Why and Wherefore of Increased Scientific Collaboration

Richard B. Freeman, Ina Ganguli and Raviv Murciano-Goroff[1]

July 19, 2013

Fewer and fewer scientists write a scientific paper by themselves. Scientific collaborations with increasing numbers of authors on a paper have become the predominant form of scientific activity (Jones, Wuchty and Uzzi, 2008; Wuchty, Jones and Uzzi, 2007; Adams, Black, Clemmons, and Stephan, 2005). Seemingly justifying the trend in terms of productivity, the papers with larger numbers of authors garner more citations and publication in journals with higher impact factors (Lawani, 1986, Katz and Hicks, 1997; deB. Beaver, 2004, Wuchty et al., 2007; Freeman and Huang, 2013).

Similarly, fewer and fewer scientists coauthor papers solely with persons in their own country. International scientific collaborations, defined as those which produce papers with addresses from more than one country, have increased substantially (Indicators, 2012; Adams, 2013). In the US and other advanced economies, the proportion of papers with international coauthors has increased since the 1990s, while the proportion of papers with domestic coauthors has stabilized. By contrast, in emerging economies, where collaboration has not yet reached the proportions in the US and other advanced countries, the share of papers with domestic collaborations has increased along with international collaborations.

This paper combines data from the Web of Science (WoS) (Thomson Reuters, 2012) to examine the similarities and differences between international and domestic collaborations in three growing fields of science – Nanoscience and Nanotechnology, Biotechnology and Applied Microbiology, and Particle and Field Physics – and from a 2012 survey of corresponding authors on collaborations in those fields. The WoS data allows us to examine the patterns of collaborations over time and to compare our fields to those found in scientific publications

[1] We appreciate assistance with the survey from John Trumpbour and input from Paula Stephan and Andrew Wang. We received helpful comments from Adam Jaffe and Ben Jones and participants at the Changing Frontier pre-conference.

broadly. The survey allows us to investigate the connections among coauthors in collaborations and the views of corresponding authors about the collaboration.

1. Factors in Decisions to Collaborate

The main reason for scientists to collaborate rather than to undertake research on their own, or to collaborate with more rather than fewer people, is to combine expertise or equipment/resources to produce better research outcomes that are reflected in the numbers of papers or quality of papers, measured by citations or some related metric. Having more papers and citations enhances scientific careers, and thus provides strong incentives for collaborations. As science has become increasingly complex and specialized (Jones, 2010), the amount of knowledge that goes into creating new knowledge has grown, producing increased numbers of authors on papers and increased numbers of references within papers.

But there are also costs to collaborations. One set of costs relate to the production of the paper itself – problems of coordination of persons in different locations or with different expertise and viewpoints, of the expenses and difficulty of getting collaborators together or linking them with data and key pieces of machinery. Some costs of collaborations have fallen, which should increase the amount of collaborative research work. For instance, the cost of travel has generally trended downward and the advent of the Internet has reduced costs of communication (Agrawal and Goldfarb, 2008).

The spread of scientific workers and research and development activity around the world (Freeman, 2010) has also affected the pattern of collaborations. With an increasing number of PhDs in science and engineering graduating outside of the major research centers, the huge growth of higher education in developing countries, and increases in R&D spending in Asia, particularly China, it is possible to find collaborators in places far from the usual suspect research centers. The increased number of international students and post-docs returning to country of origins (Scellato, Franzoni, and Stephan, 2012) but with connections to researchers at the institution that educated them has created links for new collaborations. The easiest way to think of these developments is that, like lower cost of travel and communication, they have reduced the cost of collaborating with persons in foreign countries.

Finally, the location of scientific equipment and materials, such as the CERN Giant Hadron Collider, or special data sets or biological creatures, or geological or climatological data available only in certain parts of the world, has also affected collaborations. It is striking that while the US was not a prime funder for CERN, the largest group of scientists and engineers at CERN are Americans.

There is another less well studied phenomenon regarding collaborations that affects how scientists work together and presumptively the effectiveness of collaborations as well. This is in the division of credit for collaboration. In a one-author paper, the one takes credit or blame. In a two-author paper, the convention in many sciences is that the senior person's name comes last and the junior person comes first. Freeman and Huang (2013) find that the impact factor of the placement of the paper and citations of the paper depend more on the characteristics of the last named author, suggesting that this author has greater importance on average. In papers with more than two authors, many view the decision on who is the first author and the placement of the second, third, or other non-first or last authors as potentially affecting their career. With papers having huge numbers of names, gaining credit for one's work goes beyond the list of names on a publication.

How might these factors – productivity, cost, and allocation credit – affect decisions to collaborate and the outcomes from collaboration?

As a first step toward answering this question, we consider a scientist who at a point in time, chooses between doing research and publishing a paper in their name or working with others with whom they share credit for the work. For simplicity, we assume the goal is to maximize citations, which have an ultimate payoff in reputation, employment and earnings in the labor market. While today, decisions about collaborations are more about working with larger or smaller collaborations or choosing among alternative collaborations, we use the 'collaborate or not' decision to illuminate some basic issues in collaborations that will inform ensuing analysis.

First, we link the credit for doing research alone in a period to the number of papers produced and the citations each paper receives (over some future time period):

1) Citations = $\sum C_i$, where i goes from 1 to n for n papers written and C_i is the number of citations to the ith paper.

Linearity gives the scientist the same credit for a paper with, e.g. 10 citations as two papers with 5 citations, and so on. It is chosen for convenience.

On the other side, if the scientist writes a paper with someone else, each author is credited with:

2) Citations $=\sum \gamma C'i$, where the prime refers to a paper with 2 authors and $\gamma < 1$. Here i can go from 1 to N, with $N > n$ (you write more papers by sharing the work with the other author) or $N < n$ (your collaboration produces more cites to single paper).

The fact that $\gamma < 1$ implies that the collaboration pays off only if the number of citations goes up more than commensurately with γ, through either more papers ($N>n$) or through more citations to individual papers. To make matters easier, we will assume further that the number of papers written is the same whether someone collaborates or writes solo. This directs attention at the number of cites per paper, on which our data analysis focuses. In this case the collaboration pays off when:

3) $C'/C > 1/\gamma$ – the increased number of citations from the collaborative paper must exceed the loss due to the discounting of credit among authors.

This relation implies that papers with more authors should obtain greater citations, which factually they do, and suggests that the magnitude of the estimated relation between numbers of authors and citations provides information on the discount factor. In counting papers to authors, the natural metric is to fractionalize papers. If A writes 2 papers, each with two others, we credit A with 1 full paper. It is tempting to divide citations similarly – so that a 2-authored paper would have to gain twice the number of citations as a single-authored paper to be as valuable to the scientist. But the estimated effects of numbers of authors on citations falls so short of proportionality, that it is clear γ is not proportional to numbers of authors.

The costs of collaboration also give us some insight into the likely relation between collaborations and the output of collaborations. On the innocuous assumption that it is more costly to collaborate with individuals far away, we get the prediction that collaborations with non-colocated coauthors should produce better papers – measured as citations in our model. The notion is that the authors bear the extra cost in a collaboration only if the collaboration produces value that exceeds extra costs. Put differently, you seek a distant coauthor if they offer something your neighbor doesn't have. The prediction is thus that non-colocated teams of researchers – those working in different countries or cities in the US – produce 'better' science. Moreover, to the extent that this science makes greater use of earlier scientific knowledge, we would expect those papers to have greater references (which we test) or tables, figures, or formula than papers written by authors in the same locale (which is testable but not with our data).

2. Collaboration in Web of Science Publications

We analyze data from publications and from a survey of corresponding authors on collaborations in three growing fields of science – Nanoscience and Nanotechnology, Biotechnology and Applied Microbiology, and Particle and Field Physics. We have selected our fields to cover a range of rapidly growing scientific areas with different knowledge production functions, and where changes in labor supplies and capital equipment are likely to affect the level and trend of collaborations between US-based researchers and overseas researchers.

