementary expertise, equipment, or resources can influence with whom a project is pursued. The simple framework captures an increasingly relevant challenge: to be able to solve problems of rising complexity, teams of specialized experts have to be put together (Jones 2009), but this often involves collaboration over distance.

We take advantage of a quasi-experiment—the introduction of new routes by a major low-cost airline—to test the predictions of the theory within the context of collaborations between scientific labs. The setting allows us to observe the full set of scientists at risk for collaboration in any given year as well as important characteristics about them such as their age, career stage, past productivity, area of specialization, and departmental funding.

The cheaper fares brought by the expansion of the low-cost airline (Southwest Airlines)[4] are part of a broader, 50% reduction in the cost of air travel that took place in the United States over the last 30 years (Perry 2014).[5] Furthermore, they provide a source of plausibly exogenous variation in the cost of conducting research between scientists at the affected airports.

Using a difference-in-differences empirical strategy, we are able to recover a causal estimate of the effect of a reduction in travel costs not only on the rate of collaboration but also, more importantly, on the types of projects scientists pursue. Results show that travel costs are an important friction to collaboration: after Southwest entry, the number of collaborations increases between 0.3 and 1.1 times, a result that is robust to multiple falsification tests and causal in nature. The reduction in geographic frictions is particularly beneficial for high-quality scientists that are otherwise embedded in worse local environments, although women scientists do not seem to benefit. Consistent with the theory, lower travel costs also endogenously change the types of projects scientists engage in locally versus over distance. After the shock, we observe an increase in higher quality and more novel projects, as well as projects that take advantage of complementary knowledge and skills between subfields or that rely on specialized equipment. We test the generalizability of our findings within chemistry to a broader data set of scientific publications and to mathematics, a field where specialized equipment is less likely to be relevant. Last, we discuss implications for the formation of collaborative R&D teams in the presence of geographic frictions.

The rest of the paper is organized as follows: in Section 2 we provide additional institutional details about scientific collaboration, on how chemistry differs from others fields, and the data we use. Section 3 introduces our empirical strategy and main results, together with a series of robustness tests and extensions targeted at assessing the generalizability of our findings to different samples. Section 4 develops a model to guide the interpretation of the findings, as well as the exploration of more nuanced hypotheses about the type of projects pursued in response to a reduction in geographic frictions. Section 5 tests these additional predictions, and Section 6 concludes.

The reason why we first explore the difference-in-differences results on the rate of collaboration (in Section 3) and then present the model and test hypotheses about the type of collaborations that emerge (in Sections 4 and 5) is because we follow the natural evolution of the project. We started from an empirical assessment of the presence of an effect on the rate of collaboration, moved to theory development to form predictions about the types of projects affected and to identify additional data to be collected (e.g., on

novelty, equipment intensity), and then finally brought these new predictions to the data.

2. Scientific Collaboration

Scientific research is an increasingly collaborative endeavor, as reflected in the growing number of authors per paper over time (Wuchty et al. 2007). Collaborations are typically formed to combine skills and knowledge (Freeman et al. 2014), to access complementary generalist or specialist talent and resources (Sauermann and Haeussler 2017, Teodoridis 2018), and to expand the knowledge frontier when information is tacit and difficult to transfer or recombine without extensive, direct interactions (Stephan 2012). Increasing complexity has also been linked to a rising need for interdisciplinary teams (Falk-Krzesinski et al. 2011, Milojević 2014), with Wu et al. (2019) showing that both small and large teams play an important but different role in pushing the knowledge frontier. Using a large-scale data set of papers, patents, and software products developed over 60 years, the authors show that whereas smaller teams are associated with more disruptive work and exploration, larger ones are systematically linked to advancing existing ideas and execution.

In terms of team formation, empirical evidence shows that researchers typically source collaborations through their professional networks (Freeman et al. 2014), through serendipitous interactions with colocated individuals (Catalini 2018) and conferences (Boudreau et al. 2017, Campos et al. 2018, Chai and Freeman 2018), and by relying on the information disclosed in scientific publications. As Walsh and Lee (2015) highlight, science is organized around increasingly complex teams that resemble the operations of small R&D-intensive firms, with knowledge as their core output.

Substantial differences, however, exist across scientific disciplines in how collaboration and research is organized: for example, whereas mathematicians and theoretical physicists rarely work in labs, most research in chemistry, life sciences, and experimental physics—also because of different capital, talent, and infrastructure requirements—takes place in labs (Stephan 2012).

Our core analysis is focused on collaborations within chemistry. Whereas chemistry largely remains a laboratory-based science, it has also not embraced the larger-scale, big science projects observed in physics. Team size in chemistry—as measured by the number of coauthors—is lower than in biology and physics though higher than in mathematics (Adams et al. 2005). Chemistry labs are run by a faculty member (principal investigator) who obtains funding for the laboratory, directs research projects, appears as a coauthor on all publications, oversees resource allocation, and effectively decides whether to collaborate with other labs. In our sample, the median number of

coauthors per paper is four, and many of the authors are graduate students, postdocs, or technicians. They perform most of the experiments and day-to-day work on a project.

Although many research projects involve a single principal investigator, collaborations between labs and principal investigators are common as well. Consistent with the findings from large-scale surveys of scientists (Freeman et al. 2014), in our conversations with U.S. principal investigators in chemistry, complementary expertise, skills, materials, or new types of experiments are all mentioned as reasons for collaboration across labs. As in other fields of science, collaborations in chemistry are sourced through the principal investigators and junior members' professional networks, serendipitous interactions at conferences, email, etc. In the paper, we focus on collaborations between principal investigators, which are essentially collaborations between different labs.

2.1. Data Sources and Key Outcomes of Interest

To examine the effect of the changes in travel costs induced by the entry of Southwest Airlines on scientific collaboration, we combine data on scientists with publication records and air transportation information. Within the chemistry and mathematics samples, biographical information on scientists enables us to effectively disambiguate publication data while also allowing us to separate faculty members from other types of authors. We now discuss in more detail the data sources we use and key outcomes we focus on throughout the paper.

Air Transportation Data. To recover information on when Southwest operated flights between different routes, as well as information on prices, passengers, and miles flown, we use data from the Airline Origin and Destination Survey (DB1B) of the U.S. Bureau of Transportation Statistics. The DB1B is a 10% random sample of airline tickets from reporting carries in each quarter. For each itinerary, the DB1B records all connecting airports (including origin and destination), the itinerary fare, and other information. These data are publicly available only from 1993; hence we will focus on Southwest entry decisions that occur after 1993.

Match Between Airports and Universities. We compute distances between airports and universities using Google Maps. The matching between universities and airports is complicated by the fact that the same metropolitan area could be served by multiple airports (e.g., O'Hare and Midway in Chicago) or that a college town could be halfway between two airports. We chose to match universities to all airports within a 50-mile radius. We code the year of Southwest

entry for a pair of universities as the first year in which Southwest operates a flight on any route whose end-points (airports) are within 50 miles of the respective universities. Results are robust to narrowing this definition further (e.g., 25 miles, 10 miles); see Online Appendix Table A-1.

Data on Scientists. Our focus is on collaborations between faculty members (and therefore effectively across labs) in the discipline of chemistry,[6] in part because of data availability and in part because of the short publications cycles in this discipline. For biographical information on scientists, our data source is the directory of graduate research published by the American Chemical Society (ACS). Intended as a source of information for prospective graduate students, this directory provides comprehensive listings of faculty affiliated with U.S. departments granting PhDs in chemistry, chemical engineering, and biochemistry. Besides faculty names and departmental affiliations, the directory provides information on year of birth, gender, and education. The directory is published biannually in print and since 1999 on the web.[7] We combine the directories from 1991 to 2013 to build a longitudinal panel of more than 20,000 scientists. We complement this information with department-level R&D expenditures from the National Science Foundation (NSF) Survey of Research and Development Expenditures at Universities and Colleges.

Publication Data. We match faculty names to publication data from Scopus covering more than 200 chemistry journals (including all journals from the American Chemical Society), multidisciplinary journals, and major journals in neighboring disciplines.[8] Within chemistry, the match between publications and scientists is facilitated by the fact that we know institutional affiliations from the American Chemical Society faculty data. We match publications to faculty based on last name, first name and (if nonmissing) middle initials, department, and university affiliation. From publication data, we construct for each scientist time-varying measures of past productivity (with a moving average over the last three years of publication counts weighted by journal impact factor). We also infer our main outcome, copublications, from bibliometric data combined with faculty data.

A key strength of our data is that we know when individuals enter and exit the profession and therefore are at risk for collaborating with others. If we were inferring copublications from publication data only, we could hardly distinguish between active scholars and individuals who have retired or are not doing research in the field. Papers are counted as a copublication between all pairs of faculty members involved.[9]

Additional Key Outcomes. For part of the analysis, we weight copublications by the citations they receive as a proxy for their impact and quality. Citation counts originate from Scopus, are at the article level, and are counted from the year of publication until 2013. We also construct two distinct groups of measures related to novelty using author keywords. These are based on the entire corpus of articles within chemistry journals and related fields. The first group of measures is based on established approaches from the innovation literature (Boudreau et al. 2016, Criscuolo et al. 2017), and it relies on calculating the share of keywords in any given paper that have not been observed before. This allows us to capture both novel uses that gain traction and those that do not. To check the robustness of our results, we also experimented with different definitions of what constitutes a novel use (e.g., bottom 5%, 10%, or 25% of the keyword use distribution), as well as with different aggregation methods (mean share of novelty, max share of novelty, total novelty for the focal papers), finding consistent results. We then replicate this approach for subfields to see if a specific use might have been considered novel in aggregate but not within a smaller community of science.[10]

The second dimension of novelty we explore is what we label as "novel trends." With this measure we are not focused on making sure we capture both failed and successful attempts at developing new concepts (i.e., the variance in outcomes); we instead prioritize identifying emerging new trends in science. A drawback of this measure is that it selects on successful cases where scientists work on concepts that end up gaining broader adoption afterward. The reason why we find this measure interesting is because it proxies for the focal researchers working on topics that were about to become "hot." To do this, for each keyword, we calculate the share of papers in a given year that contains the keyword—a proxy for how popular it is at any point in time. We then calculate the first and second derivatives of this measure relative to the previous year. If both the first and second derivatives are positive, then the keyword is classified as part of a novel trend because its use is quickly accelerating. Additionally, if the first derivative is zero and the second is positive, then we are at a local minimum right before a keyword takes off, which we also consider as a novel trend. Aggregating up at the paper level, a publication is considered part of a novel trend if it has an above-the-median number of novel-trend keywords (results are similar if we impose a higher threshold).

