A Field Experiment on Search Costs and the Formation of Scientific Collaborations

By Kevin J. Boudreau, Tom Brady, Ina Ganguli, Patrick Gaule,
Eva Guinan, Anthony Hollenberg, and Karim R. Lakhani*

Scientists typically self-organize into teams, matching with others to collaborate in the production of new knowledge. We present the results of a field experiment conducted at Harvard Medical School to understand the extent to which search costs affect matching among scientific collaborators. We generated exogenous variation in search costs for pairs of potential collaborators by randomly assigning individuals to 90-minute structured information-sharing sessions as part of a grant funding opportunity for biomedical researchers. We estimate that the treatment increases the baseline probability of grant co-application of a given pair of researchers by 75% (increasing the likelihood of a pair collaborating from 0.16 percent to 0.28 percent), with effects higher among those in the same specialization. The findings indicate that matching between scientists is subject to considerable frictions, even in the case of geographically-proximate scientists working in the same institutional context with ample access to common information and funding opportunities.

* Kevin J. Boudreau, London Business School (kboudreau@london.edu); Tom Brady, Massachusetts General Hospital and Harvard Medical School (HMS); Ina Ganguli, SITE, Stockholm School of Economics (Ina.Ganguli@hhs.se); Patrick Gaule, CERGE-EI, a joint workplace of Charles University and the Economics Institute of the Academy of Sciences of the Czech Republic (pgaule@cerge-ei.cz); Eva Guinan, Dana-Farber Cancer Institute and HMS (Eva_Guinan@dfci.harvard.edu); Anthony Hollenberg, Beth Israel Deaconess Medical Center and HMS; Karim R. Lakhani, Harvard Business School and IQSS (k@hbs.edu). KJB, IG, PG, EG and KRL designed, developed, and executed the experiment and contributed to the manuscript. IG and PG conducted data analysis. TB and AH contributed to the experimental development and execution. We appreciate helpful comments from Pierre Azoulay, Marcel Fafchamps, Ben Golub, Shane Greenstein, Nicola Lacetera, and Josh Lerner, and participants at the NBER Summer Institute Innovation Meetings, the NBER Productivity Lunch and the Georgia Tech REER Conference, Northwestern University and University of Rochester. Funding was provided by Harvard Catalyst Clinical and Translational Science Center (NIH UL1TR000170, UL1RR025758-02S4 and UL1TR001102.) and the Harvard-NASA Tournament Laboratory. We also thank Harvard Catalyst for support and cooperation in implementing the experiment, particularly Lee Nadler, William Chin, Laura Weisel, and David Frank. Amy Webber and Wei Zhong provided valuable assistance throughout the implementation process. Michael Menietti and Christoph Riedl assisted with execution of the experiment. The Harvard Innovation Laboratory generously provided us with meeting space for the experiment.

1. Introduction

The primary unit of scientific knowledge production has become the team or collaboration, rather than the lone scientist (Jones, 2009). Indeed teams are not only growing in frequency, but also in size and impact relative to single authors (Wuchty, Jones and Uzzi, 2007). Unlike settings inside of firms where executives and managers play a central role in organizing and forming teams (Lazear and Shaw, 2007), academic scientists have greater freedom and autonomy in selecting their collaborators and their topics of inquiry (Stephan 2012). Although there is a growing body of research on the productivity and outcomes of scientific teams once formed (e.g., Adams et al., 2005; Wuchty, Jones and Uzzi, 2007; Agrawal, Goldfarb, and Teodoridis, 2014), we currently know relatively little about the largely decentralized process by which scientific teams come into existence (Stephan, 2012). In this paper, we investigate the role of one particular mechanism—search costs and frictions—on these matching outcomes.

The role of search costs and resulting frictions in the formation of scientific collaborations is not well understood. On the one hand, the growing prominence of teams and falling communications and collaboration costs in science (Agrawal and Goldfarb, 2008; Ding, et al. 2010) might suggest forces favorable to novel team formation. On the other hand, geography and distance are regularly documented to play a role in shaping collaborations, even today (e.g., Rosenthal and Strange 2001; Glaeser 2010, Catalini, 2012); and, rather than continually forming novel collaborations, scientists most often work with partners in the same institution, in similar knowledge domain and within pre-existing social networks (Baccara and Yariv, 2013; Freeman, Ganguli and Murciano-Goroff, 2014; Freeman and Huang, 2014; Fafchamps, Goyal, and Van der Leij 2010; Azoulay, Liu and Stuart 2009). Moreover, past collaborations remain an important predictor of future ones. Although these patterns might be explained by any number of factors, they raise the question of whether search costs play a first order role in shaping the organization of scientists into teams.

The high information requirements for forming matches suggest that search frictions may be an important consideration. A large number of factors can play a role in decisions to collaborate—and these factors may be nuanced or difficult to always observe. This includes factors such as the complementarity of skills of prospective partners, current research interests

and priorities, access to broader sets of relevant resources (funding, equipment, research personnel), timing and scheduling constraints, and personal chemistry and disposition (Stephan, 2012). If acquiring and evaluating this information is costly, significant search frictions will appear, as has been found in other matching markets (Mortensen and Pissarides 1999). Observed patterns of collaboration might then be interpreted as reflecting limited information in decision-making—and therefore may constitute a suboptimal allocation of human resources.

To understand whether and to what extent search costs can impact the formation of collaborations among research collaborators, we carried out a field experiment with the goal of introducing exogenous variation in the information available to research scientists concerning potential collaborators. We worked closely with HMS clinical and administrative executives to modify and redesign existing internal grant processes so that causal inferences could be drawn in the context of a $800,000 grant opportunity for researchers at Harvard University and Harvard Medical School's (HMS) system of hospitals and research centers to encourage clinical applications of advanced medical imaging.

The experiment involved designing a research symposium (repeated on three consecutive nights) that was part of the grant process, where investigators were to get details about the grant rules and administration, learn about advanced technologies underlying the grant, and meet other researchers through structured information-sharing sessions. Participation in one of the symposia (and only one) was mandatory for the grant application, which was due four weeks after the symposia. Each symposium consisted of a 30-minute general introduction followed by 90 minutes of information-sharing in independent and physically separated "break-out" rooms. Break-out rooms facilitated face-to-face interactions by having researchers circulate about the room while their research ideas were "broadcast" in a standardized poster format. We reduced the cost of initial face-to-face interactions for random subsets of scientists by randomly assigning the roughly 400 researchers who took part to independent break-out rooms. Therefore, we can evaluate the effect of the treatment by simply comparing the likelihood of collaboration for pairs of researchers assigned to the same room (treatment) with the likelihood of pairs assigned to different rooms (control).

Important to note, estimates of implications of search costs in this context might be interpreted as occurring under "best case" conditions in that this context involves studying prospective collaborators operating within a shared institutional context, with funding

availability, within the same geographic area and in a context in which sophisticated information systems and tools have already been deployed to facilitate search for prospective collaborators. This might be indeed be the case here, where we study effects of those already choosing to participate in a research symposium intended to stimulate new research.