Nanotechnology is a general interdisciplinary applied technology, where engineers often collaborate with material scientists and where the electron microscope is a pivotal research tool. The US and other countries have made sizable investments in nanotechnology from at least the early 2000s, when President Clinton called for greater investment in nano-related science and technology, which eventually produced the 21st Century Nanotechnology Research and Development Act that President Bush signed in 2003. Other countries undertook similar initiatives.

Biotechnology is lab-based, where life scientists largely funded by NIH dominate research and where researchers often have close links to big pharma firms. The most important change in research technology has been the US-sponsored Human Genome Project, the result of international efforts, which now provides data freely on the web.

Particle physics has two parts – theoretical and experimental, with much leading edge research requiring massive investments in accelerators and colliders. The recent development of the Large Hadron Collider (LHC) has changed the geographic locus of empirical research while the US decision to forgo the proposed large collider in Texas arguably spurred the greater growth of string theory in the US than in Europe.

We begin with an analysis of collaboration patterns in publication data from the Web of Science (WoS). We identified all papers in the WoS database published from 1990-2010 with a US coauthor with journal subject categories of *Nanoscience & Nanotechnology, Biotechnology & Applied Microbiology, Physics, Particles & Field*s. From these papers, we identify teams by the names of coauthors, and we identify the location of the authors using the author affiliations. This sample includes 115,027 papers. Using the location of the authors on each paper, we then define the following types of collaborations:

US-Only Collaborations, divided into those in which all US authors are in the same city (US Colocated) and those in which US coauthors are in at least two different cities (US Non-Colocated).

International Collaborations, divided into those in which US coauthors are in the same city with at least one foreign coauthor (US Colocated/International) and those in which US coauthors are in two or more cities with at least one foreign coauthor (US Non-Colocated/International).

By distinguishing between US collaborations that occur in different locations and those that occur in the same city, we are better able to identify the impact of international collaborations per se on papers as opposed to the impact of collaborations at different locations on papers, be they in the US or overseas.

Figure 1 records measures of the geographic and national composition of collaborations in Nanoscience and Nanotechnology, Biotechnology, and Particle/Field Physics taken together. The top line gives the share of papers in which a US-based author collaborates solely with authors colocated in the same city, which makes it easy for them to meet face to face. It shows a marked decrease in collaborations between persons at the same locale. The next line gives the share of papers that include at least one person located in another country – the international collaborations that have attracted great attention. It shows a twelve percentage point increase in the proportion of international collaborations from 1990 to 2010. The third line is for collaborations across locations in the US. It also shows an increase in the proportion of papers, which implies that the increased geographic scope of collaborations involved more than cross national boundaries, though the increase is less than for international collaborations. The bottom line is the share of papers for solo authors, which fluctuates from 5% to 10% in the period studied.

Figure 2 examines the proportion of papers with an international collaboration for the three fields separately. The data for particle physics in Figure 2a displays the highest level of international collaborations, due presumably to the importance of particle accelerators and other equipment that are available at only some sites. Figure 2b and Figure 2c show that in nano and in biotech, the most common form of collaborations are US colocated teams, while international collaborations are second most common and collaborations across locations in the US third in frequency. US non-colocated collaborations are roughly as common in both nano and biotech as international collaborations until the late 2000s, when international collaborations rise sharply. The pattern of increase in international collaborations resembles the patterns reported in National Science Board (2012) and in Adams (2013) in science more broadly. In all of the fields, the proportion of papers coauthored by people in the same city falls. Though our sample covers just three fields, the results are representative of the increased geographic dispersion of coauthorship within the US as well as between Americans and persons working in other countries.

In what ways, if any, do papers with international collaborations differ from those in which collaborations occur solely in the US?

Katz and Hicks (1997), Rigby (2009) and Adams (2013) note that international collaborations tend to be more highly cited than collaborations limited to persons in a single country. Our data shows a similar pattern when the US is the single country. For papers published in 1990-2000 (the dates chosen so as to allow time for papers to gain citations), US papers with foreign authors obtain about 1 more citation in our three fields compared to US collaborations with fellow residents of the US (26.59 vs. 25.65). Since the US is the leading scientific center, with US-authored papers averaging a greater number of citations than papers worldwide, it would have been reasonable for the data to show the opposite: fewer citations for US-based scientists who collaborate with persons outside the country than for US-based scientists who collaborate with other US scientists.[2]

Does this mean that international collaborations produce better science, as the headline to Adams' 2013 paper claims, or at least to more widely recognized work?

Our distinction between collaborations among persons in the US in the same city and in different locations provides one way to examine this question. It allows us to determine whether the higher citation count for international collaborations reflects the impact of the work of persons in multiple locales, international or domestic, as opposed to the distinct impact of an overseas collaboration. To compare these situations, we calculated the average citations for collaborations involving non-colocated authors in the US as well as the average citations for colocated authors in the US and for international collaborations. Figure 3 shows the results from

[2] The results of Lee, et al. (2010) that shows that among papers published by Harvard biomedical researchers from 1993-2003, geographic distance between first-last authors is associated with fewer citations, give a very different picture than ours or the other researchers who find higher citations over greater distances. There is no inherent inconsistency because of the difference in the scales of distance. All Harvard biomedical researchers would be in the same city in our analysis and same country in Adams (2013). But the difference in the patterns at different scales deserves clarification.

this analysis for the year in which a paper was published beginning with publications in 2007 (with three years of potential citations) and going back to those published in 1990 (with 21 years of potential citations). Papers written by authors in the same city have the fewest citations. But the number of citations to the papers coauthored by people in different cities in the US is roughly comparable to the number of citations to international collaborations. This suggests that the greater cites of international collaborations reflects multiple locations more than cross-national borders.

Table 1 highlights two other differences between international and domestic collaboration that may account for the high number of citations on international collaborations. The first difference is in the number of coauthors on a paper – a factor that researchers have found positively associated with citations (and with the impact factor of the journal of publication or other such metrics). Column 1 records the regression coefficients and standard errors from a regression of the number of coauthors on a dummy variable for whether or not the paper had an international coauthor. The positive coefficient says that papers with international coauthors had 15 more coauthors on average. Column 2 replaces the international collaboration dummy with more detailed measures of the collaboration: whether the collaboration is among US persons in different cities with no authors outside the country; whether the collaboration is international with US authors in the same city; and whether it is international with US authors in different locations. The reference group in the analysis is papers written by authors in the same US city. Each of the regressions includes dummy variables for field and a time trend.

The estimated coefficients in column 2 show that collaborations across US cities has a comparable effect on number of coauthors as collaborations between US scientists in a single city with scientists in another country. The largest number of coauthors occurs when the paper has both a US non-colocated author and a foreign-based author, in part for arithmetic necessity: such papers must have at least 3 authors whereas the other collaborations can be found in 2-authored papers. But this is not the whole story. Limiting our sample to papers with three or more authors, we still find that papers in which there is an international coauthor and US authors still have more coauthors than those without an international coauthor, and that those with US coauthors in two or more cities and at least one foreign coauthor have by far the most coauthors. These papers are almost surely produced under different technologies than papers with 2 or 3 or 4 or 5 authors.

Columns 3 - 5 turn to the relation between the type of collaboration and another variable that is positively related to citations (and impact factors) – the number of references in a paper (Freeman and Huang, 2013). The regression in column 3 shows that international papers have more references than papers written by authors solely in the US; while the regression in column 4 shows that this relation holds with the addition of the number of coauthors on the paper. The remaining column gives results for the disaggregated measures of collaboration. The coefficient on US papers with non-colocated authors in the US adds about 1.1 additional references to the number of references in the base group of papers written with US colocated researchers. The number of references on papers written with authors from the same US location and authors overseas add 1.6 additional references over papers with authors from the same city. But the biggest bang comes from collaborations between persons in different US locales and overseas – a gain of 3.3 references relative to the base group.

How far, if at all, does disaggregation of US only papers between those written by authors in the same city and those written in different cities and the number of authors/references involved in international papers and other collaborations explain the higher citation rate of international papers?