We also constructed proxies for the equipment intensity of publications by first collecting a large-scale list of keywords associated with chemistry equipment[11] and then checking this list against the

keywords used in each paper. Papers with an above-the-median number of equipment-related keywords are classified as equipment intensive (similar results are obtained when using the count of equipment keywords). To test the robustness of our approach to a completely different definition, we also classified areas of chemistry as equipment intensive or not using the NSF Survey of Federal Funds for Research and Development. In particular, we first use NSF data to compute the share of departmental R&D expenditures devoted to capital. We then calculate the specialization of each department across fields of chemistry to assess which areas they are specialized in. Last, we use these measures to run a regression at the department level, linking department-level capital intensity to the relative prevalence of different subfields of chemistry, and we rely on the estimates to classify collaborations based on the type of departments from which they originated.[12]

2.2. Descriptive Statistics for the Main Sample

Our data set covers more than 20,000 scientists and their collaborations. However, we focus on a specific subset of pairs of scientists who experience Southwest entry and for whom we have variation in collaboration over time. Because all regressions include scientist-pair fixed effects, pairs that never collaborate drop out of the sample. In the online appendix, we show that our main result is robust to replacing scientist-pair fixed effects with city-pair fixed effects and including a random sample of noncollaborating pairs. Our results are also robust to replacing pair fixed effects with individual researcher fixed effects.

We have 15,244 pairs of scientists who collaborate at least once.[13] Excluding coauthors who are in the same department, we have 8,311 pairs of scientists in our sample. Only a minority (1,158) of these pairs experience Southwest entry during our analysis period of 1993–2012, either because for the other 7,153 pairs Southwest is already operating a flight or because Southwest never flies between the relevant endpoints. We drop pairs in locations where Southwest enters but then leaves within two years, as well as pairs where Southwest entry coincides with the move of a scientist.[14] Finally, we also exclude pairs that are within less than 200 miles of each other, as air travel is unlikely to be their main travel option.[15] Our final analysis covers 758 pairs of scientists corresponding to 845 individuals.

Table 1 displays descriptive statistics for our chemistry sample at different levels of analysis: individual, individual-pair, and individual-pair-year.[16] Most individuals in the sample are male (90%) with an average age at the time of Southwest entry of 49.6 years. We do not observe individual research budgets, but as a

proxy, we use departmental R&D expenses divided by the number of faculty members in the department. The average in our sample is $279,880 at the time of Southwest entry. According to the NSF survey, R&D expenses include compensation for R&D personnel, equipment, and indirect costs. In terms of specialization,[17] the largest area is physical chemistry (32%), followed by biochemistry (22%), inorganic chemistry (14%), organic chemistry (13%), and material science (11%).

We observe the 758 pairs for 17 years on average,[18] corresponding to 13,147 observations at the individual-pair-year level. Southwest entry events map to 413 distinct new routes. The median pair experiences Southwest entry in 1999, but we observe Southwest entry from 1994 to 2011. The mean number of copublications over the whole period is 1.9, but the majority of pairs copublishes once. Only 9% of pairs collaborates both before and after Southwest entry.

It is useful to compare our analysis sample to other distant pairs that do not experience Southwest entry.

Table 1. Summary Statistics (Main Sample)

Variable	Mean	Std. dev.
Individual scientist level ($n = 845$)		
Age	49.6	11.0
Female	0.10	0.30
Average R&D budget in dept. (USD 1,000s)	279.88	226.75
Speciality		
Physical chemistry	0.32	0.47
Biochemistry	0.22	0.41
Inorganic chemistry	0.14	0.34
Organic chemistry	0.13	0.34
Material science	0.11	0.31
Other	0.08	0.27
Individual-pair level ($n = 758$)		
Year of Southwest entry	2001	4.5
Distance (in miles)	1,232	808.6
Years in sample	17.3	4.6
Total copublications	1.9	3.4
Copub. both before and after	0.09	0.28
Copub. before Southwest entry	0.49	0.50
Copub. after Southwest entry	0.60	0.49
Individual-pair-year level ($n = 13,147$)		
Copublications	0.11	0.41
Local copubs	1.83	2.69
Local copubs with less productive colleagues	0.62	1.42
Different type of chemistry	0.46	0.50
One above average	0.67	0.47
Both above average	0.26	0.44
Individual-pair-year level conditional on copublication ($n = 1,177$)		
Cites	44.95	71.83
Equipment Intensive 1	0.35	0.58
Equipment Intensive 2	0.48	0.87
Novel trends	0.12	0.36
Mean share novel	0.10	0.24
Max share novel	0.12	0.26
Total share novel	0.13	0.31
Novel in field	0.30	0.40

Table 2. Comparing Pairs in the Analysis Sample to Pairs Not Experiencing Southwest (SW)

	Distant pairs not experiencing SW	Analysis sample	*p*-value for equality of means
Total copublications	1.78	1.90	0.19
Number of years observed	13.85	17.51	<0.01
Age (average in pair)	49.16	51.23	<0.01
Different type of chemistry	0.46	0.45	0.79
Average R&D budget in dept.	288.9	278.1	0.17
Observations	5,954	758	

We have approximately 6,000 such pairs. These include pairs where Southwest is already present in the relevant market prior to 1993 when our sample starts or has not entered by 2012 when it ends. They also include cases where one of the pair members is a new faculty hired after Southwest has already entered. The comparison is shown in Table 2. The pairs that experience Southwest are not statistically different from the others in terms of publications but are slightly older (51 versus 49 years) and are observed, on average, for a slightly longer period of time (17 versus 14 years).[19] It is important to note that there is no significant difference in terms of R&D budgets or propensity to be in different subfields of chemistry.

3. Empirical Strategy and Main Results
Our empirical specification is a straightforward difference-in-differences framework at the scientist-pair level where we exploit variation in Southwest entry across different airport pairs over time. It includes scientist-pair fixed effects and is estimated using a Poisson model:

$$Y_{ijt} = \beta AfterSW_{ijt} + \mu_t + \gamma_{ij} + \epsilon_{ijt},$$

where Y_{ijt} is the number of copublications between scientist i and scientist j in year t; $AfterSW_{ijt}$ is an indicator variable that takes the value 1 after Southwest entry; μ_t is a year fixed effect; γ_{ij} is a pair fixed effect to control for unobservable, time-invariant differences between pairs of scientists; and ϵ_{ijt} is an idiosyncratic error term.

Our analysis examines the change in the rate of collaboration and in the types of papers that emerge over time for pairs that coauthor at least once. Because our unit of analysis is the scientist-pair-year, and we include pair fixed effects, our main source of variation is the change in Southwest status for treated pairs, where control pairs are constituted by pairs that never experience entry or will experience it in the future. The pair fixed effects completely capture pairs of scientists for which we never see activity, and thus we remove these from the analysis without empirical consequences. Robust standard errors are clustered at the pair level.

3.1. Southwest Entry and Changes in Passengers, Prices, Miles, and Transfers
Before our main analysis, we check how the arrival of Southwest affects some of the key passenger and fare metrics of interest in the air travel industry. In this exercise, we run regressions at the airport-pair level, and we compare a number of outcomes before and after Southwest entry. Regressions include airport-pair fixed effects and year fixed effects. The coefficients in Table 3 reflect the types of changes one would expect to take place after the arrival of a low-cost competitor: the increase in the number of passengers is between 54% and 57%,[20] and prices drop by 17%–19%. We do not find any effect on the average miles flown[21] or on direct flights, and the reduction in the number of transfers is extremely small. Overall, results are consistent with Southwest lowering the cost of air travel without drastically changing the types of

Table 3. Effects of Southwest Entry on Price, Passengers, and Routes

	(1)	(2)	(3)	(4)	(5)
	Passengers (log)	*Mean price* (log)	*Average miles flown* (log)	*Direct flight*	*Number of transfers*
Southwest entry	0.4437***	−0.1910***	0.0007	0.0002	−0.0174***
	(0.0050)	(0.0024)	(0.0006)	(0.0004)	(0.0017)
Airport-pair fixed effects	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Mean of dep. variable	4.238	5.454	7.066	0.007	1.239
Number of pairs	55,750	55,750	55,739	55,750	55,750
Number of observations	956,029	956,029	955,983	956,029	956,029

Notes. Robust standard errors are in parentheses. *Southwest entry* is an indicator variable that takes a value of 1 if Southwest has started operating a flight between airports. All specifications include airport-pair fixed effects and year fixed effects. Estimation by ordinary least squares.
   ***$p < 0.01$.

routes available or the number of miles passengers have to fly to connect between two endpoints.

3.2. Changes in Collaboration and Evidence for a Causal Interpretation

After a reduction in travel costs, the relative attractiveness of the global pool of potential coauthors should increase, because working with distant collaborators becomes more cost effective. This should lead to an increase in collaboration between the affected locations. As can be seen in column (1) of Table 4 (which uses our main econometric specification), after Southwest enters, we observe a large and significant increase in collaboration between scientists at the connected endpoints.[22] Relying on a 95% confidence interval, we estimate that scientific collaboration increases between 0.3 and 1.1 times.[23] Although the magnitude of the effect is large, it is off a small base (the mean of the dependent variable is approximately 0.1) and comparable with previous studies on the impact of communications, search costs, and colocation on scientific collaboration: Agrawal and Goldfarb (2008) find that Bitnet increased the likelihood of collaboration between pairs of universities by 40%, Boudreau et al. (2017) find that a 90-minute structured information sharing session led to a 75% higher probability of coapplying for a grant, and Catalini (2018) estimates that exogenous colocation increased the chance of a collaboration between labs on the Jussieu campus of Paris by 3.5 times.

One may worry that Southwest entry is systematically correlated with time-varying factors such as growth of the universities (or the regional economies) at both ends of the routes, and therefore that collaboration would have increased even in the absence of a reduction in travel costs. Although our main specification already controls for aggregate time trends through year fixed effects, the validity of our results could be threatened by systematic, time-varying factors that affect the target locations around the time of Southwest entry. In column (2), we mitigate these concerns by controlling for two possible time-varying confounders: the age of the scientist pair and the (log of) departmental R&D budget per faculty member. The first one accounts for changes in the incentives to collaborate as scientists progress in their careers; the second, for changes in the local economies. Whereas the coefficients for the controls are positive and significant, our main result is unaffected. In column (3), we additionally control for the number of years that have passed since both scientists obtained their PhD, a proxy for their ability to both decide with whom they want to collaborate. This estimated coefficient is negative and significant but, again, does not affect the estimate for *Southwest entry*. In column (4), we study the dynamic effects of the reduction in travel costs by replacing the treatment

indicator for *Southwest entry* from column (1) with a set of four dummy variables capturing the years around the treatment. For example, the indicator *Southwest entry* (−1) is equal to 1 if the focal scientist-pair observation is recorded one year prior to the treatment. The other indicator variables are defined analogously with respect to the year of treatment (0), the first year after treatment (1), and two or more years after treatment (2+).[24] The coefficient for *Southwest entry* (−1), which would capture any "effect" of the new airline routes before their introduction, is insignificant, suggesting that there is no collaboration pretrend in the data; that is, it is only once travel costs are reduced that the coefficients turn positive and statistically significant.

A graphical version of a similar exercise with a full set of coefficient estimates for the five years before and five years after Southwest entry is displayed in Figure 1. There is again no collaboration pretrend before Southwest launches a route, and it is only after the new route is available that the estimated coefficients are positive and steadily increasing in magnitude.[25] It is useful to highlight that publication lags in chemistry are substantially shorter than in the social sciences: when studying the 10 major analytical chemistry journals (1985–1999), Dióspatonyi et al. (2001) find median lags between submission and publication of 3–10 months, with some journals publishing papers within 2 months of first submission.