The results suggest that matching between scientists is subject to considerable frictions even in this "best case" context. We estimate that assignment to the same break-out room increased the probability of forming a collaboration by 75%, increasing the probability from 0.16 percent in the control group to 0.28 percent in the treatment. We estimate the effect to be significant at the 5% or 10% level, depending upon model specification. (The 95% confidence interval around the point estimate ranges from +4% to +112%). In addition to the increase in the baseline probability being large, the treatment effect is considerable as it represents a 30 percent of the boost in the probability of a collaboration associated with those working in the same hospital and performing research in the same area. This is a substantial effect for what is arguably a small (90-minute) treatment. It is plausible the effect might have been larger still had we not implemented the experiment under these "best case" conditions.

This main finding is consistent with large search costs and frictions playing a first-order role in shaping the process of searching for collaborators and suggests the important role played by information-rich face-to-face encounters in catalyzing collaborations. Consistent with the interpretation of a significant effect of search costs, the treatment effect was especially strong for pairs of researchers working in the same clinical area, where presumably search costs might be construed as lower given similar backgrounds and training. The findings therefore suggest the possibility that current observed patterns of collaborations in academic science are perhaps highly constrained by the availability of matching-relevant information and search costs. This is plausibly an important source of inefficiency. We cannot observe implications of this inefficiency within this analysis.

The finding of the first-order role played by search costs also offers one plausible explanation for the prevalence of homophily in forming collaborations, where like scientists tend to coauthor, and repeatedly, as both tendencies may economize on search costs.

The findings also imply potentially important differences between the *execution* of distributed collaborations versus the distributed *formation* of collaborations. The formation and execution of collaborations may be considered as representing altogether different kinds of

coordination problems—one of joint production and the other of matching. Whereas evidence suggests research collaborations may be able to be carried out at a distance through decreased communication and travel costs and increasingly sophisticated collaboration platforms (see Agrawal and Goldfarb 2008; Jones, Wuchty, and Uzzi 2008; Adams, Clemmons, Black, and Stephan 2005; Catalini, Fons-Rosen and Gaule 2014), the process of *forming* collaborations may still be especially highly influenced and informed by information-rich, interpersonal interactions.

The paper proceeds as follows. We first describe our experimental design, including details of the grant program and research symposia in Section 2. In Section 3 we describe the data. The empirical strategy and results follow in Sections 4 and 5, respectively. Section 6 concludes.

2. The Field Experiment

A. Harvard Medical School and its affiliated hospitals

Our field experiment involved faculty and researchers from Harvard University and its affiliated hospitals and institutions. Harvard Medical School and its 17 affiliated hospitals and research institutes (including Massachusetts General Hospital, Children's Hospital Boston, Brigham and Women's Hospital, Beth Israel Deaconess Medical Centre, and the Dana-Farber Cancer Institute) are a major force in biomedical research. Collectively, they employ more than 11,000 faculty and receive in excess of $1.5 billion in annual funding from the United States National Institutes of Health (NIH). Harvard researchers account for around 5 percent of scientific articles published in the top four medical journals, a larger share than Germany or Canada as a whole.[1] Fifteen researchers have shared in 9 Nobel prizes awarded for work done while at Harvard Medical School.

While our experiment is set entirely within the Harvard University system, in fact its researchers work in distinct organizations and research centers. The Harvard-affiliated hospitals are separately owned and managed and appear as separate entities in hospital rankings and lists of NIH recipients. Four of the five largest hospitals are located in the Longwood Medical Area campus in Boston while Massachusetts General Hospital has its own campus about 3 miles away

[1] Journals included are the *New England Journal of Medicine, Journal of the American Medical Association (JAMA), Nature Medicine,* and *Lancet*. Authors' calculations based upon research articles published during the period 2000–2009. Fractional counting was used when coauthors belonged to different institutions.

(and approximately 20 minutes by institutional shuttle bus). See Figure 1 for a map showing the locations of the largest Harvard-affiliated hospitals.

B. Harvard Catalyst and Advanced Imaging

Closing the gap between research findings and clinical applications ("bench to bedside") is a major priority for the NIH. This has resulted in the establishment of a new institute (National Center for Advancing Translational Sciences) that provides significant research funding to universities and hospitals that undertake collaborative "translational" activities to accelerate treatment development. As part of Harvard's efforts to promote clinical and translational research, the Harvard Clinical and Translational Center, Harvard Catalyst, provides seed funding in the form of pilot grants to support nascent research efforts. These pilot grants are awarded competitively to faculty within Harvard University. They emphasize early-stage research with the potential to improve human health. Pilot grant funding enables researchers to generate the preliminary data that is essential for larger grant applications to the NIH.

Our field experiment was layered onto a Harvard Catalyst pilot grant program. This particular grant opportunity, which offered $50,000 per award, was centered on proposals to devise or improve methods for using advanced imaging technologies (specifically, Physiological Magnetic Resonance (MR), Positron Emission Tomography (PET), and Optical Imaging) to address unmet clinical needs. A major challenge in the field of advanced imaging is that progress requires both expertise in the latest imaging tools and technologies and a deep understanding of the health problems to which they could be applied, with these different types of knowledge typically being held by people with different disciplinary backgrounds. Thus, advanced imaging is an archetypical example of a problem often found in modern science where advancing the knowledge frontier requires combining knowledge embodied in different individuals (Jones, 2009).

We worked in close collaboration with HMS administrators and executives to redesign their pilot grant process so that we could obtain causal inferences about the role of search costs in finding collaborators. While the grant process was primarily focused on identifying and funding promising early-stage translational research in the field of advanced imaging, Harvard Catalyst leaders also perceived a need for familiarizing clinicians with recent developments in advanced imaging and for Harvard-wide community building amongst researchers. This

provided us with the opportunity to create a new interactive research symposium where we could exogenously shift search costs for certain pairs of individuals by building in randomized face-to-face interactions. Hence we modified the Harvard Catalyst grant process by requiring potential applicants to attend an interactive research symposium that would be a forum to learn about new technologies, understand the grant process and exchange ideas amongst fellow researchers across Harvard. This was the first time such an interactive Harvard-wide symposium on a new research grant opportunity was offered.

In November 2011, all Harvard life sciences faculty and researchers were invited to participate in a unique funding opportunity centered on advanced imaging technologies. A total of up to $800,000 was available to support 15 pilot grants. There was the additional potential for researchers to apply for several concept development prizes of $2,000 each. The concept prizes were meant to stimulate innovative thinking and future investigation in areas in which imaging had not been previously considered as an intervention and did not require any implementation plan.

In the first stage, investigators who were interested in applying for the grants were asked to submit a Statement of Interest in which they briefly described a specific medical problem that advanced imaging techniques could potentially address. Basic biographical information (e.g. degree, institution, department appointment) was collected at this stage. Information distributed about the funding opportunity specified that eligibility to submit a final application was conditional on attending an advanced imaging symposium on one of three pre-announced dates. Applicants could indicate at this stage if there were any dates during which they could not attend a symposium. It was also communicated to applicants that the symposia would be studied by Harvard Catalyst to develop better insights about scientific team formation and that data on interaction patterns amongst individuals would be collected.