To answer this question we used a regression model to examine the relation between the citations received by a paper compared to others in the same field and year and the nature of the collaboration that produced the paper. Since citations have a life cycle – with the number of citations increasing sharply in the first 5 to 7 years after publication and then tapering off – we include dummy variables for the year of publication in the analysis.

Column 1 of Table 2 documents the critical fact that international papers gain more citations than papers with only US collaborations, although the standard error is large. Column 2 shows that about half of the effect of international collaborations on citations comes about by having a larger number of authors. Column 3 shows that about 90% comes about from having more references. And column 4 shows that conditional on both number of authors and number of references, international collaborations have slightly fewer citations than US-only collaborations. The similarity in numbers of citations among papers with the same number of authors and different locations indicates that there is nothing special about the number of citations given to international collaborations. The number of coauthors/references and reliance on researchers outside a single locale are the pathway for the international collaboration effect.

Column 5 of Table 2 expands the international/domestic collaboration distinction to cover within US collaborations as well. It shows that the largest number of citations are to papers with authors from multiple locations in the US and with authors in international locations as well. As noted, many of these papers have sufficiently large numbers of authors to differ substantively in their technology from papers with small numbers of authors. While it is possible to learn more about the non-linearities in these relationships from the WoS data, we have chosen a different route to gain insight into what goes on in collaborations and its link to scientific outcomes. This is the development of our own survey of corresponding authors about their collaboration.

5. Survey Evidence on Collaborations

"*I think the best example of collaboration I have done is …where all the authors are from different countries and we met at the Bellagio Conference Center of the Rockefeller Foundation.*"

"*I think that it is absolutely indispensable to meet people in person to have effective collaborations.*"

"*Skype was not available…at the time we completed this work. We now use Skype or ITV connection to meet and discuss data with collaborators on a weekly basis.*"

"The international collaboration worked so well because of my frequent trips to Brazil during the project." (Comments from the open-ended section of our survey)

In 2012 we conducted an online survey of the authors of papers published in 2004, 2007, and 2010 and appearing in the Web of Science database. We selected all publications with two or more authored papers in the Nano, Biotech, and Particle Physics subject categories. We identified all unique authors within this sample of publications and randomly selected one paper if the author had more than one paper appearing in the database in 2004, 2007 and 2010. Using the email address of the corresponding author, in July 2012 we sent a personalized email in English to each corresponding author, inviting them to complete the survey by responding to a link that connected them to the online survey instrument. If there was more than one corresponding author, we selected the one that appeared first. We sent 2 follow-up email reminders in August and September 2012. We developed the survey using Qualtrics Survey Software and respondents accessed it from the Qualtrics server.

We customized each survey to ask the respondent about the specific collaboration and individual team members. The survey had 25 questions and was designed so that respondents would complete it in approximately 10-15 minutes. The questions related to how the team formed, the way the team communicated and interacted during the collaboration, the contribution of each coauthor, types of funding received, and the advantages and disadvantages of working with the team. The survey also included an open-ended question at the end for respondents to make comments. Several respondents sent emails with additional thoughts and information about the collaboration.

We sent the survey to approximately 19,000 individuals. The actual number of individuals who received the email is lower, since some email addresses had expired, changed, or some individuals were deceased. We received 3,920 responses, which implies a response rate of approximately 20%, which is in line with other surveys of scientists (Sauermann and Roach, 2013). However, this underestimates the response rate of the emails that reached respondents.[3] For individuals who had published in the most recent year (2010), we had a better response rate of 26%. Our final sample includes 3,452 respondents, as we dropped several responses that failed to meet our sample requirements of having at least one US author.

Table 3 provides a summary of our main data samples and contrasts the distribution of the characteristics of the papers on which we obtained survey responses to the distribution of the characteristics of all papers in 2004, 2007 and 2010 papers in the 3 fields in the WoS database. The column giving the difference between the distribution of our sample shows that our survey sample is slightly overrepresented by US colocated teams, the more recent publication year (2010), and publications from biotechnology.

The survey has one distinct virtue over WoS location data to identify international collaborations. We obtain a more accurate measure of whether teams are international, since we ask the respondent which country each coauthor was *"primarily based in during the research and writing"* of the article. Most studies define international teams based only on author affiliations, which can produce errors if affiliations change between the time the research was undertaken and the time of publication or because some people have affiliations from more than one country. Comparing the row "Int'l Collaboration Survey", which is based on the

[3] We are estimating the number of emails that actually got to potential respondents to obtain an alternative response rate.

respondent's answers regarding the location of coauthors, and "Int'l Collaboration" in Table 1 shows that using only author affiliations overestimates the number of international teams by about 4 percentage points. Our analysis of the survey data uses the respondents' information to define US colocated, US non-colocated and international teams.

Figure 4 displays the proportion of persons of each collaboration type who the corresponding author first met through the five specified mechanisms. Most of the first meetings occurred when the corresponding author and the other person worked in the same institution. For papers written in the same location, the predominant contact came through advisor-student or postdoc relationships, but over one third of the meetings came about as colleagues. For papers with authors from other US locations or foreign locations, the corresponding author met them through working in the same place, primarily working as a colleague. But nearly ten to sixteen percent met the person as a visitor. Conferences accounted for a large proportion of the first meetings between corresponding authors on papers written with persons in other locations or in foreign locations.

Overall, the most striking pattern in Figure 4 is the similarity between modes of meeting among non-colocated US authors and US and foreign-located authors, which contrasts with the mode of meeting coauthors in colocated collaborations. The time series data in Appendix Figure A and B show that conferences have become a less important way to meet future coauthors, while students/postdocs have become more important, possibly due to their increased importance in the scientific production process.

Seeking to contrast Internet communication and face-to-face meetings in making collaborations work, we asked corresponding authors the frequency with which they communicated with one or more of their coauthors from "every week" to "never". Because papers that extend beyond a local area allow for the corresponding author to meet face-to-face with coauthors in their locale but only infrequently in other areas, the question does not pin down differences associated with distance as well as it might. To overcome this problem, we contrast the frequency of communication between two-authored papers with colocated authors with the frequency of communication on two-authored papers with non-colocated US and foreign coauthors.

The results in Figure 5 show that the corresponding author relies extensively on face-to-face meetings when all authors are in the same location, and are much lower but still important in communicating with authors across distances. Among 2-author papers, for example, just over 50% of corresponding authors on international teams report meeting face-to-face at least a few times per year, while 64% of those on US non-colocated papers reported face-to-face meetings at least a few times a year. By contrast there are no noticeable differences in using e-mail: corresponding authors in all forms of collaborations use them frequently, approximately 40 weeks during the year. The differences in use of telephone vs. Internet (e.g. Skype) between US-based teams and international teams are readily explained by cost differences.

To understand some of the factors that underlie the collaborations, we asked the corresponding author about what each team member contributed to the research. The major factor cited for all types of collaborations was “unique knowledge, expertise, capabilities” (Figure 6). Non-colocated and international teams were more likely to have a coauthor contributing data, material or components – a pattern that has been increasing over time, according to Appendix Figure C. Looking more closely at the specific contributions of particular team members, Figure 7 compares the contribution of foreign-located coauthors and domestic coauthors on two-authored papers. The distribution of authors across the contribution types is similar for foreign and US coauthors.

To get an overall assessment of the effects of the different forms of collaboration on the production and output of scientific activity, we asked the corresponding authors their views of the particular collaboration in our survey and then related some of the attributes of the collaboration to the citations given to the papers.