Figure 1. (Color online) Dynamics of the Effect of Southwest Entry: Individual-Pair Level
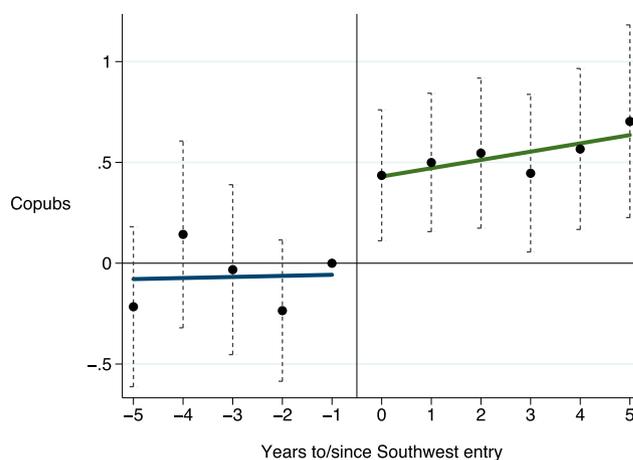
Notes. To generate this graph, we regress individual copublications on year fixed effects, pair effects, and a set of indicator variables corresponding to five years before Southwest entry, four years before Southwest entry, etc., up to four years after Southwest entry, five years after Southwest entry (one year before Southwest entry is omitted). We then plot the coefficients associated with these indicator variables against the time to and from Southwest entry, superimposing a linear fit line before entry and after entry. The vertical bars represent 95% confidence intervals. The coefficient for the year immediately before entry is set to 0 and displayed without a confidence interval because it is our baseline year.

In column (5) of Table 4, we conduct a placebo test where we randomly allocate Southwest entry events to scientist pairs. The coefficient for *Fake Southwest entry* is not significant and close to 0, suggesting that it is not the structure of the panel or changes in the data over time that are driving the result. In column (6) of Table 4, we conduct one more falsification test by looking at entry events (not included in the other regressions) where Southwest withdraws from the market within two years. For these cases, the point estimate of Southwest entry is close to 0 and insignificant.[26]

Overall, we believe results in Table 4 and Figure 1 provide robust support for a causal interpretation of our main effect, and reassure us that we are not simply measuring some underlying, unobservable process that takes place with each entry event[27] and drives both Southwest decisions and the increase in scientific collaboration.

While Southwest is the largest U.S. low-cost carrier in terms of number of passengers transported, there are other low-cost airlines operating within the same market. In Online Appendix Table A-12, we explore how our results vary depending on whether a low-cost airline is already operating on a route, as well as whether they differ when other airlines (low-cost or not) start operating a flight in the same year as Southwest. Consistent with the impact of Southwest on travel costs being largest when no low-cost alternatives existed on the same route, estimates are larger when Southwest is the first low-cost to enter (Table A-12, column (2))[28] and are positive but nonsignificant when another low-cost was already operating between the same airports (column (3)). Results are instead essentially

Table 4. Effect of Southwest Entry on Copublications at the Individual-Pair Level

DV = Copublications	(1) Baseline	(2) Controls	(3) Controls	(4) Timing	(5) Placebo 1	(6) Placebo 2
Southwest entry	0.505*** (0.121)	0.526*** (0.121)	0.526*** (0.121)			−0.029 (0.216)
Mean age		0.153*** (0.008)	0.268*** (0.015)			
Dept R&D budget per faculty (log)		0.364*** (0.127)	0.364*** (0.127)			
Years since both have a PhD			−0.230*** (0.022)			
Southwest entry (−1)				0.078 (0.152)		
Southwest entry (0)				0.485*** (0.150)		
Southwest entry (1)				0.518*** (0.166)		
Southwest entry (2)				0.582*** (0.181)		
Fake Southwest entry (random timing)					0.095 (0.121)	
Individual pair FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Number of pairs	758	758	758	758	758	171
Number of obs.	13,147	13,147	13,147	13,147	13,147	2,945

Notes. Robust standard errors are in parentheses. The dependent variable is the number of copublications between pairs of scientists. *Southwest entry* is an indicator variable that takes a value of 1 if Southwest has started operating a flight from airports close to the respective scientists. All specifications include individual-pair fixed effects and year fixed effects. Column (1) is our baseline specification. Column (2) adds controls for the age of the pair members and departmental R&D budget per faculty (both variables are means across the two pairs' members). Column (3) additionally controls for the numbers of years that have passed since both pairs' members obtained their PhD's. Column (4) replaces *Southwest entry* with a set of indicator variables corresponding to different times from or since entry: *Southwest entry* (−1) is an indicator variable if the observation is in the year preceding Southwest entry; *Southwest entry* (0), *Southwest entry* (1), and *Southwest entry* (2+) are defined analogously for the year of the Southwest entry, the year after the Southwest entry, and two years or more after the Southwest entry, respectively. Column (5) is a placebo where we pretend Southwest entry has occurred in a random year for each pair. Column (6) is a placebo where we look at the set of pairs (not included in the baseline specification) who experienced a Southwest entry followed by a Southwest exit event shortly thereafter. Estimation by Poisson quasi-maximum likelihood. DV, dependent variable; FE, fixed effects.

***$p < 0.01$.

Table 5. Effect of Southwest Entry on Collaboration Among Mathematicians

DV = *Copublications*	(1) *Mathematics*
Southwest entry	0.247**
	(0.123)
Pair fixed effects	Yes
Year fixed effects	Yes
Number of pairs	431
Number of observations	5,514

Notes. Robust standard errors are in parentheses. These regressions are based on a data set of U.S. mathematicians constructed using MathSciNet and the Mathematics Geneaology Project. DV, dependent variable.

$**p < 0.05$.

unchanged if we exclude cases where other low-cost airlines enter at the same time (column (4)), other major airlines[29] enter at the same time (column (5)), or any other airline enters at the same time (column (6)). We conclude that our results are robust to considering concurrent entry by other airlines.[30]

Results are also not driven by the fact that our sample includes only pairs that ever collaborate: when we include a random sample of noncollaborating pairs and replace individual-pair fixed effects with university-pair fixed effects,[31] we find comparable effects of Southwest entry (see Online Appendix Table A-3). In Online Appendix Table A-14, we decompose the main effect by pairs of scientists who collaborate both before and after Southwest entry (intensive margin pairs) versus pairs of scientists who collaborate either before or after entry, but not both (extensive margin pairs).[32] We find a stronger effect for intensive margin pairs (column (3)), although the cheaper fares also seem to enable experimentation in the form of new collaborations over distance (column (2)).[33]

In the online appendix, we perform additional robustness to different econometric approaches, functional forms, clustering of standard errors, treatment of outliers, and inclusion in the sample of noncollaborating pairs. In brief, we obtain qualitatively and quantitatively similar results using ordinary least squares instead of Poisson (see Table A-16, column (2)). We also obtain a positive and significant coefficient for Southwest entry (though of a somewhat smaller magnitude) when we run a linear probability model with an indicator variable for any copublication in the focal year as the dependent variable (see Table A-16, column (3)). Clustering at the city-pair level, rather than at the individual-pair level, hardly impacts the standard errors (see column (3) of Table A-2). The coefficient on Southwest entry remains significant when we exclude pairs that have more than two copublications over the entire observation period or winsorize observations with more than two copublications (see Table A-17).

3.3. Extensions and External Validity in Different Samples

The analysis and results presented in the previous section describe the effect of Southwest entry on the rate of collaboration between chemistry faculty members. Although this approach has the advantage of leveraging rich individual-level data and offers a cleaner identification strategy, one may also be interested in replicating the analysis within a field with slightly different characteristics, as well as testing external validity within a broader set of fields. To do so, we first perform a deep-dive within mathematics (a field for which we have also collected individual-level data) and then explore regressions at the region-pair level for biology, physics, and engineering. Results show that the effect we have identified within chemistry is also present across these samples.

3.3.1. Increases in the Rate of Collaboration Within Mathematics. The data set we use for mathematics includes all U.S. faculty members that have advised at least one PhD student.[34] We observe 431 pairs of individuals that experienced Southwest entry between 1993 and 2012 and have at least one copublication in that period. We adopt the same empirical strategy as in the chemistry sample and regress copublications on an indicator variable for Southwest entry, controlling for pair fixed effects and year fixed effects. Results show that Southwest entry significantly increases copublications in mathematics too.

3.3.2. Increases in the Rate of Collaboration Across Regions. To test whether the availability of cheaper flights had an effect on scientific collaboration across a broader set of fields, we also use a large-scale publication data set covering close to a million papers matched to U.S. regions (defined in terms of core-based statistical areas, or CBSAs).[35] Specifically, we explore how collaboration between any two CBSAs changed after Southwest starts operating a new route between them. The unit of analysis is the CBSA-pair-year (48,274 pairs), and we include CBSA-pair fixed effects and year fixed effects to control for underlying differences across regions that are consistent over time and for overall time trend, respectively.[36] The regressions also include linear time trends for the origin and destination CBSA. For the estimation, we use a Poisson model with standard errors clustered at the CBSA-pair level.

Results are displayed in Table 6: the point estimates for Southwest entry at this more aggregated level of analysis are significant not just in chemistry but also in biology, physics, and engineering. Although the estimated coefficients for chemistry, physics, and engineering are not statistically different from each other, the difference between chemistry and biology is significant.

Table 6. Southwest Entry and Collaborations Between U.S. Regions (CBSAs)

	(1)	(2)	(3)	(4)	(5)
DV = *Copublications*	*All*	*Chemistry*	*Biology*	*Physics*	*Engineering*
Southwest entry	0.503***	0.159***	0.494***	0.141***	0.238***
	(0.020)	(0.033)	(0.032)	(0.031)	(0.055)
CBSA pair fixed effects	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
City trends	Yes	Yes	Yes	Yes	Yes
Testing H_0			(3) = (2)	(4) = (2)	(5) = (2)
p-value			0.045	0.701	0.216
Number of pairs	48,274	15,303	22,079	15,872	7,635
Number of observations	965,480	306,060	441,580	317,440	152,700

Notes Robust standard errors are in parentheses. These regressions are run at the CBSA-pair level. The dependent variable is the number of copublications between pairs of CBSAs. *Southwest entry* is an indicator variable that takes a value of 1 if Southwest has started operating a flight from airports close to the respective cities. Column (1) is based on copublications in all journals in our sample. Columns (2), (3), (4), and (5) are based on chemistry, biology, physics, and engineering journals, respectively. All specifications include CBSA-pair fixed effects, year fixed effects, an origin-CBSA time trend, and a destination-CBSA time trend. We also report *p*-values of statistical tests for the equality of *Southwest entry* coefficients across samples. Estimation by Poisson quasi-maximum likelihood. DV, dependent variable.
    ***$p < 0.01$.

Overall, we conclude that the results from the chemistry sample are generalizable to other fields and that the increase in collaboration is larger within biology. We now turn to developing a simple model to place our main finding into the broader context of how geographic frictions shape collaboration and to guide the empirical exploration of additional predictions.