C. Randomization and the Advanced Imaging Symposium

The initial call generated 471 Statement of Interest applications, of which 435 applicants were invited to attend an advanced imaging symposium and thus proceed in the grant application process.[2] Forty-one applicants (9.4 percent) failed to RSVP or otherwise show up at the event.[3]

[2] Thirty-six statements of interests were outside the parameters of the request for applications in terms of area of inquiry (e.g. proposing Ultrasound or X-Ray Computed Tomography (CT) techniques) and the submitters were not invited to attend the symposium.

Additionally, invitations were extended to several individuals with world-class expertise in advanced imaging, bringing the total number of participants to 402.

The symposium was structured so that participants would come to the event prepared to discuss their idea with other participants in small break-out rooms of 30 to 40 people. The treatment was intended to introduce exogenous variation in search costs to some pairs of participants at the symposium by having them be present in the same break-out rooms at the event. Each participant was randomly allocated to a break-out room in advance so that a random subset of all possible pairs among all participants would receive the treatment. Three symposia were held on sequential nights and were identically structured, with four break-out rooms per night.[4] We also randomized the participants across nights, however we respected the 'black-out' dates for which applicants had previously indicated they would not be available.[5]

The events were held January 31, February 1, and February 2, 2012 at the Harvard Innovation Lab, located on Harvard's Allston campus. The program began with a 30-minute address by the program leadership describing the pilot grant opportunity and the agenda for the evening, including an introduction to advanced imaging tools and technologies. The break-out sessions then began in separate rooms. The number of participants in each room varied from 28 to 43.

The break-out room sessions were split into two periods of 45 minutes each, with a 15-minute break in the middle during which all participants could mingle in a common space where refreshments were provided. The rooms provided a venue for presentation of the participants' ideas in the form of posters. Each poster followed a standard format describing each participant's submitted idea from the Statement of Interest (based on information they had provided prior to the event) and was placed in the break-out room in advance.[6] The posters were intended to foster information sharing among participants, and included the following details related to the Statement of Interest idea: (1) What is your question? (2) Why does it matter? and

---

[3] We do not include these individuals in the analysis.

[4] The randomization was carried out by generating a unique random number for each participant, ranking the numbers, and then assigning participants to break-out rooms within nights based on their rank. We assigned 32 participants to the first 3 rooms each night, and the remainder (41-48) to the last room, which was slightly larger.

[5] Participants with black-out dates were a minority but to guard against the potential endogeneity of selection into nights for this group, the analysis will focus on comparisons *within* nights.

[6] Participants were provided with the following information in the emailed invitations to attend a symposium: "You do not need to bring any particular items with you. We have a poster prepared with your submitted answers to the three questions based on your statement of interest. Posters will be displayed at the symposium to facilitate talking about your idea with other attendees. There will be no formal presentations of any kind."

(3) What is needed for your research to succeed? A 300-character limit was imposed for each question. Posters were prepared in a standard size and format by Harvard Catalyst and each was placed on a separate white board that allowed for the possibility of visual explanations and note taking. Appendix Figure A1 provides several representative examples of participant posters from the event.

Participants within each break-out room were randomly split into two groups. Participants from Group 1 were asked to stand by their poster during the first period, while Group 2 participants circulated. The two groups then switched roles during period two (i.e. Group 1 participants circulated around the room while Group 2 participants stood near their own posters). The placement of each individual's poster in the room was also randomly determined in advance.

D. Grant Applications

Shortly after the symposia, all participants received via email an invitation to submit applications for the pilot grants or concept awards by the deadline of March 8, 2012. At this time, they also received PDF booklets with the names and contact information of all researchers who participated over the three nights and their posters[7]. The intention was to provide identical information to all participants apart from information acquired specifically in the break-out rooms.

Consistent with previous Harvard Catalyst pilot grant processes, applications had to include a principal investigator and at least one co-investigator. Concept award applications similarly had to include at least two individuals. Researchers with faculty appointments could apply as principal investigator on only one pilot grant, but could apply as co-investigator on an unlimited number of additional applications. Researchers without a faculty appointment could not be principal investigators on a pilot grant application, but they could be co-investigators on an unlimited number of applications. All attendees were eligible to apply for a concept award grant and could appear on an unlimited number of applications. Finally, at least one applicant on

[7] The following information was included in email communication with participants immediately following the event: "Attached to this email is a PDF booklet with the names and contact information of all researchers who participated over the three nights and their posters. We hope this is of use in contacting individuals that you met during the evening and in identifying additional potential collaborations and collaborators…. As described at the symposium, your proposal or your collaborators can be the same as suggested in your Statement of Interest or can be somewhat or entirely different. You can participate in multiple applications."

any grant application had to have attended the symposium. The grant application did not need to be based upon the initial Statement of Interest.

Extra care was taken to ensure that the symposium process did not somehow prime participants to seek collaborations only in their break-out rooms. Participants were informed that the composition of their teams would not be communicated to reviewers and would not be considered as a criterion for awarding the grant. They were also told to remove any personally identifying information about the submission teams from their proposals (including self-references and indications of special access to technologies).[8] This differed from the typically single-blinded process used in NIH and Catalyst grants, in case the identification of submission applicants might have an impact on collaboration choices. In the end, the majority of participants chose not to apply with other symposium participants: 66 percent of the applications included only one symposium participant as a co-applicant.

3. Data

A. Sources

To examine the impact of search costs on collaboration, we created a dyad-level dataset using a variety of data sources.

Registration data. Faculty and researchers interested in taking part in the funding opportunity were asked to submit a short Statement of Interest describing in 250 words or less a specific medical problem that advanced imaging techniques could potentially address. Registration data also included basic biographical information (rank, education history, hospital affiliation, department). Participants were also asked to identify themselves as primarily an imager or primarily a clinician. Clinical area and imaging modality were coded from the Statement of Interest documents.

Publications. We matched participants to Harvard Catalyst Profiles, an online, publicly accessible database that includes individual publication records and other information for faculty and researchers across Harvard Medical School. From the publication records, we deduced whether scientist pairs in our sample were previous coauthors.

[8] The following directions to applicants were highlighted in the grant request for applications: "As the initial review will be blinded in regard to the applicant(s), do not refer to yourself, other participants or institutions by name (e.g. substitute "our optical imaging experts", "our cardiology collaborators", "our laboratory" or "the genomics core" for specific individuals or facilities)."

Grant applications. Our main outcome variable comes from the pilot grant and concept award applications. Two hundred and twenty-four applications for pilot grants or concept awards were received.[9] Of those, 148 included one symposium participant in the applicant list, 49 included two symposium participants, and 27 included more than two symposium participants. We measure collaboration as any pairs of symposium participants appearing on the same application.

Location geocoding. Figure 2 shows the distribution of geographic distances between pairs of participants, which we created by geocoding the exact location of their offices and calculating the intervening distance in miles. The distribution has two clear peaks, with more than 30 percent of pair members located less than 0.5 miles apart and 25 percent of pair members located between 2.5 and 3.0 miles apart. The first peak corresponds to pairs where both members were based either in the same hospital or in different hospitals from the Longwood Medical Area. The second peak corresponds to pairs where one member was at the Longwood Medical Area and the other was at Massachusetts General Hospital.