Table 4 summarizes the responses of corresponding authors on the advantages and challenges of the collaborations, organized by type of collaboration. It records the average score on a five-point scale of agreement or disagreement with a set of statements that reflect the attributes of the collaboration. The teams were in substantial agreement that the collaboration had substantial advantages in harnessing human capital to produce a scientific outcome. “Complementing our knowledge, expertise and capabilities” and “learning from each other” are the only two items with average scores greater than 4 in the table, with little variation among the types of collaborations: the US non-colocated and international teams giving slightly higher scores than the co-located teams. Similarly, the groups ranked highly “Gaining access to data, material or components”, though here the highest assessment came from the corresponding authors of US-non-collated teams. Another area of difference was in the higher score given by the corresponding authors of international teams and to lesser extent of US-non-colocated teams of the advantage of having "Our research reached a wider audience" compared to US-only teams. Viewing "wider audience" in terms of the geographic distribution of citations, this suggests that the wider geographic distribution of authors, the wider is the distribution of citations, possibly even among papers with the same numbers of citations – a pattern which can be investigated further in the WoS data.

As for the challenges, non-colocated and international teams tended to agree more that there was "Insufficient time for communication", "Problems coordinating with team members' schedules", and "Insufficient time to use a critical instrument, facility or infrastructure", but international teams did not report any greater problems in this regard than national non-colocated teams. Consistent with some of our earlier WoS results, geography would appear to be more than national boundaries in the way teams operated.

We also asked whether the teams were of appropriate size. The responses, given in Appendix Table A, show that most corresponding authors viewed their team as having the right size – presumably the principal investigator(s) would have modified the team if they did not think that was the case. But there are some differences by collaboration type. US colocated teams were more likely to say that they needed additional collaborator (7.58% vs. 3.48% and 3.38% for US non-colocated and international); whereas international teams were more likely to say that fewer team members were needed (6.67% vs. 3.37% for US colocated). Twenty four percent of the international teams received funding specifically aimed at supporting cross-country collaboration, with 6.65% receiving US government funding, 4.64% receiving EU funding, and the remainder from other government sources. These teams were slightly more likely to report that their team size was optimal.

Finally, we combine our survey data with information on the surveyed paper from the WoS to examine how the views of corresponding authors relate to the citations received by the paper. We do this by adding to the regressions of the number of citations on the characteristics of a paper in Table 2 survey-related information of how the authors met, the coauthor contribution, mode of funding, and other variables, with covariates for the field, year, and the nature of the collaboration that produced the paper.

Because publication of the paper preceded the survey, we cannot determine the causality of some of the corresponding author views of the collaboration outcomes. Ideally, corresponding authors would give the same assessment of the collaboration regardless of the ensuing impact of the paper but we cannot rule out that the response to the paper affected how the corresponding authors viewed the collaboration afterward.

Table 5 gives the results for survey responses that seem least prone to the dual causality problem: the way corresponding authors met coauthors; the funding support; and what coauthors contributed. Of these three areas, the coauthor contribution question is probably the most prone to being colored by the outcome of the paper. Columns 1 and 2 replicate the regression estimates in Table 2 for the dichotomous international collaboration variable. They show that the basic pattern found in the larger WoS data set is mirrored in the smaller survey paper data set. The coefficient on international collaborations has a positive effect on citations in column 1 that is larger than the coefficient in the comparable regression in Table 2. The coefficients on the number of coauthors and number of references variables are positive and significant in column 2 but with the impact of coauthors larger than that of references in this regression, contrary to the result in Table 2.

The regressions with the survey variables show that papers in which at least one coauthor met at a conference had higher citations; that papers for which a coauthor contributed funding had lower citations, and that papers that got funding specifically for cross-country collaborations had lower citations. We also estimated the model including dummies for whether the corresponding author didn't view the team size as optimal, and an average of the scores assessing the advantages and disadvantages to the collaboration, but found no effect of these measures on citations.

6. Conclusion: The Geographic Spread of Scientific Collaborations

This examination of scientific collaborations has found that US collaborations have increased not only with scientists in other countries, but also across US cities; that the nature of collaborations across cities resembles that across countries; and that the higher citation rate of international collaborations operates through similar pathways as the higher citation rate given to papers with non-colocated authors within the US rather than any “international magic.” While there may be some special attributes to international collaborations, the best framework for examining them is that associated with the more general issue of collaborations across space.

References

Adams, J. (2013). “Collaborations: The Fourth Age of Research.” *Nature*, 497(7451), 557-560.

Adams, J. D., Black, G. C., Clemmons, J. R., & Stephan, P. E. (2005). “Scientific Teams And Institutional Collaborations: Evidence From US Universities, 1981–1999.” *Research Policy,* 34(3), 259-285.

Barrantes Bárbara S. Lancho, Vicente P. Guerrero Bote, Zaida Chinchilla Rodríguez, Félix de Moya Anegón (2012). “Citation Flows in the Zones of Influence of Scientific Collaborations” *Journal of the American Society for Information Science and Technology*, Volume 63, Issue 3, pages 481–489, March 2012.

deB. Beaver, Donald (2004). “Does Collaborative Research Have Greater Epistemic Authority?” *Scientometrics*, 60(3), 399-408.

Freeman, Richard B. 2010 “Globalization of Scientific And Engineering Talent: International Mobility of Students, Workers, and Ideas and The World Economy.” *Economics Of Innovation And New Technology*, Volume 19, issue 5, 201 pp. 393-406.

Freeman, R.B. and Huang, W. (2013). “Collaborating With People Like Me: Ethnic Co-authorship within the US.” Harvard mimeo.

Guerrero Bote, Vicente P. Carlos Olmeda-Gómez Félix de Moya-Anegón (2013). “Quantifying the Benefits of International Scientific Collaboration,” *Journal of the American Society for Information Science and Technology*, Volume 64, Issue 2, pages 392–404, February 2013

Hsu, J.W., & Huang, D.W. (2011). “Correlation Between Impact and Collaboration.” *Scientometrics*, 86(2), 317–324.

Jones, B. F., Wuchty, S., & Uzzi, B. (2008). “Multi-University Research Teams: Shifting Impact, Geography, And Stratification In Science.” *Science*, 322(5905), 1259-1262.

Jones, Benjamin (2010). “As Science Evolves, How Can Science Policy?” NBER Working Paper No. 16002 Issued in May 2010.

Katz, J.S. And D.Hicks (1997). “How Much Is a Collaboration Worth? A Calibrated Bibliometric Model,” *Scientometrics,* 40:3, 541-554.

Lawani, S. M. (1986). “Some Bibliometric Correlates Of Quality In Scientific Research,” *Scientometrics,* 9:1-1 13-25.

Lee K, Brownstein JS, Mills RG, Kohane IS (2010). “Does Collocation Inform the Impact of Collaboration?” PLoS ONE 5(12): e14279. doi:10.1371/journal.pone.0014279

National Science Board (2012). Science and Engineering Indicators, 2012.

Rigby, John (2009). "Comparing the Scientific Quality Achieved By Funding Instruments For Single Grant Holders And For Collaborative Networks Within A Research System: Some Observations," *Scientometrics* 78:1, 145-164.

Sauermann, H., & Roach, M. (2013). "Increasing Web Survey Response Rates In Innovation Research: An Experimental Study Of Static And Dynamic Contact Design Features." *Research Policy,* 42(1), 273-286.

Scellato, G., Franzoni, C., & Stephan, P. (2012). "Mobile Scientists and International Networks," NBER Working Paper 18613.

Thomson Reuters (2012). Web of Science ®, prepared by THOMSON REUTERS ®, Inc. (Thomson®), Philadelphia, Pennsylvania, USA: © Copyright THOMSON REUTERS ® 2012. All rights reserved.

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). "The Increasing Dominance Of Teams In Production Of Knowledge." *Science*, 316(5827), 1036-1039.