4. Theoretical Framework

The objective of this section is to develop a simple theoretical framework to highlight key trade-offs scientists face when deciding whether they should collaborate with a local or a distant coauthor and how much effort they should dedicate to a collaboration based on its intrinsic potential. The model generates novel predictions about how travel costs shape collaboration decisions, which we then test using our data.

We start by assuming that because the global pool of potential coauthors offers more variety than the local one, it is, on average, possible to find better matches when team formation is not constrained by geographic distance. The quality of a match may depend on complementary ideas, knowledge, skills, equipment, and resources that a coauthor brings to a project. Of course, because of agglomeration forces, as the size, specialization, and quality of a region's local pool increases, scientists will rely less on distant coauthors. To account for this, in an extension of the baseline model, we allow for the share of "first-best" coauthors available locally to vary.[37]

Our setup is straightforward: ideas are born with intrinsic quality q but require effort e to be developed and achieve their full potential v. Because scientists observe a noisy signal of q before starting a project, they will allocate more effort, time, and resources to projects that have higher potential (i.e., in our model, effort is endogenous to potential). At the same time, because research constitutes an uncertain endeavor, even when scientists apply effort, projects are only successful with probability p, which depends on the quality of the coauthor match. Thus, the realized value of a project can be expressed as $v = p_i q e$, where p_i (with $i = G, B$) is higher when a good match between coauthors is achieved (p_G) relative to a bad match (p_B). Although a two-sided matching framework would be more realistic, in the model, we abstract away from a setup where collaboration decisions are influenced by both sides. Our approach follows a partial equilibrium model in which the pool of potential applicants always accepts a collaboration when invited and where proposers invite potential coauthors only if they know the project is a fit for them and an interesting one for them to pursue.

Whereas the global pool may offer a better match between coauthors (i.e., p_G) and increase the chances of realizing a project's full potential v, collaborating over distance introduces additional costs, as scientists have to travel for face-to-face interactions and may be less effective at communicating complex information remotely. As a result, scientists face a trade-off between less choice locally and increased communication and travel costs over distance.

It is important to highlight that the model is not focused on the decision to collaborate (see, e.g., Bikard et al. 2015) nor on the type of project to pursue (this is discovered by the scientist at the start) but is explicitly centered on a situation where a scientist is looking for the best coauthor for a particular idea. Because we cannot empirically measure search behavior and search frictions, the model also abstracts

away from search costs and assumes that the broader talent pool a scientist is considering is formed by all the individuals a focal researcher is already aware of, has previously met at a conference, has been colocated with, or has read work by. Boudreau et al. (2017) find that search costs constitute a key friction to collaboration even within the same institution, so we make the simplifying assumption of these costs being present both for local and for distant collaboration decisions. In the model, we also abstract away from scientists' budgets: although in the regressions we are able to use departmental R&D budget data to explore heterogeneous effects, in the theory we do not account for the fact that scientists at better-funded institutions, or more productive scientists in general, may have access to larger budgets and may be therefore less sensitive to changes in travel costs. If that were the case, then the reduction in travel costs could disproportionately help lower-productivity researchers. We also do not model endogenous time to project completion as a function of travel intensity or team dynamics beyond two authors. Nevertheless, the stylized framework allows us to obtain several additional predictions that we then take to the data.

In the next sections, we perform comparative statics and explore the main tensions of the model in more detail.

4.1. Local vs. Distant Collaborations

The scientist's payoff from developing an idea with a local coauthor for a given level of effort e is

$$\pi_L(e) = p_B q e - c(e), \tag{1}$$

where $c(e)$ is the cost of effort, which we assume for tractability to have the following convex function: $c(e) = \frac{\alpha}{2} e^2$. Thus, Equation (1) can be rewritten as

$$\pi_L(e) = p_B q e - \frac{\alpha}{2} e^2. \tag{2}$$

The first-order condition yields an optimal effort level of $e_L^* = \frac{p_B q}{\alpha}$, which is increasing both in project quality q and in the quality of the coauthor match p_B. Intuitively, scientists are more willing to apply effort to projects with higher potential and to projects they are working on with better-matched coauthors. Inserting e_L^* back into (2), we obtain a scientist's payoff for a local collaboration given the optimal effort level as

$$\pi_L^* = \frac{(p_B q)^2}{2\alpha}. \tag{3}$$

How does this compare with a distant collaboration? In our setup, over distance, scientists have a higher chance of securing the ideal coauthor because the global pool offers more variety. At the same time, this does not happen all the time, and scientists have to incur additional communication and travel costs t_i to develop a project over distance. We assume that with probability z, scientists find a first-best coauthor and secure p_G, and with probability $(1 - z)$, they land a coauthor of the exact same level they would have found in the local pool p_B. Thus, the payoff for a distant collaboration can be written as

$$\pi_D(e, t) = (1 - z)\left[p_B q e_B - \alpha \frac{e_B^2}{1 + t_B} - \beta t_B^2 \right] + z\left[p_G q e_G - \alpha \frac{e_G^2}{1 + t_G} - \beta t_G^2 \right], \tag{4}$$

where e_i and t_i with $i = G, B$ are the optimally chosen levels of effort and travel for perfectly matched coauthors (p_G) versus imperfectly matched ones (p_B).

Traveling not only enters as a convex cost[38] ($t_i = [0, 1]$, scaled by a parameter β),[39] but also increases the chances of success because it improves the ability to communicate complex information, coordinate work, and make progress on a project through face-to-face interactions. This trade-off allows for interesting cases to emerge where temporary colocation between distant coauthors is expensive but also helpful, and it can therefore lead to both higher and lower payoffs relative to a collaboration on the same project with a local coauthor.

The basic dynamic we want to capture is one in which collaboration and communication require less effort when scientists are colocated but where travel can also be strategically used to recreate the same efficiencies experienced in local collaborations. When communicating over email or phone, coauthors may need more time and effort to convey the same concepts and avoid misunderstanding, and when in-person meetings are infrequent, it may take more time for a team to get everyone up to speed and make progress. In the model, distant coauthors can either spend more effort and communicate over distance or invest in travel and rely on more effective face-to-face interactions.[40]

For simplicity, we assume that once a local versus distant coauthor has been chosen for a project, it is too costly to switch types without starting a completely new project.[41] We also assume that before a substantial amount of effort and travel is dedicated to a project, the quality of a coauthor match has been revealed. The first-order conditions with respect to effort and traveling are respectively

$$e_D(t_D) = \frac{p_i q(1 + t_D)}{2\alpha},$$

$$\alpha \frac{e_D^2}{(1 + t_D)^2} - 2\beta t_D = 0,$$

where $i = B, G$ depends on whether the distant coauthor has led to a first-best or a second-best match. Combining both first-order conditions, one can show

that the optimal levels of travel and effort for a distant collaboration are

$$t_i^* = \frac{(p_i q)^2}{8\alpha\beta}, \tag{5}$$

$$e_i^* = \frac{p_i q}{2\alpha}(1 + t_i^*) = \frac{p_i q}{2\alpha}\left(1 + \frac{(p_i q)^2}{8\alpha\beta}\right). \tag{6}$$

If we plug these back into the payoff function, we obtain

$$\pi_i^* = \frac{(p_i q)^2}{4\alpha}\left(1 + \frac{1}{2}t_i^*\right) = \frac{(p_i q)^2}{4\alpha}\left(1 + \frac{(p_i q)^2}{16\alpha\beta}\right). \tag{7}$$

Thus, the overall payoff over distance is

$$\pi_D^* = (1 - z)\pi_B^* + z\pi_G^*. \tag{8}$$

Comparing the payoff equation (3) for local collaborations with Equation (8) for distant ones is informative independent of travel costs (which we discuss in detail in the next section). For example, it allows us to explore how the relative appeal of a local versus a distant collaboration changes as the comparative advantage of the global pool (z) over the local one varies:

$$\frac{\partial[\pi_D^* - \pi_L^*]}{\partial z} = \pi_G^* - \pi_B^* > 0. \tag{9}$$

Intuitively, an increase in the likelihood of finding a first-best coauthor in the global pool will lead to a relative increase in the payoff for distant over local collaborations. Similarly, if scientists enjoy a high-quality local environment with good matches (e.g., they are in an agglomerated research cluster), they will find limited benefits from collaborating over distance.

Until now, we have assumed that all scientific projects have the same probability of failure. At the same time, novel, exploratory, and cross-disciplinary projects are more likely to fail relative to incremental research or work that does not attempt to recombine knowledge across different disciplines (National Academies 2004, Wang et al. 2017). To account for this, we introduce γ and link it to the overall probability of success through $p_G = (1 + \gamma)p_B$. What we want to capture with γ is a tension between exploitation (low γ) and exploration (high γ). Exploratory projects, whether because they are novel or because they bring together disciplines that rarely interact with each other, are more likely to fail, and they benefit disproportionately from finding the right coauthor. The intuition here is that for cross-disciplinary research, a scientist needs to find the exact specialist with whom to pursue the project,[42] and similarly, when novelty is high, the returns from working with a better coauthor are also higher. Novel projects are very likely to fail to begin with and may be particularly sensitive to the weakest member of a team, as discussed in Kremer's (1993) O-ring theory. Whereas in the baseline model we

discuss novelty and across-field specialization together—as they share many similarities and both fit under the broader framework of exploratory versus exploitative research—in the extensions we separate the two constructs further by incorporating uncertainty and higher variance about the potential states of the world in Online Appendix D and by modeling coauthor specialization directly in Online Appendix E. Because the implications are similar, in this paper we focus on the simplified implementation based on γ.

For a given p_B, a low γ means that the quality of the match between coauthors will have a minor influence on the chances of realizing a project's full potential. Low γ exploitation projects are therefore relatively more straightforward research where most of the techniques and ideas are established (or everyone has access to similar infrastructure to work on them), and the gap between working with the best possible coauthor versus anyone else is small. When γ is low, the relative appeal of the global talent pool is more limited. For high γ exploration projects, instead, scientists will be more willing to travel to work with the ideal coauthor and increase their chances of success:

$$\frac{\partial[\pi_D^* - \pi_L^*]}{\partial \gamma} = z\frac{\partial[\pi_G^*]}{\partial \gamma} > 0. \tag{10}$$

An extreme example of this from a specialization perspective is a project for which there are only a few leading experts or key labs with the right equipment (e.g., CERN (European Organization for Nuclear Research), LIGO (Laser Interferometer Gravitational-Wave Observatory)), and the difference between working with them relative to working with a local alternative is large.

Last, when comparing local versus distant collaborations, it is useful to point out that increases in the underlying, intrinsic project quality (q) have an ambiguous effect on the choice of coauthor type. As shown in the online appendix, which type of collaborations prevail still depends on the basic trade-off between the quality of the match between scientists and travel costs (because a distant collaboration can still leave a scientist with a match of similar quality to the local alternative).