[Insert Figure 2 Here]

B. Summary Statistics and Randomization Check

Table 1 provides individual-level summary statistics for symposium participants.[10] Of the 402 attendees, 29 percent were females, 42 percent identified themselves as imagers, and 73 percent held Harvard faculty appointments (the others being postdoctoral fellows or clinical fellows). Over 80 percent of attendees came from the four largest Harvard-affiliated hospitals: Massachusetts General Hospital, Brigham and Women's Hospital, Children's Hospital Boston, and the Beth Israel Deaconess Medical Centre. The most prevalent clinical expertise areas were neurology (25 percent), oncology (25 percent), and neuropsychiatry (10 percent).

[Insert Table 1 Here]

[9] Seventy-eight percent of applications were for pilot grants and the 22 percent were for concept awards.
[10] Across the three nights, 394 individuals were in attendance. However, five individuals with special expertise in advanced imaging attended the event on more than one night; we count them as different participants on each night, bringing the total number of participants to 402.

We can also compare the participants to the general population of researchers at Harvard Medical School. Appendix Table A1 provides summary statistics on participants and non-participants based on information in the Harvard Catalyst Profiles database. In terms of degree types, there was no significant difference in the share of MDs among attendees and the overall HMS population, but there was a larger share of PhDs among attendees (49 percent PhDs among attendees vs. 38 percent at HMS). We would expect a greater representation of PhDs at the event since it was part of a research grant opportunity, and academic PhDs are very often focused on research while academic MDs have a larger array of potential roles. Attendees also had more prior publications on average (approximately 4 publications more than the typical HMS researcher). We also see some significant differences in the distribution across ranks, with attendees more likely to be instructors and assistant or associate professors relative to the overall distribution at HMS, and less likely to be full professors and postdocs. Attendees were also more likely to come from MGH. One reason for this is that MGH houses a large advanced-imaging center, the Martinos Centre for Biomedical Imaging, and the focus of the grant opportunity was advanced imaging. For the same reason, individuals from radiology departments were overrepresented among attendees.

To verify that the randomization generated balance across covariates, we present summary statistics in Table 2 for the pairs in our sample assigned to the same break-out room and those assigned to different break-out rooms. Treated pairs and control pairs look very similar, with the exception of pairs of previous coauthors, pairs with both members from the same hospital, and pairs including one female, which are statistically different across treated and control pairs.[11] In our regression analysis, we will control for these covariates.

[Insert Table 2 Here]

The last row of Table 2 includes our outcome variable, collaboration. The incidence of collaboration is significantly larger in the treated group, which we investigate in a regression framework in the next section. It is notable that the incidence of collaboration is less than 0.2 percent in our sample. While this may seem low, the likelihood that two HMS faculty members

[11] The relative large difference between the percentage of pairs of previous coauthors across treatment and control groups can be explained by the very small number of pairs of previous coauthors in our sample (40 out of more than 20,000). Thus, randomization could easily result in a different incidence of pairs with coauthors across treatment and control groups, as it did in our case.

will co-publish in a given year is 0.06 percent and, thus, of the same order of magnitude.[12] Viewed through the lens of all pair-wise combinations of scientists who could collaborate, collaboration is indeed a relatively rare event.

[Insert Table 3 Here]

Table 3 shows characteristics of the collaborating dyads. Among attendees who attended on the same night but were not in the same break-out room, there were 33 pairs that co-applied. Among pairs in the same room at the event, there were 19 pairs that co-applied.[13] T-tests show that among the same-room collaborations, there was a higher incidence of pairs with one postdoc and of pairs researching the same clinical area. It is important to note that some of the within-room collaborations would have occurred in the absence of any treatment effect. Extrapolating the across-room incidence rate (0.16 percent) to the number of within-room pairs (7,149), we would expect 11 collaborations to have occurred within rooms in the absence of any treatment effect.

4. Estimation Strategy

A. Specification

We use the simplest possible estimation strategy to describe differences between treatment and control groups—and the effect of exogenous variation in search costs in our context. The approach of our statistical analysis is to study the incidence of collaborations among all possible pairs of participants attending on the same night within our experimental group of 402 individuals. This reduced-form approach suits our interest in studying the extent to which observed behaviors deviate from fully informed equilibrium outcomes.[14] This approach also allows us to deal with relatively small numbers of actual within-room collaborations in a most straightforward and conservative manner.

[12] Authors' calculation based upon publication data from Harvard Catalyst Profiles.

[13] The 19 pairs that co-applied from the same room correspond to 18 separate grant applications.

[14] Structural matching models that contemplate competitive equilibria in matching are an alternative approach to modelling the equilibrium formation of collaborations. However, pursuing such an approach requires we make structural assumptions regarding equilibrium search process and outcomes—which goes against our interests in this study, given our interest in investigating frictions. Therefore, it is more appropriate in this instance to proceed with a reduced form description of patterns to better describe any implications of search costs. Although this creates the possibility of downwardly biased estimates on the treatment effect, any such effect is likely to be vanishingly small: competition in matching is likely to have played only a small role, if much at all, as the absolute incidence of collaborations in these data is rather low and individuals were not limited in the number of collaborations they can form.

Thus, the unit of analysis is the scientist *pair* and the data set includes *every* possible pair of scientists across all nights. We use a linear probability model to describe how the incidence or probability of collaborations differs across treatment and control groups (i.e., those in the same versus different break-out rooms). Random assignment of pairs within the research design allows us to interpret differences as causally related to exogenous variation in search costs. We are also able to regress the incidence of collaborations on other covariates of researcher pairs, to further describe associations with the incidence of collaborations. To measure whether treatment effects varied across subgroups, we interacted *Same Room* indicator with pair-level variables.

Thus, to estimate the impact of being in the same room at the event on the likelihood of collaboration between pairs, we ran linear regressions with the following specification:

$$(1) \quad Collaboration_{ij} = \alpha + \beta Same\ room_{ij} + \theta Same\ room_{ij} * Distance_{ij} + \pi Distance_{ij} + \delta X_{ij} + \varepsilon_{ij}$$

where the key explanatory variable associated with the treatment effect, *Same Room*$_{ij}$, is an indicator variable that equals 1 if both researcher *i* and *j* were randomly assigned to the same break-out room at the symposium.[15] *Collaboration*$_{ij}$ is an indicator variable for whether *i* and *j* appeared on any common pilot grant or concept award applications. X_{ij} is a vector of observable pair-level characteristics that can impact the likelihood of collaboration and includes measures of gender and professional rank. The vector $Distance_{ij}$ includes measures of differences in professional rank, as well as geographic, scientific and past coauthoring, described below. The model also includes fixed effects for each night of the symposium.

We estimate and report equation (1) using a linear probability regression (OLS) and standard errors using Eicker-White heteroskedasticity-robust standard errors. We also estimated each model using grouped dyadic standard errors as suggested by Fachamps and Gubert (2007) and found virtually identical t-statistics and confidence intervals. We do not cluster standard errors by night of attendance, since assignment to nights is itself random (conditional on black-out dates for a minority of participants) (Cameron and Miller, 2013).

[15] There are several other ways to study and model search costs in this setting. We could, for example, study the effect of attending the symposium on the same night. Furthermore, since participants' posters were also randomized within the break-out rooms, we could study if immediate neighbors in the break-out room at the event had an impact on collaboration. However, neither of these approaches had a significant impact on our outcome of interest, grant co-applications.