TABLES

Table 1: Estimated Relation Between Number of Coauthors and Number of References on Papers by Nature of Collaboration

	Number of Coauthors		Number of References		
	(1)	(2)	(3)	(4)	(5)
US Collaboration Only					
US Colocated					
US Non-Colocated		0.945^{**}			1.091^{**}
		(0.067)			(0.141)
Int'l Collaboration	15.093^{**}		1.668^{**}	1.608^{**}	
	(0.304)		(0.124)	(0.125)	
US Colocated/Int'l		2.242^{**}			1.565^{**}
		(0.215)			(0.143)
US Non-Col./Int'l		50.295^{**}			3.320^{**}
		(1.115)			(0.218)
Number of Coauthors				0.004^{**}	
				(0.001)	
Particle Physics	19.222^{**}	15.680^{**}	6.229^{**}	6.152^{**}	6.042^{**}
	(0.479)	(0.397)	(0.170)	(0.173)	(0.171)
Biotechnology	0.641^{**}	0.068	7.408^{**}	7.405^{**}	7.379^{**}
	(0.084)	(0.105)	(0.124)	(0.124)	(0.124)
Year Trend	-0.061^{+}	-0.084^{*}	0.871^{**}	0.871^{**}	0.866^{**}
	(0.035)	(0.034)	(0.013)	(0.013)	(0.013)
Constant	122.530^{+}	169.298^{*}	$-1.7e+03^{**}$	$-1.7e+03^{**}$	$-1.7e+03^{**}$
	(70.554)	(67.554)	(25.779)	(25.764)	(25.777)
R2	0.062	0.129	0.059	0.059	0.060
Nb. of Obs.	115,027	115,027	115,027	115,027	115,027

Notes: + p < 0.10, * p < 0.05, ** p < 0.01, OLS estimation. Includes all papers in the Web of Science with a US coauthor with journal subject categories of Nanoscience & Nanotechnology, Biotechnology & Applied Microbiology, Physics, Particles & Fields.

Table 2: The Estimated Relation Between Number of Citations to a Paper and the Type of Collaboration That Produced the Paper

	(1)	(2)	(3)	(4)	(5)
US Collaboration Only					
US Colocated					
US Non-Colocated					1.368**
					(0.319)
Int'l Collaboration	0.535	0.261	0.062	-0.196	
	(0.339)	(0.340)	(0.340)	(0.340)	
US Colocated/Int'l					-0.783*
					(0.348)
US Non-Col./Int'l					3.447**
					(0.787)
No. Coauthors		0.018**		0.017**	0.011**
		(0.003)		(0.003)	(0.003)
No. References			0.286**	0.285**	0.283**
			(0.017)	(0.017)	(0.017)
Particle Physics	-0.681*	-1.030**	-2.508**	-2.835**	-3.080**
	(0.316)	(0.311)	(0.345)	(0.337)	(0.339)
Biotechnology	0.653*	0.638*	-1.558**	-1.570**	-1.605**
	(0.285)	(0.285)	(0.341)	(0.340)	(0.339)
Constant	28.039**	28.137**	23.391**	23.490**	23.245**
	(1.731)	(1.725)	(1.703)	(1.697)	(1.700)
Year FE	Yes	Yes	Yes	Yes	Yes
R2	0.024	0.024	0.035	0.036	0.036
Nb. of Obs.	115,027	115,027	115,027	115,027	115,027

Notes: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, OLS estimation. Includes all papers in the Web of Science with a US coauthor with journal subject categories of Nanoscience & Nanotechnology, Biotechnology & Applied Microbiology, Physics, Particles & Fields.

Table 3. Distribution of papers by characteristics, Web of Science Papers and Survey Respondents

	(1) Papers, 1990-2010	(2) Papers in 2004, 2007, 2010	(3) Survey Sample	(3)-(2) Difference
Collaboration Type				
US Collaboration Only	65.08	63.65	62.25	-1.4
US Colocated	43.38	41.55	46.84	5.29
US Non-Colocated	21.7	22.1	15.41	-6.69
Int'l Collaboration	34.92	36.35	37.75	1.4
US Colocated/Int'l	24.82	26.04	26.94	0.9
US Non-Col./Int'l	10.1	10.31	10.81	0.5
Int'l Collaboration Survey			34.01	
Year				
2004	6.64	25.36	18.42	-6.94
2007	8.8	33.6	29.46	-4.14
2010	10.75	41.04	52.11	11.07
Field				
Nano	25.88	32.87	30.5	-2.37
Particle Physics	26.71	21.75	19.55	-2.2
Biotechnology	47.41	45.38	49.94	4.56
N	115,027	31,187	3,452	

Notes: (1) includes all papers in the Web of Science with a US coauthor with journal subject categories of Nanoscience & Nanotechnology, Biotechnology & Applied Microbiology, Physics, Particles & Fields 1990-2010. (2) includes those papers in 2004, 2007, and 2010. (3) includes the respondents to our survey, which was a sample based on unique corresponding authors appearing in (2) that had more than 1 author.

Table 4. Advantages and Challenges to Working with the Team

	US Colocated	US Non-Colocated	Int'l
Advantages			
Learning from each other	4.26	4.33	4.36
Complementing our knowledge, expertise and capabilities	4.39	4.58	4.57
Gaining access to data, materials or components	3.21	3.56	3.32
Gaining access to data, materials or components protected by IP	2.14	2.30	2.29
Our research reached a wider audience	3.24	3.37	3.48
Challenges			
Insufficient time for communication	1.82	2.13	2.11
Less flexibility in how the research was carried out	1.73	1.99	1.93
Unable to unequivocally portray my contribution	1.55	1.59	1.65
Problems coordinating with team members' schedules	1.96	2.18	2.11
Insufficient time to use a critical instrument, facility or infrastructure	1.45	1.67	1.67
Observations	1,693	585	1,174

Notes: 5 = Agree, 1 = Disagree

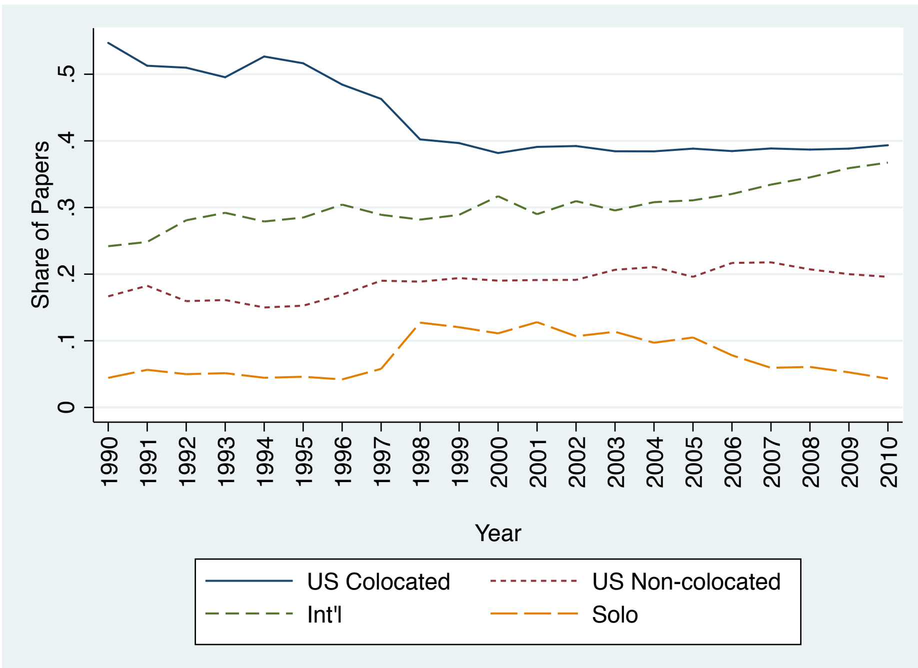
Table 5: The Estimated Relation Between Number of Citations to a Paper and Survey Variables