4.2. Reductions in Travel Costs and Changes in the Types of Collaborations

How does a reduction in travel costs affect the types of collaborations scientists engage in? In this section, we perform comparative statics to see how cheaper fares such as the ones brought by a low-cost airline change the relative attractiveness of local versus distant collaborations, and how this effect varies for projects of different types (higher versus lower potential, novelty, interdisciplinarity, etc.). To simplify the notation and exposition, we define $\theta = \frac{1}{\beta}$ (which is

the inverse of travel costs) as the "ease of travel." One can think of an improvement in θ as better infrastructure that allows scientists to meet with their distant coauthors at a lower cost and with lower frictions. The derivative of relative returns with respect to θ is

$$\frac{\partial[\pi_D^* - \pi_L^*]}{\partial \theta} = \left(\frac{q^2}{8\alpha}\right)^2 [(1-z)p_B^4 + zp_G^4] > 0; \qquad (11)$$

θ does not matter for the returns to local collaborations (π_L^*) as no travel is required, but it makes face-to-face interactions with distant coauthors less expensive. Therefore, it is intuitive that with better travel technology the relative attractiveness of the global talent pool increases,[43] as accessing it is now more cost effective.

But how does this effect vary with the ex ante relative competitiveness of the local pool? That is, how does this vary for regions that offer better versus worse alternatives to begin with? Remember that this is captured in our framework by the share of first-best coauthors who are in the global pool z. Taking the first-order condition with respect to θ and z, we obtain

$$\frac{\partial[\pi_D^* - \pi_L^*]}{\partial \theta \partial z} = \left(\frac{q^2}{8\alpha}\right)^2 \left(p_G^4 - p_B^4\right) > 0, \qquad (12)$$

which leads to the following prediction.

Prediction 1. *A reduction in travel costs will be especially beneficial for researchers who have access, ex ante, to a relatively worse pool of local coauthors.*

From an empirical standpoint, if highly productive researchers embedded in worse local environments start substituting local collaborations with better-matched ones over distance, we should see evidence of crowding-out behavior.

If we take the derivative of relative returns with respect to quality too,

$$\frac{\partial[\pi_D^* - \pi_L^*]}{\partial \theta \partial q} = q\left(\frac{q^2}{4\alpha}\right)^2 [(1-z)p_B^4 + zp_G^4] > 0, \qquad (13)$$

we see that after an improvement in the ease of travel, higher-quality projects are more likely to be undertaken with better-matched coauthors (which are more abundant over distance). The intuition here is that as travel costs fall, scientists are more likely to travel to match with a better coauthor. This is disproportionately valuable when the returns to travel and effort on a project are high to begin with (i.e., for ideas of high potential). This leads us to our second prediction.

Prediction 2. *A reduction in travel costs will be especially beneficial for distant collaborations on higher-quality projects.*

Last, if we do not assume that all projects have the same probability of failure and separate exploratory from exploitative projects by introducing γ, we obtain

$$\frac{\partial[\pi_D^* - \pi_L^*]}{\partial \theta \partial \gamma} = z\left(\frac{q^2}{4\alpha}\right)^2 (1+\gamma)^3 p_B^4 > 0, \qquad (14)$$

which shows that a reduction in travel costs makes the global pool disproportionately more appealing for exploratory projects (high γ),[44] which can be restated as follows.

Prediction 3. *A reduction in travel costs will be especially beneficial for distant collaboration on novel or cross-disciplinary projects.*

Empirically, to proxy for γ, we will rely on how novel the keywords used by the authors on a focal paper are, as well as explore results for collaborations that span different subfields of chemistry versus those that not. We now return to the data to test these predictions.

5. Testing the Theoretical Framework
5.1. Types of Scientists Affected and Crowding Out
Having shown evidence in Section 3 that Southwest entry led to a plausibly causal increase in collaboration between the affected scientists, we now take advantage of this source of exogenous variation to test the additional predictions from the theoretical framework.

The first prediction of the model focuses on the impact travel costs have on scientists embedded within better versus worse local research environments. Intuitively, agglomerated regions with a greater number of potential collaborators offer, on average, better local matches to begin with, which makes the global scientist pool relatively less appealing. Because it is difficult to build accurate proxies for the number of ideal coauthors a specific scientist may have access to without traveling, we rely on past productivity to assess whether a scientist from a given department is more or less likely to find a good match locally.

As can be seen in panel A of Table 7 (columns (1)–(3)), the increase in collaboration we observe in chemistry after the arrival of Southwest is driven by scientist pairs where at least one member is more productive than her local peers, and it is even more pronounced when both scientists are more productive than their colleagues. In the mathematics sample, where we only have a small number of observations in column (1), the effect is positive and significant only for pairs that are both more productive than their local peers (column (3)), possibly because distant collaborations are more rare and selected in this field to begin with.

Overall, the cheaper fares seem to be particularly helpful for individuals who are talented but potentially do not have access to coauthors of comparable quality within their local environment. They might be in

Table 7. Effect of Southwest Entry on Copublications: Which Pairs Are Most Affected?

	Panel A: Chemistry				
	(1)	(2)	(3)	(4)	(5)
DV = Copublications	Both less productive	One more productive	Both more productive	Same type of chemistry	Different type of chemistry
Southwest entry	0.228	0.566***	0.863***	0.340**	0.668***
	(0.292)	(0.153)	(0.272)	(0.164)	(0.165)
Pair fixed effects	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Testing H_0		(2) = (1)	(3) = (1)		(4) = (5)
p-value		0.015	0.003		0.002
Number of pairs	154	403	101	417	341
Number of observations	2,498	6,597	1,630	7,183	5,964
	Panel B: Mathematics				
	(1)	(2)	(3)	(4)	(5)
DV = Copublications	Both less productive	One more productive	Both more productive	Same type of math	Different type of math
Southwest entry	0.196	0.002	0.374**	0.260	0.204
	(0.364)	(0.192)	(0.155)	(0.173)	(0.158)
Pair fixed effects	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Testing H_0		(2) = (1)	(3) = (1)		(4) = (5)
p-value		0.638	0.653		0.900
Number of pairs	61	155	263	180	299
Number of observations	859	2,001	3,318	2,223	3,955

Notes. Robust standard errors are in parentheses. Panel A corresponds to our main chemistry sample. Panel B corresponds to the mathematics sample. The dependent variable in all specifications is the number of copublications. Different columns correspond to different subsamples in terms of productivity (columns (1)–(3)) and whether both pair members are specialized in the same subfield (column (4)). Productivity is measured at the time of Southwest entry. All specifications are estimated by Poisson quasi-maximum likelihood and include year fixed effects and pair fixed effects. DV, dependent variable.
    * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

peripheral institutions because of imperfections in the labor market or simply because of their geographic preferences. With lower travel costs, these individuals are able to find and sustain better matches over distance.

As mentioned in the theoretical framework, a natural consequence of highly productive scientists prioritizing distant coauthors in their collaboration portfolio because of the lower fares is a crowding-out effect on local collaborations. In Table 8, we explore whether the cheaper fares have a negative impact on the local collaboration environment. Whereas local copublications are slightly increasing (column (1)), this result is really a composition of two different, countervailing effects. On the one hand, less productive pairs seem to be working together more with each other (column (3)). On the other hand, we see a sharp decline in collaborations between above-average productivity scientists and their local, below-average productivity peers (column (2)).[45] This makes sense, as higher-productivity individuals are also the ones who respond the most to Southwest entry to begin with. Interestingly, we find this crowding-out pattern both in the chemistry and in the mathematics samples, suggesting that when better options become available over distance, highly productive scientists substitute local

collaborations with potentially better-matched ones over distance.

In Online Appendix Table A-19, we present additional splits of the data beyond those predicted by the theory. The effect of Southwest is stronger for younger scientists (panel A) and scientists who are more distant from each other (panel C). We do not find a statistically significant difference in effect size by departmental R&D budget (panel B), even though the estimate for departments with low budgets is almost twice as large as the other ones. Finally, pairs where one or both scientists are female do not respond to lower travel costs, possibly because women may have more constrained travel schedules.

5.2. Changes in the Type of Projects
The next set of predictions of the model link the reduction in geographic frictions to an increase in the amount of time and effort allocated to higher quality (Prediction 2) and more novel or cross-disciplinary projects with distant coauthors (Prediction 3). As discussed in Section 2.1, we proxy for the quality of projects using citations and for high γ projects by looking both at projects that span different subfields and at research that uses novel keywords or belongs to an

Table 8. Southwest Entry and Local Copublications

	(1)	(2)	(3)
	All local pairs	More productive pairs with less productive local colleagues	Less productive pairs with less productive local colleagues
		Panel A: Chemistry	
Southwest entry	0.092**	−0.686***	0.169**
	(0.040)	(0.262)	(0.070)
Pair fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
No. of pairs	741	126	547
No. of obs.	12,939	2,305	9,533
		Panel B: Mathematics	
Southwest entry	0.196***	−0.838*	0.471**
	(0.064)	(0.457)	(0.219)
Pair fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
No. of pairs	896	60	186
No. of obs.	11,665	848	2,484

Notes. Robust standard errors are in parentheses. Panel A corresponds to our main chemistry sample. Panel B corresponds to the mathematics sample. In both samples, we construct the set of local copublications of pairs affected by Southwest entry and use it as the dependent variable in column (1). We also tag the set of local copublications with local colleagues whose productivity is below departmental average in the years preceding Southwest entry and use it as the dependent variable in columns (2) and (3). The specification of column (2) is run on the sample of pairs where both members are above departmental average in productivity. The specification of column (3) is run on the sample of pairs where one or both members are below departmental average in productivity. Pairs that have all zero outcomes are dropped from the respective regressions, which results in the number of observations in columns (2) and (3) not summing up to the number of observations in column (1). All specifications are estimated by Poisson quasi-maximum likelihood and include pair fixed effects and year fixed effects.
*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

emerging novel trend. If exploratory projects (high γ) are more likely to fail, then our estimate will likely underestimate the full impact of a reduction in travel costs on this set of projects, as many will be abandoned and never turn into a publication to begin with.

In terms of project quality, in column (1) of Table 9, we condition on collaboration and weight the dependent variable, *copublications*, by citations received (a proxy for scientific impact and quality). Consistent with Prediction 2 and with the idea that lower travel

costs induce scientists to allocate disproportionately more effort to distant collaborations as quality increases, we observe a larger effect of Southwest entry on right tail projects.

In terms of interdisciplinarity, in panel A, columns (4) and (5) of Table 7, we see that after Southwest enters, collaborations between scientists specialized in different subfields of chemistry increase disproportionately relative to other types of collaborations.[46] These interdisciplinary projects may benefit more from face-to-face interactions because of a greater

Table 9. Effect of Southwest Entry on the Type of Collaborations

	(1) Cites	(2) Novel trends	(3) Mean share novel	(4) Max share novel	(5) Total share novel	(6) Novel in field	(7) Equipment intensive 1	(8) Equipment intensive 2
Southwest entry	0.420*	1.175*	0.057*	0.069*	0.093*	−0.008	0.839***	0.647*
	(0.234)	(0.687)	(0.034)	(0.040)	(0.048)	(0.085)	(0.303)	(0.356)
Pair fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of pairs	189	37	758	758	758	758	74	74
Number of obs.	606	137	1,177	1,177	1,177	1,177	261	261

Notes. Robust standard errors are in parentheses. These regressions are conditional on the pair having collaborated in the focal year. The dependent variables are the number of cites received (column (1)), various measures of novelty based on keywords (columns (2)–(6)), and the number of equipment-intensive collaborations based on keywords (columns (7) and (8)). Pairs that never have a nonzero value of the dependent variable are dropped from the regressions. All specifications include individual-pair fixed effects and year fixed effects. Estimation by Poisson quasi-maximum likelihood.
*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

need to exchange complex information that may be new for at least one of the participants or because these pairs cannot rely on a shared, discipline-specific vocabulary to streamline communications over distance.