B. Covariates

Several additional covariates describing pairs are also included in the model. Inclusion of these covariates should not affect the point estimate of the treatment effect, but should increase its precision and offer further opportunity for interpretation. Our vector of pair-level covariates, X_{ij}, includes variables for gender and professional rank. Gender is captured by indicator variables *Both female*, *One female*, and *Both male*. Past research indicates that women have a greater propensity to work with other women (Boschini and Sjogren 2007) and have more limited academic networks, more generally (see Ding, Levin, Stephan, and Winkler 2010). For professional rank, we include indicators for *One postoc* in the pair and *Both postdocs*. Postdocs were eligible to apply for either the concept or pilot grants; however, two postdocs could collaborate on a pilot grant application only if a third team member with a faculty appointment assumed the role of principal investigator.

The vector $Distance_{ij}$ includes measures of differences in professional rank, and geographic, scientific and past coauthoring. Given the potential relevance of these various forms of distance to search costs, coefficients estimated on these variables provide at least some broad and rough means of judging the importance of any estimated treatment effect by direct comparison with coefficients on these variables.

Distances in professional range are measured with indicator variables corresponding to possible combinations of differences. In relation to geographic distance, we create an indicator for *Same hospital*, which indicates whether pair members' primary appointments are in the same Harvard-affiliated hospital or institute. We also create an indicator for *Both Longwood*, indicating that both members of the pair work on the same campus,[16] as the largest concentration of researchers are located in hospitals and institutes either on the Longwood Medical Area (LMA) campus or at the Massachusetts General Hospital (MGH) campus. The campuses are located approximately three miles apart (with a travel time of about 20 minutes during normal traffic). We also create a direct measure of geographic distance by geocoding exact locations of offices and calculating pairwise distances in miles.

[16] The LMA includes eight hospitals/institutes in our sample and the MGH campus includes two hospitals/institutes. The other hospitals/institutes in the sample are considered to be individual campuses.

In relation to scientific or intellectual distance, we create indicator variables for *Both imagers, One imager + one clinician*, and *Both clinicians*. We construct this variable using the information attendees themselves reported during the initial stage of the application process. We also constructed indicator variables for *Same clinical area* and *Same imaging modality* (Physiological MR, PET, or Optical Imaging). These were coded from the Statement of Interest documents submitted in the first stage of the application process. We also create measures of scientific distance using overlap in the Medical Subject Heading (MeSH) terms from each individual's publications, and overlaps in the keywords of each individual's Statement of Interest.[17]

A final measure of distances is whether the pair had previously collaborated, indicator variable *Previous coauthors*. We also distinguish cases of one single past co-publication with more than one past co-publication with indicator variables.

5. Analysis & Results

A. Does reducing search costs increase the propensity to collaborate?

We first analyze whether our 90-minute breakout treatment had an effect on the incidence of collaborations and the magnitude of any such effects. OLS estimates with robust standard errors are presented in Table 4. (The same results are presented using probit estimation in Table A2.) Column 1 shows the basic result, regressing the incidence of collaborations on our treatment effect indicator and a constant. The baseline probability of collaborations is captured by the constant coefficient of 0.0016 or 0.16 percent. The point estimate shows that the treatment increases the likelihood of collaborating on an application by approximately 75 percent (increasing the likelihood of a pair collaborating from 0.16 percent to 0.28 percent).[18] The estimate is significant at the 10 percent level.

[Insert Table 4 Here]

[17] We include these other measures of scientific distance in our regression analysis, but since these measures rely on prior publications (and some individuals in the sample have zero or few publications), our preferred measure is self-reported clinical area.

[18] However, our point estimates regarding the magnitude of the effect are imprecise. The confidence interval ranges from +4 percent to +112 percent.

The advanced imaging symposia were held on three different nights. We thus include fixed effects for the night of the event (January 31, February 1, or February 2) in Column 2 to account for any differences across nights. The night fixed effects are not significant and their inclusion has very little impact on the same room coefficient (or its standard error).

In Column 3 we introduce pair-level variables to account for gender composition, differences in rank as well as geographic, scientific, and past coauthoring distance. The random assignment ensures that being in the same room is orthogonal asymptotically to any observable or unobservable pair characteristic.[19] Correspondingly, introducing covariates does not statistically change the estimated treatment effect. Standard error is not palpably changed, but significance marginally increases, on account of a small increase in the point estimate. The point estimate for the effect of being in the same room increases slightly from 0.0012 to 0.0014. (Note that we also include additional controls in other specifications, including dummies for whether a pair was in the same group within a break-out room (group 1 or 2) and their proximity to one another in the room (whether the pair had posters next to each other) but the results do not change.)

The estimated coefficients on the additional pair-level variables correspond with what would be expected (Section 5.B). The results show that working in the same clinical area, being affiliated with the same hospital and being a co-author in the past are positively and significantly correlated with collaboration. Consistent with the related literature, these results suggest that geographic, scientific, and past coauthoring are all positively related to collaboration. Given the likely complementarities of skills and knowledge between imagers and clinicians, our prior was that collaborations would be more likely to form when one pair member was a clinician and the other was an imager. The results show that such pairs of one imager and one clinician were significantly more likely to collaborate than pairs of clinicians only, but collaborations were even more likely to form when both members of the pair were imagers. Collaboration was significantly less likely to occur between pairs consisting of two postdocs, which is possibly explained by the fact that two postdocs could collaborate on a pilot grant application only if a third team member with a faculty appointment assumed the role of principal investigator.

[19] Being in the same room is orthogonal to pair characteristics *ex ante*. However, *ex post*, being in the same room at the event could be correlated with pair characteristics by chance. While this is much less of a concern than in observational data (Leamer 2010), it is nonetheless useful to control for relevant, observable pair characteristics to address the possibility that the effect of being in the same room is affected by differences in observable pair characteristics. Introducing controls has the added benefit of improving the precision of the *Same room* estimate by reducing the unexplained variance.

Overall, the single largest correlate is whether scientists had previously coauthored a publication. This association is at least an order of magnitude larger than for each of the other correlates.

Therefore, our estimated treatment effect of being in the same break-out room on collaboration is over 30 percent of the effect of being from the same hospital (0.0044) and of researching the same clinical area (0.0040). Relative to the single most important correlate, past co-authorship, it is only about 1 percent of the magnitude (0.1126). (The probit estimates in Table A2 of the appendix show similar results.)

B. For which pairs does reducing search costs have the greatest effect?

Next, we investigate whether the treatment had an effect for different types of pairs. Unlike earlier estimates of correlations with covariates, interaction terms can be interpreted causally. Probit estimates are reported in Table A3 of the appendix. We introduce the interactions between covariates with the treatment effect individually in Columns 1–7 and then simultaneously in Column 8 of Table 5. In introducing each of the interaction terms, we also of course re-introduce the direct effect of the covariate in the regressions; however, our focus here is on interactions terms.