	(1)	(2)	(3)	(4)	(5)
US Noncolocated	0.915	0.454	-0.348	0.444	0.459
	(0.784)	(0.765)	(0.776)	(0.771)	(0.766)
Int'l Collaboration	1.145*	0.349	-0.547	0.406	0.535
	(0.545)	(0.562)	(0.652)	(0.597)	(0.582)
No. Coauthors		0.155*	0.152*	0.155*	0.156*
		(0.063)	(0.063)	(0.063)	(0.063)
No. References		0.100**	0.099**	0.099**	0.099**
		(0.015)	(0.014)	(0.015)	(0.015)
How They Met					
Advisor-Stu./Postdoc			-0.710		
			(0.659)		
Colleagues			0.629		
			(0.543)		
Visiting			0.744		
			(0.888)		
Conference			2.914**		
			(0.991)		
No introduction			0.557		
			(0.890)		
Coauthor Contributions					
Knowledge, etc.				0.474	
				(0.677)	
Funding				-1.308*	
				(0.553)	
Data, etc.				-0.293	
				(0.522)	
IP Data, etc.				0.058	
				(0.635)	
Instrument, etc.				0.221	
				(0.571)	
Cross-country funding					-1.204*
					(0.611)
Constant	16.283**	13.092**	13.272**	13.471**	13.234**
	(1.137)	(1.206)	(1.312)	(1.276)	(1.219)
Year FE	Yes	Yes	Yes	Yes	Yes
R2	0.074	0.111	0.115	0.112	0.112
Nb. of Obs.	3,452	3,452	3,452	3,452	3,452

Notes: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, OLS estimation. Includes all papers in the Web of Science with a US coauthor with journal subject categories of Nanoscience & Nanotechnology, Biotechnology & Applied Microbiology, Physics, Particles & Fields. "How They Met" and "Coauthor Controbution" variables are dummies indicating whether any coauthor on the team met that way/contributed the resource.

FIGURES

Figure 1. Share of Papers by Collaboration Type



Notes: Includes papers with a US coauthor with journal subject categories of Nanoscience & Nanotechnology, Biotechnology & Applied Microbiology, Physics, Particles & Fields.

Figure 2a: Share of Papers by Collaboration Type, Particle Physics

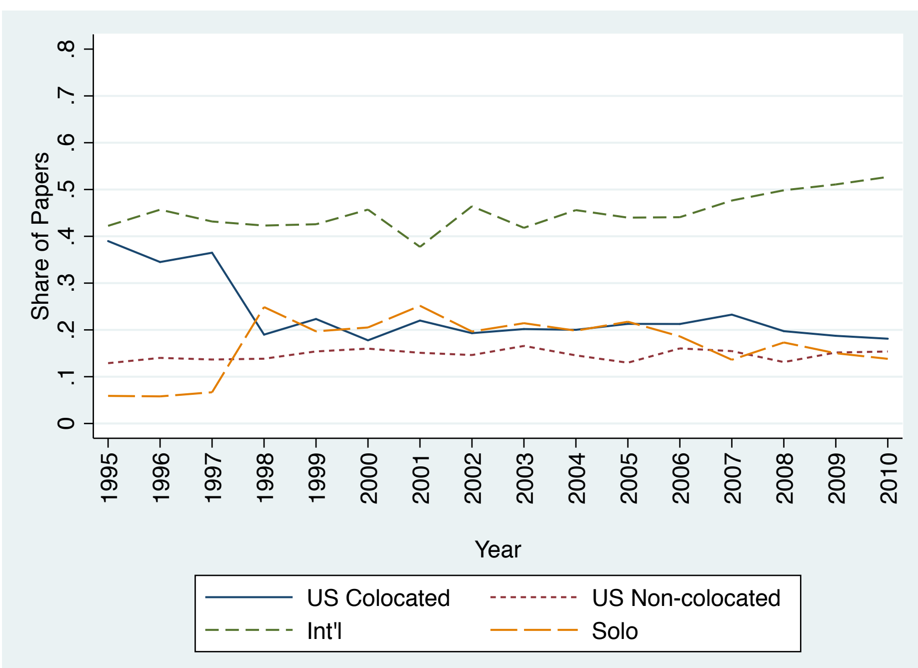


Figure 2b: Share of Papers by Collaboration Type, Nano

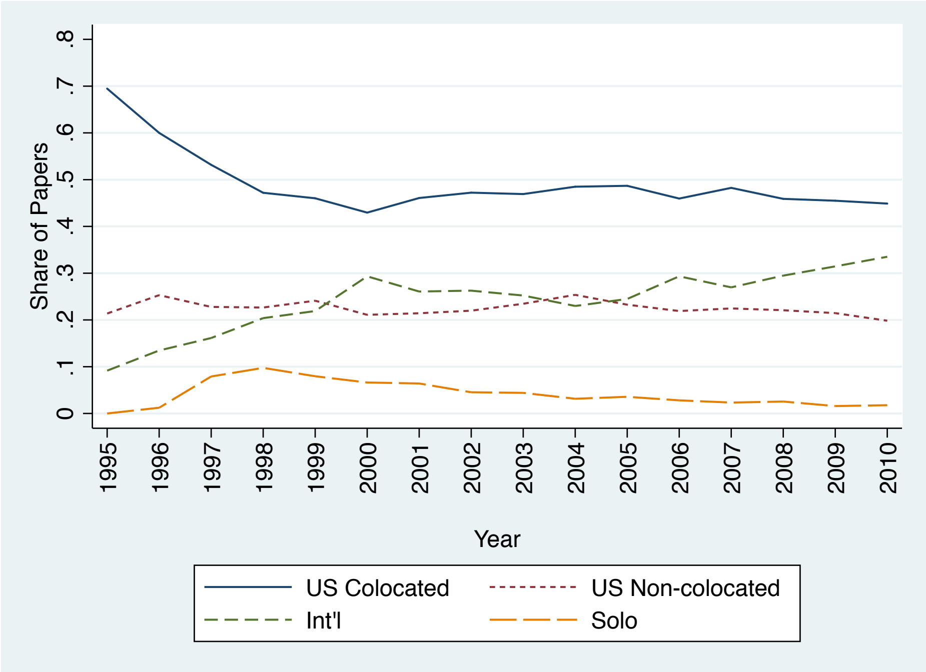


Figure 2c: Share of Papers by Collaboration Type, Biotech

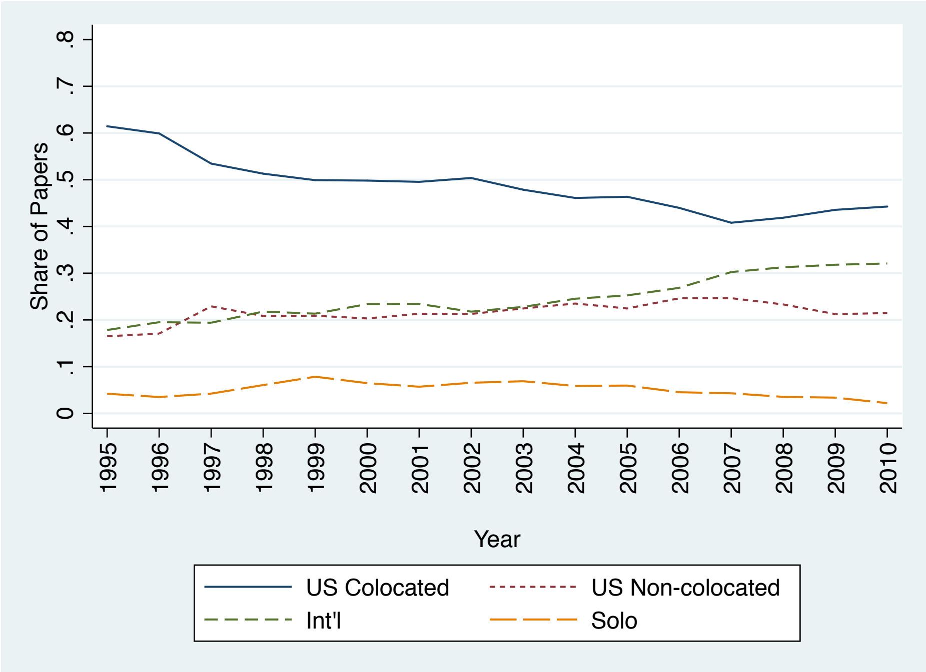


Figure 3. Citations By the Nature of Collaboration, All Fields by Year of Publication