Beyond complementarities in ideas and knowledge, cross-disciplinary work between specialized labs can also be captured through complementarities in equipment and infrastructure. In columns (7) and (8) of Table 9, the dependent variables are the number of equipment-intensive copublications (*Equipment 1*) and the count of equipment-related keywords used in the focal papers (*Equipment 2*), respectively.[47] Although these types of collaborations are more rare, the effect of Southwest entry is large and significant, suggesting that, at least within chemistry, specialization driven by equipment may play a key role in how scientists select into distant collaborations. In Online Appendix Table A-20, we perform a similar regression without relying on equipment-related keywords but taking advantage of data on capital intensity by department: results show that the lower fares have the largest effect on collaborations where one of the scientists belongs to a capital-intensive group and the other one does not.

Finally, in columns (2)–(6) of Table 9, we look directly at different measures of novelty. Results are consistent across the dependent variables and provide further support for Prediction 3. After Southwest enters, we see an increase in collaborations that focus on emerging novel trends and topics that are about to become "hot" (column (2)), as well as an increase in the use of novel keywords (columns (3)–(5)). Interestingly, when we define novelty within the more narrow confines of a subdomain (column (6)), the result is insignificant, possibly because some of the novel uses from columns (2)–(5) may represent ideas that are being slowly incubated within a subdomain but have not diffused more broadly yet.

Next, we study how Southwest entry changes the types of collaborations that are pursued at the regional level. The analysis of collaborations at the dyadic level between chemistry faculty members already suggests that lower travel costs are particularly beneficial for higher-quality, interdisciplinary, equipment-intensive, and more novel projects. However, the estimates are significant but noisy because of the smaller sample size when looking at these rare outcomes. We therefore replicate our analysis within a broader set of papers in chemistry and related fields. The analysis is at the CBSA-pair-year level and includes CBSA-pair fixed effects and year fixed effects.

The effects (see Table 10) are consistent with our previous findings and highlight that lower travel costs have a disproportionate effect on the right tail of the quality distribution and on more novel, cross-disciplinary, and equipment-intensive projects. The impact of these changes is large, with increases in aggregate output between 9% and 40% (column (1)). This corresponds to roughly 300 extra copublications per year.[48] We also find support for the other predictions. Novel ideas increase between 15% and 80% (column (4)), but the estimate is noisy (as in Table 9) when we estimate novelty within subfields, and equipment-intensive collaborations increase between 5% and 50%.

Our analyses so far have focused on how reductions in travel costs induced by Southwest entry affect pair-level outcomes. Our model, however, also predicts that higher impact projects should mostly result from distant rather than local collaborations. To test this implication, we focus on individual researchers and change the unit of analysis to a paper, which we flag as local or distant.[49] We then regress the number of citations a paper receives on an indicator variable for whether the paper resulted from a distant collaboration (the omitted category being a local

Table 10. Effect of Southwest Entry on the Type of Collaborations (CBSA Level)

		Quality	Type						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Copubs	Cites-weighted copubs	Novel trends	Novel overall	Novel in field	Across-field copubs	Equipment intensive 1	Equipment intensive 2
Southwest entry	0.208***	0.240**	0.235***	0.365***	0.125*	0.266**	0.264***	0.236**
	(0.0647)	(0.0995)	(0.0710)	(0.112)	(0.0687)	(0.116)	(0.0871)	(0.0939)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CBSA-pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	40,227	39,517	34,847	28,616	36,704	27,379	29,767	29,767

Notes. Robust standard errors are in parentheses. These regressions are run at the CBSA-pair level and are based on a large sample of publications in chemistry and chemistry-related fields. The dependent variable in column (1) is the number of copublications between pairs of CBSAs. Column (2) uses citation-weighted copublications as the dependent variable. Columns (3)–(5) use our different specifications of novelty. Column (6) counts across-field collaborations and columns (7) and (8) equipment-intensive ones, as inferred from the keywords associated with the papers. FE, fixed effects.
*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Table 11. Quality of Distant and Local Collaborations at the Scientist Level

	(1)	(2)
	Cites	*Cites*
	Chemistry	Math
Distant collaboration	0.0674***	0.2277***
(local collaboration omitted)	(0.0026)	(0.0069)
Individual fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Number of pairs	4,737	2,104
Number of observations	46,060	12,187

Notes. Robust standard errors in parentheses. The data underlying these regressions are the same as in the main analyses but are now structured at the individual scientist level. An observation is a paper that we tag as either a distant or a local collaboration; papers not involving a collaboration with another faculty member are excluded. The dependent variable is the number of cites received, and the variable of interest is whether the paper is a distant collaboration, with local collaborations as the omitted category. All specifications are estimated by Poisson quasi-maximum likelihood and include individual scientist fixed effects and year fixed effects.

    ***$p < 0.01$.

collaboration), controlling for year fixed effects and scientists fixed effects. We find that long-distance collaborations get 6%–7% more cites than local ones in chemistry and 21%–24% more in mathematics (Table 11). These results are consistent with the often-reported stylized fact that distant collaborations are more heavily cited (Jones et al. 2008).

6. Conclusions

The paper explores how geographic frictions, and in particular travel costs, shape the rate and direction of scientific research. Whereas previous work has mostly focused on communication costs and their impact on the rate of collaboration, our paper emphasizes other effects distance-related frictions can have on innovative outcomes, including the type of projects that are pursued with local versus distant teams.

Whereas both Gaspar and Glaeser (1998) and Kim et al. (2009) have suggested that the secular decline in air travel costs might have led to an increase in scientific collaborations, they do not take their prediction to the data, making this the first study to do so to our knowledge. We build on a vibrant literature that has looked at how scientists respond to reductions in communication costs and how changes in the infrastructure of collaboration can counterbalance preexisting geographic frictions (Agrawal and Goldfarb 2008, Ding et al. 2010). In particular, our finding that highly productive scientists who are embedded in worse local environments disproportionately benefit from reductions in travel costs is complementary to the Agrawal and Goldfarb (2008) result that reductions in communication costs allow for better

matches between top-tier and middle-tier institutions from the same region. It is also consistent with the Ding et al. (2010) finding that lower communication costs help scientists from non-elite institutions. Relative to reductions in communication costs, which Ding et al. (2010) show have a positive effect on female scientists, in our setting, lower travel costs only help men—possibly because female scientists have more constraints on their travel schedules.[50]

The paper also extends work that studies the effect of geographic frictions at a much smaller scale, as it provides insights on how easier access to better, distant collaborators influences local collaboration decisions. Whereas studies at the microgeographic level have shown that colocation influences the probability and quality of collaboration (Catalini 2018), we show that additional, and at times opposing forces may be at work at a larger scale through travel costs. Our findings also call for more research on the exact form search costs take when scientists explore collaborations with local versus distant coauthors: whereas our model abstracts away from these frictions, Boudreau et al. (2017) show that they are a major obstacle even for colocated individuals.

Our theoretical framework builds on the tension between lower collaboration costs when colocated and the availability of a broader set of potential collaborators over distance. We start from this basic trade-off and then explore some of the key choices scientists face when deciding whether they should collaborate locally versus over distance, how much effort to allocate to projects of different potential, and with whom they should pursue a more novel or interdisciplinary project. We test the predictions from this framework by taking advantage of a source of plausibly exogenous variation in travel costs: the differential timing of entry by a low-cost carrier across multiple U.S. airports. Our difference-in-differences empirical strategy, combined with a series of robustness and falsification tests, supports the idea that the availability of lower fares has a causal effect on the probability and intensity of collaboration between scientists.[51] The effect is particularly pronounced for scientists who are less likely to find coauthors of the same quality within their local environment; is present across multiple fields of science (chemistry, physics, biology, engineering, mathematics); and is robust to controlling for idiosyncratic scientist-pair characteristics, trends in collaboration over time, and department R&D budgets. Moreover, we do not observe a pretrend in collaboration between scientist pairs who are going to experience lower air travel costs in the future.

Consistent with the theory presented, the reduction in geographic frictions also transforms the types of projects that emerge, influencing the direction of innovation:[52] our estimates suggest a sizable increase

in higher-quality papers and in projects that span different subdisciplines, are more intensive in their use of specialized equipment, and are more novel. Comparisons between our findings in chemistry and mathematics suggest that complementarities in specialized equipment—although important for collaboration decisions between distant labs—are not the only driver behind the observed increase in joint projects over distance. Scientists also launch more experimental projects and projects that seem to take advantage of the complementary skills, ideas, and knowledge that a distant laboratory may contribute to a collaboration.

Beyond the lower fares introduced by the low-cost airline we study in the paper, the cost per mile in the United States has dropped by over 50% in the last 30 years (Perry 2014),[53] and convenience and routes have greatly improved. Our results should be therefore interpreted within this broader context of improvements in our ability to travel and work with distant collaborators. Whereas we cleanly estimate the impact of only part of these changes, improvements in air travel are likely to affect a much larger population of individuals. This includes inventors and researchers working within firms or public organizations that have multiple sites and face a similar trade-off between the ability to form ideal teams when not constraining their search for participants to one location and the additional communication, coordination, and travel costs geographically dispersed teams entail.[54] As advancements in communication technology make online interactions increasingly closer in latency and fidelity to offline ones, more research is needed to understand why face-to-face exchanges still appear to be a complement rather than a substitute to remote ones. In particular, whereas online exchanges seem to work well for executing on existing ideas, offline ones may still offer greater serendipity (Catalini 2018). Moreover, trust between participants—often a prerequisite for collaboration when uncertainty makes it difficult to precisely evaluate individual contributions and effort—seems to still depend on individuals having spent enough unstructured time together in the same location.

Overall, relative to location decisions, which are extremely expensive for organizations to shape in the short run, this paper shows that investments targeted at facilitating travel and incentivizing face-to-face interactions may have higher returns than previously expected. This is because they affect not only the intensity of collaboration but also the quality and impact of the resulting work. By facilitating better matches and more productive teams, they also support novel recombinations of ideas. Although it may be tempting for firms and public funding agencies to assume that they can rely on technology to reduce costs and replace travel, our results support the view that—at least in the case of innovative outcomes—this is unlikely to be the case. The findings also show that this constitutes an opportunity for these organizations, as support for travel is a flexible policy lever that can be adjusted over time to shape R&D trajectories.