Reviewing results of Columns 1 through 7, the only interaction term found to be significant is that in Column 6, which reports a positive and significant coefficient on the interaction between the treatment effect and the indicator for researchers being in the same clinical area. The coefficient on the direct treatment effect term *Same Room* becomes statistically indistinguishable from zero when introducing this interaction. Results in Column 8 corroborate this result, as introducing all covariates and all interaction terms at once in the model produces an almost identical estimate on this interaction terms. In Column 8, which includes all interactions, pairs researching the same clinical area increases the likelihood of a pair collaborating from 0.35 percent (the sample average incidence of collaboration for pairs working in the same clinical area but not in the same room) to 0.94 percent relative to pairs researching different clinical areas.

There are several possible explanations for the effect, but it suggests that researchers had limited information about these potential collaborations—either about who else was working on applying advanced imaging to the same clinical area, or about the potential benefits of collaborating with these individuals. If they did, the information provided at the event should not

have any independent effect for these pairs. It may also be the case that discussions were more beneficial for clinically-proximate pairs because they shared common ground, allowing them to convert their discussions into collaborations. Another possible explanation is that it is quite costly to switch clinical areas (specialization and training in medicine occurs on the basis of clinical areas, e.g. Dermatology, Neurology, Oncology), and therefore, even if researchers talked to people with interesting ideas in other clinical areas at the event, the benefits to collaboration were highest for those in the same clinical area.

We fail to detect evidence of the significance of other interactions. Our results on the interaction between being in the same room at the event and other pair characteristics are not conclusive. While the point estimates for some interactions are positive, they are not significant up to the 10 percent level.[20]

[Insert Table 5 Here]

We also investigated various alternative specifications such as including more fine-grained measure of geographic distance, scientific distance, or past coauthoring,[21] as well as controlling more flexibly for ranks and rank differences between pair members. We included dummies for whether a pair was in the same group (1 or 2), proximity in the room (whether the pair had posters next to each other), and the number of total individuals in the room (to test whether density mattered), but the results were not significant and the same clinical area result is consistent and stable across these specifications.

6. Summary and Conclusions

Teams are a primary unit of knowledge production and scientists, in large part, self-organize into research teams. Yet we know little regarding the matching of scientists into teams. In this paper, we present the results of a field experiment to investigate the role of search costs in

[20] The interaction between being in the same room and pairs with one woman are marginally significant with p-values of 0.093 in the probit specification and 0.133 in the OLS specification. A differential effect for pairs with a woman would be consistent with the findings of Ding et al. [2010], who show that the introduction of IT benefited collaborations more for female scientists than for male scientists, since women tend to have less diverse networks, have lower job mobility, and more constraints to attending conferences and seminars. These factors would similarly lead women to benefit more from mixing with other researchers at the event in terms of finding coauthors.

[21] We considered, for instance, whether pair members investigated the same imaging modality, the extent of the overlap of scientific keywords in previous publications, and whether pair members shared a common coauthor.

the formation of scientific teams by comparing the incidence of collaborations among researchers who participated in the same break-out rooms within an interactive research symposium as part of a grant proposal process, versus those who were assigned to different break-out rooms. We thus randomly varied search costs for a set of prospective collaborators, observing both the collaborations that did form along with those that did not.

We find that the small, focused treatment significantly increased the incidence of collaboration on subsequent grant proposals by 75% in relation to the baseline probability of collaboration between pairs of researchers (increasing from 0.16 percent in the control group to 0.28 percent in the treated group). The magnitude of this effect is equivalent to roughly a third of the boost in probability of collaboration associated with working in the same hospital or, alternatively, the probability of working in the same clinical area. In these regards, the point estimate can be viewed as rather large, despite the relatively small and focused nature of the treatment (i.e., a 90-minute break-out session). It is in fact notable we find any effect at all, let alone such a large effect in the context of scientists who are already geographically proximate and working within a common institutional context, where online resources and information systems already exist to facilitate collaboration.

We interpret these large effects as showing that even when working in relatively favorable conditions, search costs and frictions continue to powerfully shape (and limit) the formation of collaborations between scientists. Whereas a great deal of collaborative work might potentially be *performed* at a distance, the *formation* of collaborations appears to be highly sensitive to information-rich face-to-face interactions. In this sense, the question of the "death of distance" and the role of collocation and information technology, for example, might be reconsidered at least in relation to questions of *forming* collaborations.

The finding is consistent with the complex and manifold set of variables upon which collaboration decisions might be based and the effectiveness of face-to-face interactions in rapidly conveying information through high frequency, rapid feedback and visual and non-verbal cues (Storper and Venables, 2004). For example, given our existing communications technologies, it may remain difficult to wholly codify current research interests, complementarity in knowledge and skills, access to resources, timing and scheduling constraints—let alone questions of personal chemistry and disposition or more subtle questions of one's intellectual outlook. The result is also consistent with face-to-face interactions potentially triggering or

credibly signaling commitments, establishing trust and personal chemistry (Azoulay, Liu, and Stuart 2009).

Further consistent with the role of search costs in our results, the treatment effect was most pronounced on subsets of scientist pairs who are less "distant", working within the same clinical area, and therefore perhaps needing to overcome lower information and search cost hurdles. We also found positive associations between the likelihood of forming collaborations as prospective collaborators coming from the same hospital and the single most important predictor of collaborations—in terms of coefficient magnitude—was whether individuals had previously collaborated.

In documenting an important role played by search costs in influencing the formation of collaborations, we leave open a range of related questions. For example, in this paper, we did not study nor observe longer-run outcomes of scientific productivity such as subsequent publications. (Initial analysis of reviewers' assessments of the grant applications indicates no statistical difference between scores of applications submitted by pairs in the same room versus pairs not in the same room at the event (see Appendix Table A4)). Also, we demonstrate what are arguably large effects of the particular treatment we implemented here. However, the treatment exploited here is not necessarily optimal and could be subject to further improvement. Such insights could be relevant in devising improved means of designing supporting information systems and matching facilities. An additional, potentially rather important, series of questions falling outside the scope of this study concerns how individuals develop their own stock of matching-relevant stock of information and heuristics in the first place (apart from effects of situational or episodic shocks in information, as were explored here).

The patterns documented here also raise questions regarding the extent to which "homophily", (Lazarsfeld and Merton, 1954; Baccara and Yariv, 2013) exemplified by increased likelihood for scientists to form ties with other scientists possessing similar personal characteristics, might, at least in large part, be the result of search costs—rather than reflecting collaboration preferences (Boudreau and Lakhani, 2012) or lower coordination costs when collaborating with similar partners (Reagans and Zuckerman, 2001; Reagans, Zuckerman, and McEvily, 2004).

Despite these limitations, we see the present study as a step toward opening the black box of how scientific collaborations form. In recent years there has been considerable interest in the policy arena in fostering collaborations, and especially interdisciplinary collaborations, in particular by the U.S. government agencies funding fundamental research and development (a combined budget of $36 billion in 2011), the NIH and the National Science Foundation (NSF). Yet there is scant evidence indicating how to do this in practice. On a methodological level, we are—to the best of our knowledge—the first to bring field experimental methods to a workplace setting where the participants are engaged in scientific knowledge production. Evidence from randomized experiments on the scientific community such as ours will presumably be increasingly valuable to policymakers as they consider reforms to scientific institutions (Azoulay 2012). We show that creating settings where scientists meet face-to-face to discuss early-stage research ideas can be useful for fostering collaboration. However, time spent in such "mixer" events has opportunity costs, and we thus remain agnostic on the effect of such activities on scientific productivity and on welfare more generally.