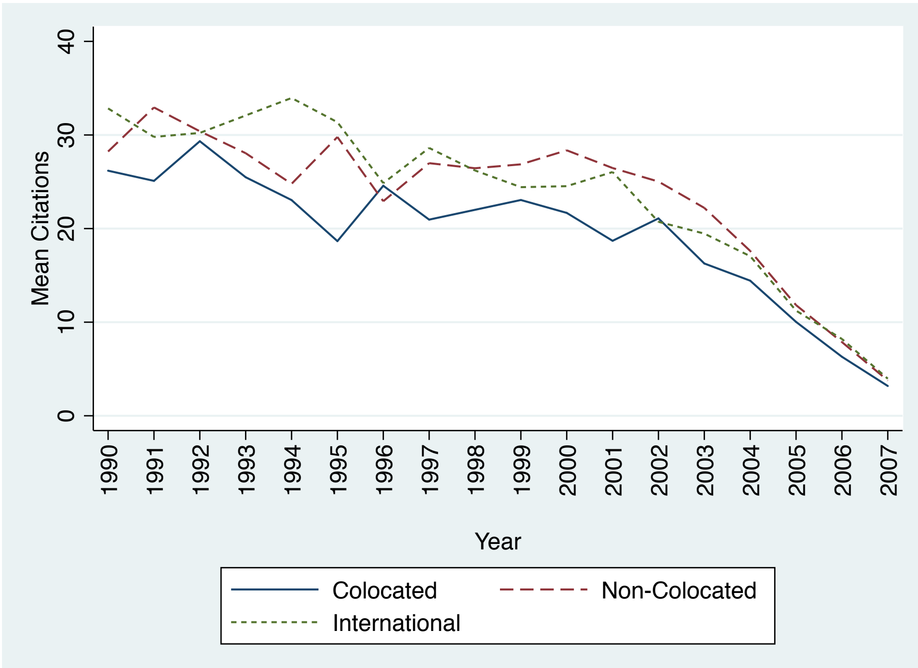
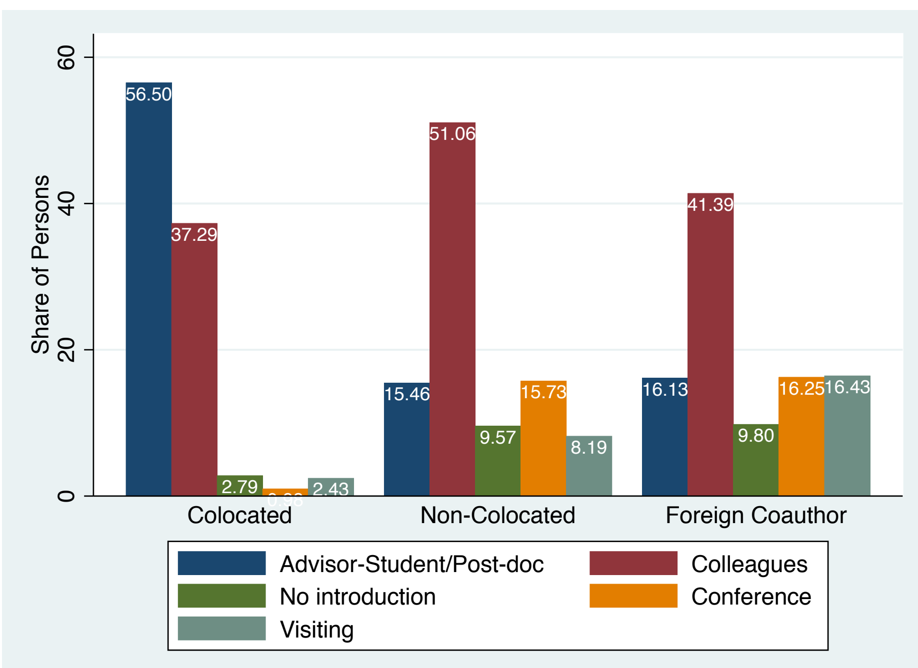
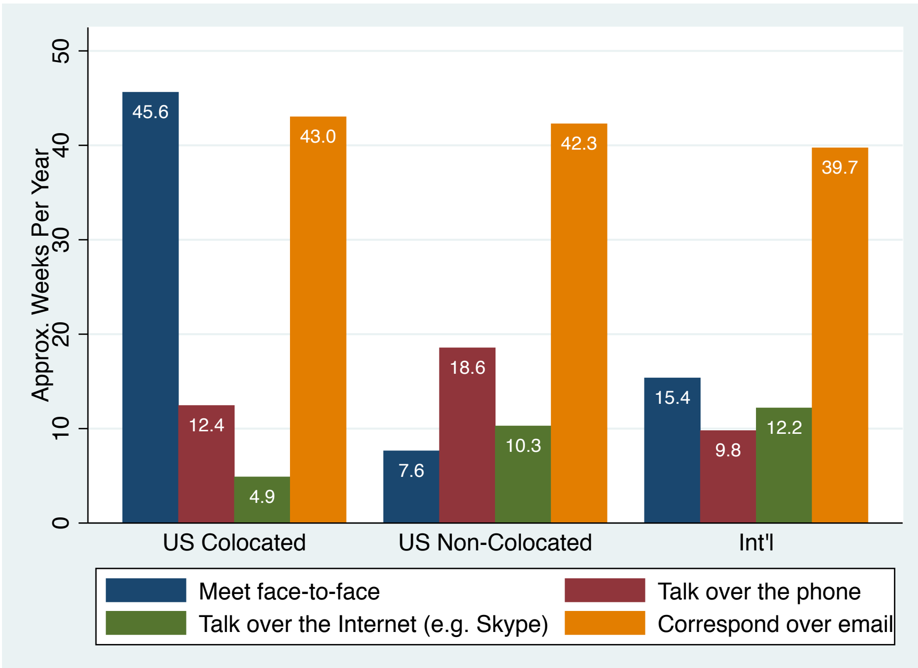


Figure 4: Share of Persons Who Were First Met in a Given Way by the Nature of Collaboration



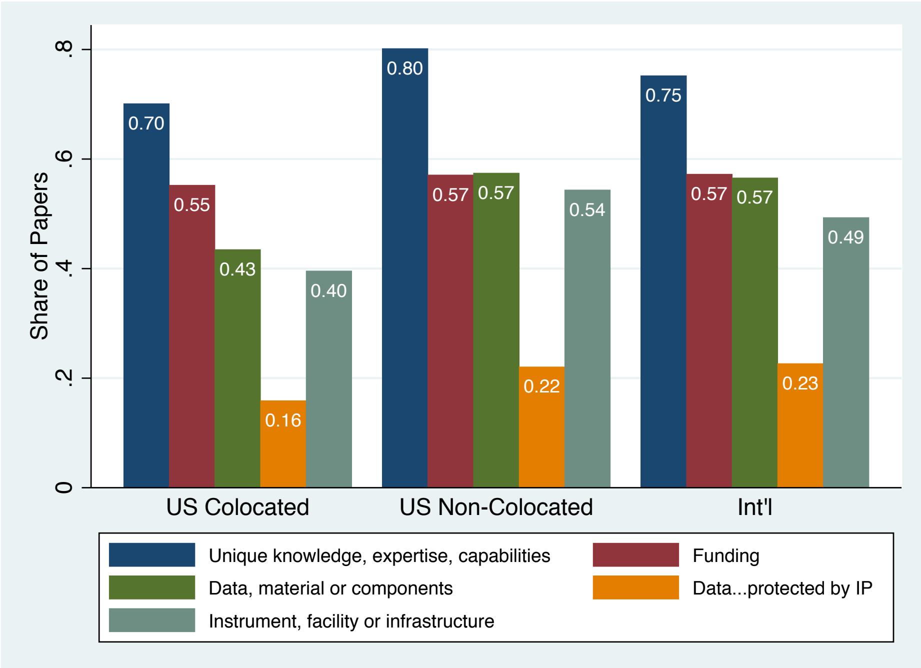
Notes: Share of all coauthors on papers for a given collaboration type. Question was phrased as "*How did you FIRST come in contact with each of these coauthors?*"