Whereas geographic distance acts as a sizable disincentive to collaboration and idea recombination, organizations can institutionalize and encourage travel to offset its effect. From a policy perspective, support for travel and better infrastructure can also be used to offer better opportunities to individuals and organizations located away from key economic and innovation hubs, to influence the strategic location decisions of firms, and to ultimately support economic growth. Firms with a geographically dispersed customer base rely on air travel (both for cargo and for employees) to remain competitive, and there have been multiple cases in recent years of headquarters being moved, among other reasons, for better access to flights: for example, in 2017, Illinois-based Caterpillar Inc. decided to move from Peoria to Chicago to better serve its customers, the vast majority of whom are international.[55] Further exploring the trade-offs geographic frictions introduce for individuals and firms in these different contexts is a fruitful area for subsequent research.

Acknowledgments

Endnotes

[1] Agrawal and Goldfarb's (2008) study of Bitnet, an internet predecessor, finds that as more academic institutions joined the network, collaboration among affected scientists increased. Interestingly, their results hint at the technology being a complement to offline interactions, as coauthorship increases disproportionately among university pairs that are colocated. Other studies have found an effect of Bitnet on collaborations in the academic life sciences (Ding et al. 2010) and of the internet on cooperative R&D between firms (Forman and Zeebroeck 2012).

[2] By 2000, less than 20% of papers in science and engineering were single-authored. Similar patterns—and in particular, the rise of coauthorship and distant coauthorship—have been documented in economics. See Gaspar and Glaeser (1998), Hamermesh and Oster (2002), and Rosenblatt and Mobius (2004).

[3] Chai and Freeman (2018) compare collaboration patterns among attendees of a Gordon conference before and after the event in a difference-in-differences framework using a carefully constructed control group of qualitatively similar nonparticipants. They find that attendees are more likely to be cited by and collaborate with other participants, especially if

they were new to this community of experts. In a related paper, Campos et al. (2018) document that a conference cancellation led to a decrease in individuals' likelihood of coauthoring together.

[4] Southwest has been described as the most significant development in the market structure of the U.S. airline industry by the Transportation Research Board (1999) and by industrial economists (Morrison 2001, Borenstein and Rose 2007, Goolsbee and Syverson 2008).

[5] Kim et al. (2009) and Freeman et al. (2014) note that secular declines in both communication costs and air travel costs may have facilitated long-distance collaborations.

[6] Chemistry, which focuses on the composition, structure, transformations, and properties of matter, is a large discipline, with chemistry PhD graduates accounting for about 15% of U.S. PhD life and physical science graduates (NSF 2015).

[7] The American Chemical Society also produced a CD-ROM for the years 1991–1993.

[8] Scopus is one of the two major bibliometric databases (along with ISI Web of Science). Our set of chemistry journals includes all journals from the American Chemical Society, as well as any chemistry journal with an impact factor above 2. Our set of multidisciplinary journals includes *Nature, Science, Cell,* and the *Proceedings of the National Academy of Sciences of the United States of America.* Our set of major journals in neighboring disciplines includes all journals with an impact factor above 6 in physics, biology, material science, and nanotechnology.

[9] The majority (75%) of papers matched to a faculty member have exactly one faculty author, 21% have two, and less than 4% have more than two authors. Both papers with one faculty author and papers with multiple faculty authors typically have several nonfaculty authors. We focus on faculty authors because they are the ones usually making the decision to collaborate. Papers in chemistry journals that are not matched to any of our U.S. faculty authors are likely to be from foreign scientists, scientists working in corporate environments, and federal labs.

[10] We are interested in the tension between aggregate and subfield-specific novelty because our theoretical framework predicts that across-field collaborations should disproportionately benefit from reductions in travel costs, and we want to test whether some of these represent arbitrage of ideas between subfields of science.

[11] This was built by scraping and compiling an inventory of equipment for sale in online catalogues and stores targeted at a wide range of chemistry labs.

[12] The regression yields the following classification: physical chemistry, analytical chemistry, and biochemistry as capital-intensive fields; and organic chemistry, inorganic chemistry, and material science as not capital intensive. Discussions with domain experts and anecdotal evidence support this classification.

[13] Our dyadic data are not directed and thus are symmetric: the pair between i and j is the mirror image of the pair between j and i. The 15,244 figure is obtained after dropping an equal number of symmetric observations.

[14] Scientists in our sample may move from one department to another, in some cases leading to a change in whether they are connected by Southwest or not. We want changes in Southwest status to be driven by Southwest entry decisions rather than by scientist location decisions, and thus we exclude pairs who happen to move in the same year as Southwest enters, the year before or the year after.

[15] Results are robust to decreasing this threshold to 100 or 50 miles.

[16] Correlation tables, as well as summary statistics for all samples, are in the online appendix.

[17] Specialization is inferred from the journals in which a scientist publishes. For instance, a faculty member who often publishes in the *Journal of Biological Chemistry* is assumed to be specialized in biochemistry. See Online Appendix Table A-4.

[18] A pair lasts in our sample for a maximum of 21 years (from 1991 to 2011). We observe some pairs for less than 22 years because of pair members starting their first faculty appointment after 1991, retiring before 2011, or otherwise no longer being listed in the ACS faculty directory (e.g., because they moved to industry or to a foreign country).

[19] This makes sense because a longer observation period mechanically increases the chances of experiencing Southwest entry.

[20] The 95% confidence interval for the number of passengers expressed in percentage change is $[\exp(0.4437 - 1.96 \times 0.005) - 1; \exp(0.4437 - 1.96 \times 0.005) + 1]$ or $[0.543; 0.574]$.

[21] The data from the Bureau of Public Transportation include the number of miles flown for each itinerary. Differences in miles flown arise from the number of connections an itinerary involves. We compute the average miles flown as the average across all passengers travelling between two airports in a given year.

[22] Collaboration between scientists is increasing over time. In our regressions, this trend is captured by the inclusion of year fixed effects. Therefore, one can interpret our estimates as the relative percentage increase in collaboration because of Southwest entry once the underlying increasing trend in collaboration has been accounted for.

[23] The point estimate is $\hat{\beta} = 0.505$, and the standard error 0.121. So the lower bound of the 95% confidence interval expressed in percentages is $(\exp(0.505 - 1.96 \times 0.121) - 1)/100 = 30.6\%$, and the upper bound is $(\exp(0.505 + 1.96 \times 0.121) - 1)/100 = 110.2\%$.

[24] We adopt this particular specification because it is the same used by Bernstein et al. (2016) in their study of venture capital monitoring and air travel costs, but we show robustness to specifications with additional years in Figure 1.

[25] We repeat the same graph within the large sample at the CBSA-pair level in Online Appendix Figure A-1.

[26] Although it may seem counterintuitive that early withdrawals are not associated with an effect, but that we also obtain a positive estimate for *Southwest entry* (0) in column (4) of Table 4, it is important to highlight that (a) 98% of early withdraws occur in the year of entry; (b) when we separate entry events by quarter of the year, we find only a positive effect for the year of entry when Southwest starts serving a route in the first or second quarter of the year; and (c) the confidence interval for *Southwest entry* (0) in column (4) and for early withdrawals in column (6) overlap—so in a statistical sense, we cannot rule out the possibility of a positive effect for these short spells associated with a withdrawal, even if the estimate is noisy.

[27] We observe Southwest entry across multiple locations and years, which makes it unlikely that some other simultaneous event is always co-occurring with the shocks we use.

[28] Our list of low-cost airlines includes AirTran Airways Corporation, JetBlue, Frontier Airlines, Spirit Air Lines, ATA Airlines, Allegiant Air, Virgin America, Sun Country Airlines, ValuJet Airlines, and Vanguard Airlines.

[29] We classify as major airlines the following: Delta, American Airlines, United Airlines, US Airways, Northwest Airlines, Continental, America West Airlines, Alaska Airlines, Trans World Airlines, and Envoy Air. These correspond to the 10 companies with the largest numbers of passengers carried between 1993 and 2012.

[30] One might also wonder about additional modes of transportation. As shown in Online Appendix Table A-13, we find no effect of Southwest entry in the Northeast corridor, where train travel has been a consistent alternative to flying.

[31] If we were to run this regression with individual-pair fixed effects, the noncollaborating pairs would be dropped from the estimation.

[32] It is useful to highlight that because our unit of analysis is a pair-year, and we observe the extensive margin pairs both before and after entry, the Southwest entry variable is not absorbed by the pair fixed effects and does have variation within these pairs. Some of these extensive margin pairs do not collaborate before the event, and others

do not collaborate after, so the estimated effect is a composition of the behavior of both types of pairs. Because the estimate is positive and significant, we infer that, on average, Southwest entry is associated with more non–previously collaborating pairs engaging in collaboration than the other way around.

[33] We also investigate whether the stronger effect for the intensive margin pairs is driven by continued collaboration between former doctoral students or postdocs and their advisors and sponsoring principal investigators. To code these advisor-advisee pairs, we take advantage of the ordering convention for author names in the field (juniors are typically listed first and principal investigators last). Results are reported in Online Appendix Table A-15 and are suggestive of stronger effects among these special pairs, but the sample size is too small to reach a conclusion.

[34] This database is based on MathSciNet, an abstracting service run by the American Mathematical Project and the Mathematics Genealogy Project, which is targeted at tracking PhD theses in mathematics. We construct a sample of U.S.-based mathematicians who advise at least one PhD student, and we deduce their location from the institution their students graduate from.

[35] The starting point for the construction of this sample is the population of scientific articles published in the top 477 scientific journals in biology, chemistry, physics, and engineering between 1991 and 2012. We have a total of 2,773,560 papers, of which 1,169,458 have at least one author with a U.S. address. Of all papers with U.S. addresses, we are able to successfully map 994,672 (85%) to a U.S. CBSA using a combination of three different geocoding services (Google Maps API, Bing Maps API, and the Data Science Toolkit). This allows us to link the vast majority of U.S. papers to the geographic regions involved in their production.

[36] Although this approach has the advantage of considering different fields of science, it also has important limitations. We can no longer include scientist-pair fixed effects and account for idiosyncratic, unobservable, and time-invariant reasons that may drive collaboration between any two scientists. CBSAs may also be too large as a unit of analysis for correctly measuring the effects of interest. Finally, our ability to test the full set of predictions of the model is limited.

[37] The fraction of first-best coauthors in the global pool is assumed to be z. Because the global pool can be seen as an average over all possible local pools, the fraction of first-best coauthors in a given local pool w can be either higher, lower, or equal to z. If $w > z$, then scientists will never collaborate over distance, as they would incur additional costs but would not be more likely to find an ideal coauthor over that distance. Therefore, the range of values of w that provides a meaningful trade-off is $0 \le w < z$. To simplify the exposition, in this paper we will assume $w = 0$. More general cases are discussed in the online appendix.

[38] Longer air travel incurs additional costs, including a longer time in the air, additional transfers, the inability to perform the trip within a day, accommodation costs, time zones, and fatigue. In the data, fare prices also increase more than proportionally with distance (i.e., when we estimate TicketPrice $= \beta_0 + \beta_1 \times$ Distance $+ \beta_2 \times$ Distance^2, we systematically obtain a coefficient $\beta_2 > 0$).