REFERENCES

Adams, J., R. Clemmons, C. Black, P. Stephan. 2005. Scientific teams and institutional collaborations: Evidence from U.S. universities. *Research Policy* 34:259–285.

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

Agrawal, A., A. Goldfarb, F. Teodoridis. 2013. Does the Accumulation of Knowledge Cause an Increase in Returns to Collaboration? Evidence from the Collapse of the Soviet Union. *NBER Working Paper 19694*.

Azoulay, P. 2012. Research efficiency: Turn the scientific method on ourselves. *Nature* 484, 31–32 (05 April 2012).

Azoulay P., C. Liu, T. Stuart. 2009. Social influence given (partially) deliberate matching: Career imprints in the creation of academic entrepreneurs. HBS Working Paper 9-136.

Baccara, M., L. Yariv. 2013. Homophily in peer groups. American Economic Journal: Microeconomics, 5(3), 69-96.

Boudreau, K. J., K.R. Lakhani. The Confederacy of Heterogeneous Software Organizations and Heterogeneous Developers: Field Experimental Evidence on Sorting and Worker Effort. In The Rate and Direction of Inventive Activity Revisited, edited by Josh Lerner and Scott Stern, 483–502. University of Chicago Press, 2012.

Boschini, A., A. Sjogren. 2007. Is team formation gender neutral? Evidence from coauthorship patterns. *Journal of Labor Economics* 25(2): 325–365.

Catalini, C. 2012. *Microgeography and the direction of inventive activity*. Mimeo, MIT Sloan School of Management.

Catalini, C., C. Fons-Rosen, P. Gaule. 2014. Air travel costs and scientific collaborations. Mimeo. MIT Sloan School of Management.

Cameron, A. C., D. L Miller. 2013. A Practitioner's Guide to Cluster-Robust Inference. Forthcoming in *Journal of Human Resources*.

Ding, W., S. Levin, P. Stephan, A. Winkler. 2010. The impact of information technology on academic scientists' productivity and collaboration patterns. *Management Science* 56(9): 1439–1461.

Fafchamps, M., F. Gubert. (2007). The formation of risk sharing networks. *Journal of Development Economics*,o*83*(2), 326-350.

Fafchamps, M., S. Goyal, M. Van der Leij. 2010. Matching and network effects. *Journal of the European Economic Association* 8(1): 203–231.

Freeman, R. B., I. Ganguli, R. Murciano-Goroff.. 2014. Why and Wherefore of Increased Scientific Collaboration. NBER Working Paper 19819. Forthcoming in A. Jaffe and B. Jones (eds), *The Changing Frontier: Rethinking Science and Innovation Policy*, National Bureau of Economic Research.

Freeman, R. B., & Huang, W. 2014. "Collaborating With People Like Me: Ethnic co-authorship within the US." NBER Working Paper No. w19905.

Glaesar, E. L. 2010. Agglomeration Economics. Chicago: University of Chicago Press.

Lazear, E. P., K. L. Shaw. 2007. Personnel economics: The economist's view of human resources. *Journal of Economic Perspectives* 21: 91-114.

Leamer, E. 2010. Tantalus on the road to Asymptotia. *Journal of Economic Perspectives* 24(2): 31–46.

Jones, B. 2009. The burden of knowledge and the 'death of the renaissance man': Is innovation getting harder? *Review of Economic Studies* 76: 283–317.

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

Katz, J. 1994. Geographical proximity and scientific collaboration. *Scientometrics* 31(1): 31–43.

Katz, J., B. Martin. 1997, What is research collaboration? Research Policy 26: 1-18.

Mairesse, J., L. Turner. 2005. Measurement and explanation of the intensity of co-publication in science: An analysis at the laboratory level". NBER Working Paper 11172.

Mortensen, D., C. Pissarides. 1999. New developments in models of search in the labor market. In *Handbook of Labor Economics*, eds. O. C. Ashenfelter and D. Card, Vol. 3b, 2567–2627. Amsterdam: Elsevier.

Reagans, R., E. W. Zuckerman. 2001. Networks, diversity, and productivity: The social capital of corporate R&D teams. Organization science, 12(4), 502-517.

Reagans, R., E.W. Zuckerman, B. McEvily. 2004. How to make the team: Social networks vs. demography as criteria for designing effective teams. Administrative Science Quarterly, 49(1), 101-133.

Rosenthal, S. S., W. C. Strange. 2001. The Determinants of Agglomeration. *Journal of Urban Economics* 50(2), 191–229.

Storper, M., A. Venables. 2004. Buzz: Face-to-face contact and the urban economy. *Journal of Economic Geography* 4(4): 351–370.

Wuchty S., B. Jones, B. Uzzi. 2007. The increasing dominance of teams in production of knowledge. *Science* 316: 1036.

Table 1. Summary Statistics, Attendees

	Sample Mean
Female	0.29
Faculty member	0.73
Imager	0.42
Longwood	0.51
Hospital	
Massachusetts General Hospital	0.37
Brigham and Women's Hospital	0.19
Beth Israel Deaconess Medical Center	0.14
Children's Hospital Boston	0.13
Other	0.17
Clinical Area (SOI)	
Neurology	0.25
Oncology	0.25
Neuropsychiatric	0.10
Cardiovascular	0.06
Gastroenterology	0.04
Transplantation	0.04
Ophthalmology	0.03
Other	0.23
Attended on Jan. 31	0.35
Attended on Feb. 1	0.32
Attended on Feb. 2	0.33
Observations	402

Notes: See Section III in the text for a detailed description of the variables.

Table 2. Dyads by Treatment Status

Sample Means	*Treatment: Same Room*	*Control: Different Room*	*Difference*
One postdoc	0.404	0.396	-0.007
Both postdocs	0.072	0.075	0.003
One female	0.403	0.418	0.015*
Both female	0.085	0.082	-0.004
Same hospital	0.198	0.208	0.010+
Both Longwood	0.266	0.258	-0.010+
One imager + one clinician	0.492	0.489	-0.003
Both imagers	0.175	0.176	0.001
Same clinical area (SOI)	0.123	0.119	-0.004
Previous coauthor	0.001	0.002	0.001*
Collaboration (Outcome variable)	0.0028	0.0016	-0.0012*
Observations	6,702	19,962	

Notes: The category *Treatment: Same Room* refers to participants in the same room at the event; it was randomized across pairs of participants attending on the same night. *Collaboration* indicates whether the pair appeared on any common pilot grant or concept award applications. See Section III in the text for a detailed description of the variables. Stars indicate the results of t-tests for equality of means. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Table 3. Collaborating Dyads by Treatment Status

Sample Means	*Collaborations within the same room*	*Collaborations across rooms*	*Difference*
One postdoc	0.421	0.212	-0.209
Both postdocs	0.000	0.030	0.030
One female	0.474	0.303	-0.140
Both female	0.158	0.061	-0.097
Same hospital	0.579	0.636	0.057
Both Longwood	0.158	0.303	0.145
One imager + one clinician	0.474	0.485	0.011
Both imagers	0.316	0.394	0.078
Same clinical area (SOI)	0.579	0.273	-0.306*
Previous coauthor	0.105	0.121	0.016
Observations	19	33	