Figure 5: Overcoming Distance: Frequency of Communication Modes for 2-Author papers by the Nature of Collaboration (Approx. Weeks per Year)



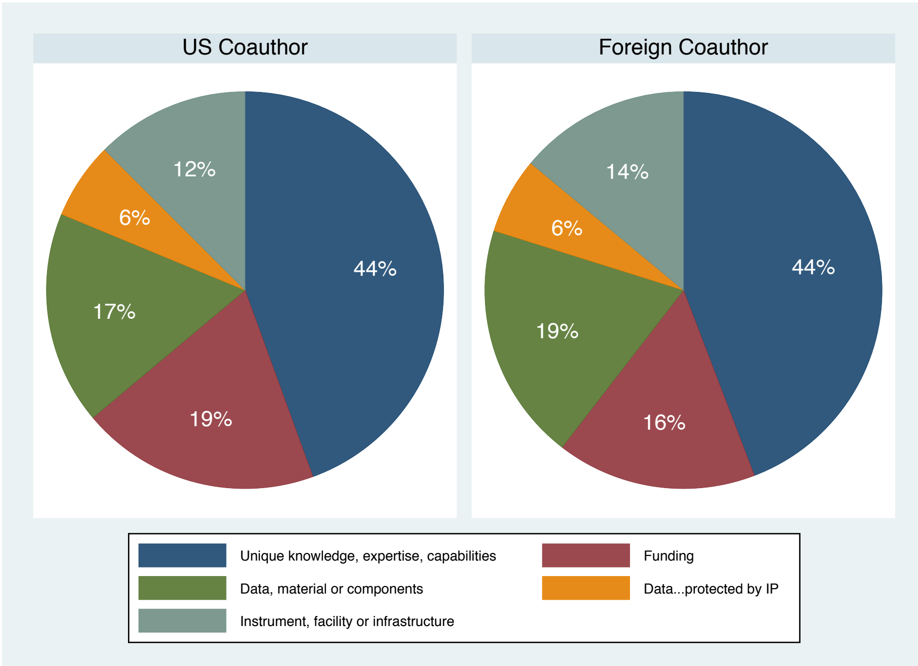
Notes: Question was phrased as "*When carrying out the research and writing for this article, how frequently did you use the following forms of communication with one or more of your coauthors?*" The possible choices were transformed into approximate number of weeks per year that each communication type was used: 6 = Every week (52), 5 = Almost every week (45), 4 = Once or twice a month (15), 3 = A few times per year (5), 2 = Less often than that (2), 1 = Never (0).

Figure 6: Contribution of Coauthors by the Nature of Collaboration



Notes: Share of papers for which the corresponding author reported at least one coauthor contributing the given resource. Question was phrased as "*Did any of the team members working on this article (including yourself) have access to one of the following resources that the other team members did NOT have, which made it important for you to all work together on this topic?*"

Figure 7. Contribution of US and Foreign Coauthors for 2-Author Papers



Notes: Share of US and foreign coauthors on 2-author papers only, as reported by the corresponding author.

APPENDIX

Table A. Optimal Team Size by Nature of Collaboration

	US Colocated	US Non-Colocated	All Int'l	Int'l with Cross-Country Funding
Yes	89.06	91.11	89.95	92.50
No, Additional	7.58	3.48	3.38	2.86
No, Fewer	3.37	5.40	6.67	4.64
N	1,663	574	1,154	280

Notes: Question was phrased as *"Do you think that the size of your team was optimal?"* The cross-country funding question was phrased as *"In carrying out the research for this article, did any of the coauthors receive funding that was specifically aimed at supporting cross-country scientific collaboration?"*

Figure A. Share of Coauthors Who Were First Met at a Conference

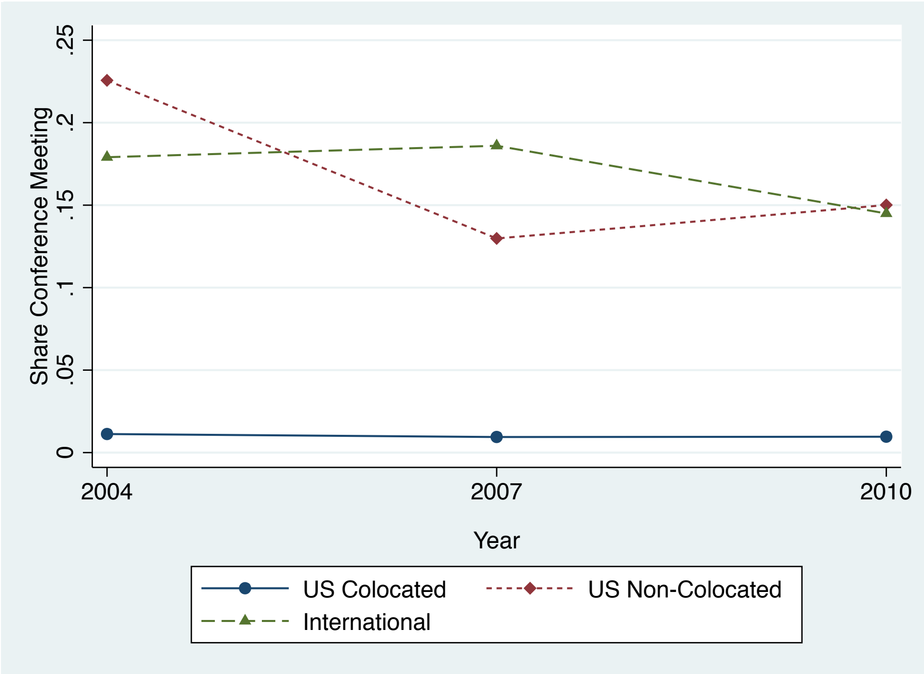


Figure B. Share of Coauthors Who Were First Met as Advisor-Student/Postdoc

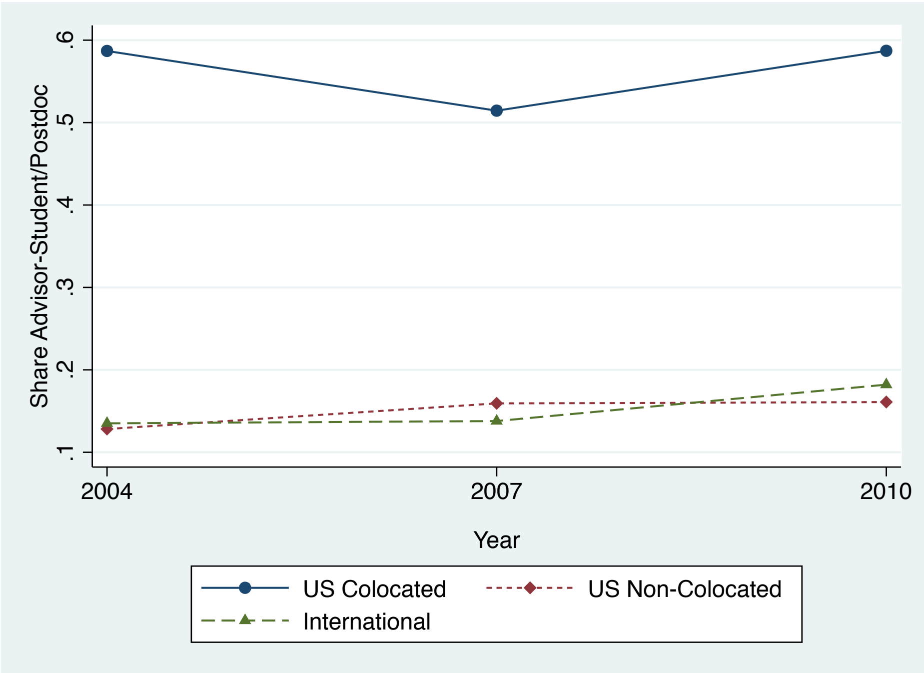


Figure C. Share of Papers With a Coauthor Contributing Data, Material or Components