[39] If $t_i = 1$, face-to-face communication is always available (as with a local collaborator), and the cost of effort would be the same under both scenarios. Advancements in communication technology and virtual reality can be therefore thought of as changes in t_i.

[40] Furthermore, because coauthors invest more in travel and in-person meetings, one could imagine that the two effort functions should look more similar; in our model, as t converges to its upper bound ($t = 1$), $e_i^2/(1 + t_i)$ converges to the cost local coauthors face, $e_i^2/2$. That is, one can think of the effort cost under colocation to be a particular case of a distant collaboration facing the minimum possible cost of effort.

[41] One intuitive way to think about changing a coauthor within our simple framework is to imagine the original project failing and a new one being launched with a different team.

[42] For example, one that is able to understand and interpret the contributions and language from another discipline.

[43] Notice that this holds for the general case of $0 \le w < z$, and it is not limited to cases where $w = 0$.

[44] Intuitively, in our model, this is a result of the complementarity between γ and the quality of a coauthor match.

[45] Additionally, when we consider the quality of the local collaborations of these above–average-productivity scientists and their local, below-average-productivity peers, we find that it goes down following Southwest entry (Online Appendix Table A-18, column (1)), although the effect on novelty is insignificant (column (2)).

[46] A result that we do not find in mathematics, possibly because some of the cross-disciplinary collaborations within chemistry may be driven by access to specialized equipment.

[47] See Section 2.1 for additional details.

[48] The sample mean for copublications at the CBSA-pair-year level is about 2.1, leading to an increase of 0.42 per CBSA-pair-year. We have about 2,100 pairs per year, of which around one-third are in treatment status. So a back–of-the-envelope estimate is $0.42 \times 2,100 \times 0.33 = 294$.

[49] Papers not involving a collaboration with another faculty member are excluded.

[50] Policies targeted at reducing geographic frictions through lower travel costs may therefore need to account for the total cost of travel (including the opportunity cost of time) different types of individuals actually face.

[51] A back-of-the-envelope calculation suggests that Southwest entry induced close to 400 copublications among chemistry faculty pairs. The sample mean of 0.1 copublications per year increases by 50% to 0.15 copublications per year after Southwest entry. We have 750 pairs and 10 postentry years on average, leading to a back-of–the-envelope estimate of $0.05 \times 750 \times 10 = 375$ copublications. Although this number is sizeable, it is small relative to the total number of copublications among chemistry faculty members in this period. However, Southwest entry corresponds to a 20% price reduction, affecting only a fraction of faculty pairs (a large fraction of pairs are served by Southwest or other low-cost carriers before our observation period). Over the last 30 years, the cost per mile for air travel across all routes within the U.S. dropped by 50% (Perry 2014). This suggests that reductions in air transportation costs overall could have had a substantial aggregate effect on collaboration, above and beyond the particular source of variation in air travel cost we use in this paper.

[52] For additional work focused on changes in the direction of research, see Furman and Teodoridis (2017) and Catalini (2018).

[53] We would expect similar effects in Europe, where low-cost airlines had even more of an effect on market structure and competition, as well as on uniting different economies.

[54] There is an interesting parallel here with the literature on communications costs and collaboration: whereas Agrawal and Goldfarb (2008) focused on academic collaborations, Forman and Zeebroeck (2012) subsequently found that the internet fostered R&D collaborations within firms. In principle, one could make progress on this related question using patenting and coinvention data together with our empirical strategy.

[55] See https://www.vox.com/the-goods/2018/11/12/18080806/air-service-small-cities-crucial (accessed December 1, 2018).

[56] The goal of the constant B is to ensure that this ratio increases with the project's degree of specialization θ.

References

Adams JD, Black GC, Clemmons JR, Stephan PE (2005) Scientific teams and institutional collaborations: Evidence from US universities, 1981–1999. *Res. Policy* 34(3):259–285.

Agrawal A, Goldfarb A (2008) Restructuring research: Communication costs and the democratization of university innovation. *Amer. Econom. Rev.* 98(4):1578–1590.

Agrawal A, Catalini C, Goldfarb A (2015) Crowdfunding: Geography, social networks, and the timing of investment decisions. *J. Econom. Management Strategy* 24(2):253–274.

Bernstein S, Giroud X, Townsend RR (2016) The impact of venture capital monitoring. *J. Finance* 71(4):1591–1622.

Bikard M, Murray F, Gans S (2015) Exploring trade-offs in the organization of scientific work: Collaboration and scientific reward. *Management Sci.* 61(7):1473–1495.

Blum BS, Goldfarb A (2006) Does the internet defy the law of gravity? *J. Internat. Econom.* 70(2):384–405.

Borenstein S, Rose NL (2007) How airline markets work...or do they? Regulatory reform in the airline industry. NBER Working Paper 13452, National Bureau of Economic Research, Cambridge, MA.

Boudreau KJ, Guinan EC, Lakhani K, Riedl C (2016) Looking across and looking beyond the knowledge frontier: Intellectual distance, novelty, and resource allocation in science. *Management Sci.* 62(10):2765–2783.

Boudreau KJ, Brady T, Ganguli I, Gaule P, Guinan E, Hollenberg T, Lakhani K (2017) A field experiment on search costs and the formation of scientific collaborations. *Rev. Econom. Statist.* 99(4):565–576.

Cairncross F (1997) *The Death of Distance: How the Communications Revolution Will Change Our Lives* (Harvard Business School Press, Boston).

Campos R, de Leon F, McQuillin B (2018) Lost in the storm: The academic collaborations that went missing in Hurricane Isaac. *Econom. J.* 128(610):995–1018.

Catalini C (2018) Microgeography and the direction of inventive activity. *Management Sci.* 64(9):4348–4364.

Chai S, Freeman R (2018) Knowledge spillover and collaboration through temporary colocation. Working paper, ESSEC Business School, Cergy Pontoise, France.

Criscuolo P, Dahlander L, Grohsjean T, Salter A (2017) Evaluating novelty: The role of panels in the selection of R&D projects. *Acad. Management J.* 60(2):433–460.

Ding WW, Levin SG, Stephan PE, Winkler AE (2010) The impact of information technology on academic scientists' productivity and collaboration patterns. *Management Sci.* 56(9):1439–1461.

Dióspatonyi I, Horvai G, Braun T (2001) Publication speed in analytical chemistry journals. *J. Chemical Inform. Comput. Sci.* 41(6):1452–1456.

Falk-Krzesinski HJ, Contractor N, Fiore SM, Hall KL, Kane C, Keyton J, Thompson Klein J, Spring B, Stokols D, Trochim W (2011) Mapping a research agenda for the science of team science. *Res. Evaluation* 20(2):145–158.

Forman C, Zeebroeck NV (2012) From wires to partners: How the internet has fostered R&D collaborations within firms. *Management Sci.* 58(8):1549–1568.

Forman C, Goldfarb A, Greenstein S (2005) How did location affect adoption of the commercial Internet? Global village vs. urban leadership. *J. Urban Econom.* 58(3):389–420.

Freeman RB, Ganguli I, Murciano-Goroff R (2014) Why and wherefore of increased scientific collaboration. NBER Working Paper 19819, National Bureau of Economic Research, Cambridge, MA.

Friedman TL (2005) *The World Is Flat: A Brief History of the Twenty-First Century* (Farrar, Straus and Giroux, New York).

Furman JL, Teodoridis F (2017) The cost of research tools and the direction of innovation: Evidence from computer science and electrical engineering. Working paper, Boston University, Boston.

Gaspar J, Glaeser EL (1998) Information technology and the future of cities. *J. Urban Econom.* 43(1):136–156.

Goolsbee A, Syverson C (2008) How do incumbents respond to the threat of entry? Evidence from the major airlines. *Quart. J. Econom.* 123(4):1611–1633.

Hamermesh D, Oster S (2002) Tools or toys? The impact of high technology on scholarly productivity. *Econom. Inquiry* 40(4):539–555.

Jones BF (2009) The burden of knowledge and the death of the renaissance man: Is innovation getting harder? *Rev. Econom. Stud.* 76(1):283–317.

Jones BF, Wuchty S, Uzzi B (2008) Multi-university research teams: Shifting impact, geography and stratification in science. *Science* 322(5905):1259–1262.

Kim EH, Morse A, Zingales L (2009) Are elite universities losing their competitive edge? *J. Financial Econom.* 93(3):353–381.

Kremer M (1993) The O-ring theory of economic development. *Quart. J. Econom.* 108(3):551–575.

Leamer EE, Levinsohn J (1995) International trade theory: The evidence. Grossman GM, Rogoff K, eds. *Handbook of International Economics*, vol. 3 (North-Holland, Amsterdam), 1339–1394.

Milojević S (2014) Principles of scientific research team formation and evolution. *Proc. Natl. Acad. Sci. USA* 111(11):3984–3989.

Morrison SA (2001) Actual, adjacent, and potential competition: Estimating the full effect of Southwest Airlines. *J. Transportation Econom. Policy* 35(2):239–256.

National Academies (2004) *Facilitating Interdisciplinary Research* (National Academies Press, Washington, DC).

National Science Foundation (2015) Doctorate recipients from U.S. universities: 2014. Special Report NSF 16-300, National Science Foundation, Arlington, VA.

Perry MJ (2014) The cost of air travel in the US has been remarkably stable for the last decade, and 17% cheaper than 20 years ago. *Carpe Diem* (blog), February 12, https://www.aei.org/publication/the-cost-of-air-travel-in-the-us-has-been-remarkably-stable-for-the-last-decade-and-17-cheaper-than-20-years-ago/.

Polanyi M (1958) *Personal Knowledge: Toward a Post-Critical Philosophy* (University of Chicago Press, Chicago).

Rosenblatt T, Mobius M (2004) Getting closer or drifting apart? *Quart. J. Econom.* 119(3):971–1009.

Rosenthal SS, Strange WC (2001) The determinants of agglomeration. *J. Urban Econom.* 50(2):191–229.

Sauermann H, Haeussler C (2017) Authorship and contribution disclosures. *Sci. Adv.* 3(11):e1700404.

Stephan PE (2012) *How Economics Shapes Science* (Harvard University Press, Cambridge, MA).

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

Teodoridis F (2018) Understanding team knowledge production: The interrelated roles of technology and expertise. *Management Sci.* 64(8):3625–3648.

Transportation Research Board, National Research Council (1999) *Entry and Competition in the US Airline Industry: Issues and Opportunities* (National Academies Press, Washington, DC).

Von Hippel E (1994) "Sticky information" and the locus of problem solving: Implications for innovation. *Management Sci.* 40(4):429–439.

Walsh JP, Lee Y-N (2015) The bureaucratization of science. *Res. Policy* 44(8):1584–1600.

Wang J, Veugelers R, Stephan P (2017) Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Res. Policy* 46(8):1416–1436.

Wu L, Wang D, Evans JA (2019) Large teams develop and small teams disrupt science and technology. *Nature* 566(7744):378–382.

Wuchty S, Jones BF, Uzzi B (2007) The increasing dominance of teams in production of knowledge. *Science* 316(5827):1036–1039.