Notes: *Collaboration* indicates the pair appeared on a common pilot grant or concept award application. See Section III in the text for a detailed description of the variables. Stars indicate the results of t-tests for equality of means. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Table 4. Main effect of Treatment on Collaboration

DV = Collaboration	(1)	(2)	(3)
Same Room	0.0012^{+}	0.0012^{+}	0.0014^{+}
	(0.0007)	(0.0007)	(0.0007)
One postdoc			-0.0008
			(0.0005)
Both postdocs			-0.0014*
			(0.0007)
One is female			0.0002
			(0.0006)
Both are female			0.0010
			(0.0011)
Same hospital			0.0042**
			(0.0010)
Both Longwood			-0.0001
			(0.0006)
One imager + one clinician			0.0008^{+}
			(0.0005)
Both imagers			0.0025*
			(0.0010)
Same clinical area (SOI)			0.0042**
			(0.0014)
Previous coauthor			0.1176*
			(0.0468)
Constant	0.0016**	0.0012**	-0.0010
	(0.0003)	(0.0004)	(0.0007)
Night fixed effects	No	Yes	Yes
R2	0.000	0.000	0.017
Nb. of Obs.	26,664	26,664	26,664

Notes: Dependent variable is *Collaboration*, an indicator variable for whether the pair appeared on any common pilot grant or concept award applications. The main variable of interest is *Same room*, which was randomized across pairs attending on the same night. All estimation is by OLS. Robust standard errors in parentheses. $^{+}$ *p < 0.10,* * *p < 0.05,* ** *p < 0.01*

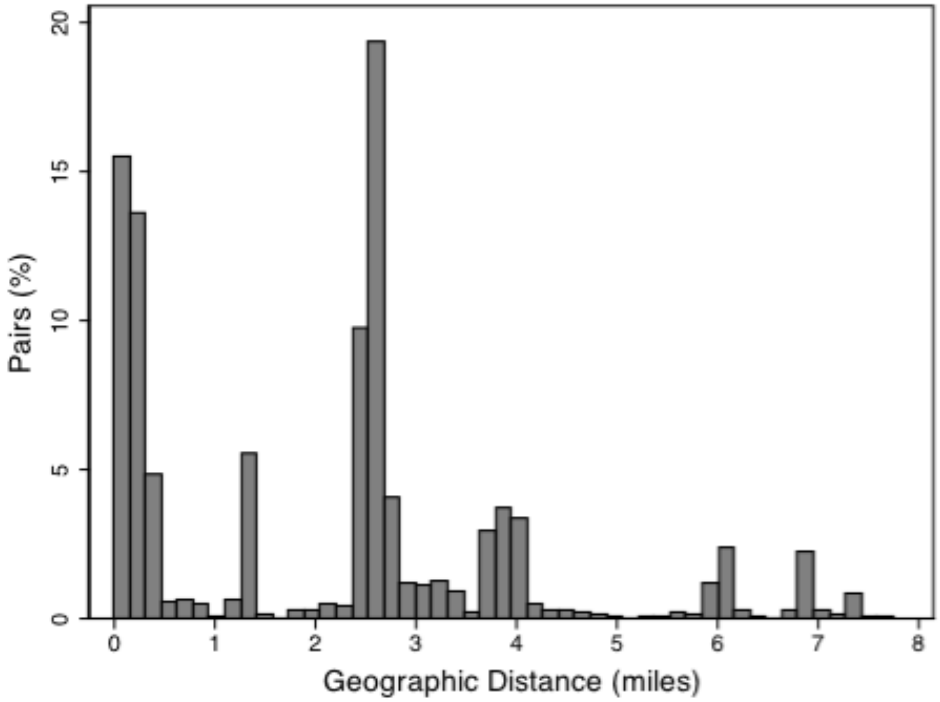
Table 5. Treatment and Interactions with Measures of Distance

DV = Collaboration	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Same Room	0.0008	0.0000	0.0008	0.0018^{+}	0.0012	-0.0000	0.0011	-0.0019
	(0.0011)	(0.0009)	(0.0006)	(0.0009)	(0.0009)	(0.0006)	(0.0007)	(0.0015)
One postdoc	-0.0016**							-0.0012*
	(0.0006)							(0.0005)
Same rm X One postdoc	0.0015							0.0019
	(0.0015)							(0.0015)
Both postdocs	-0.0016^{+}							-0.0012
	(0.0008)							(0.0008)
Same rm X Both postdocs	-0.0014							-0.0011
	(0.0012)							(0.0012)
One is female		-0.0008						-0.0004
		(0.0006)						(0.0006)
Same rm X One female		0.0021						0.0024
		(0.0015)						(0.0015)
Both are female		-0.0007						-0.0000
		(0.0010)						(0.0010)
Same rm X Both female		0.0040						0.0035
		(0.0033)						(0.0030)
Same hospital			0.0044**					0.0037**
			(0.0011)					(0.0011)
Same rm X Same hospital			0.0024					0.0023
			(0.0028)					(0.0027)
Both Longwood				0.0005				0.0004
				(0.0007)				(0.0007)
Same rm X Both Longwood				-0.0020				-0.0018
				(0.0014)				(0.0015)
One imager + one clinician					0.0009^{+}			0.0008^{+}
					(0.0005)			(0.0005)
Same rm X One imager					-0.0000			-0.0000
					(0.0014)			(0.0013)
Both imagers					0.0031**			0.0025*
					(0.0011)			(0.0011)
Same rm X Both imager					0.0002			-0.0002
					(0.0025)			(0.0024)
Same clinical area (SOI)						0.0025^{+}		0.0018
						(0.0013)		(0.0013)
Same rm X Same clin area						0.0095*		0.0094*
						(0.0042)		(0.0041)
Previous coauthor							0.0917*	0.0889*
							(0.0443)	(0.0442)
Same rm X Prev coauthor							0.2389	0.2363
							(0.1974)	(0.1967)
Night fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.001	0.000	0.003	0.000	0.001	0.002	0.019	0.024
Nb. of Obs.	26,664	26,664	26,664	26,664	26,664	26,664	26,664	26,664

Notes: Dependent variable is *Collaboration*, an indicator variable for whether the pair appeared on any common pilot grant or concept award applications. The main variable of interest is *Same room*, which was randomized across pairs attending on the same night. All estimation is by OLS. Robust standard errors in parentheses. $+ p < 0.10$, $* p < 0.05$, $** p < 0.01$

Figure 1. Map of the four largest Harvard Medical School affiliates

Figure 2. Geographic Distance Between Pairs



Notes: Distances between pairs of researchers in the sample were calculated by geocoding the exact location of their offices and calculating the intervening distance in miles.

APPENDIX: Additional Tables and Figures

Table A1. Summary Statistics, Harvard Medical School vs. Attendees

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

Notes: See Section III in the text for a description of the data. Stars indicate the results of t-tests for equality of means. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Table A2. Main effect of Treatment on Collaboration – Estimated with Probit

DV = Collaboration	(1)	(2)	(3)
Same room	0.0011^{+}	0.0011^{+}	0.0012*
	(0.0006)	(0.0006)	(0.0005)
One postdoc			-0.0008
			(0.0006)
Both postdocs			-0.0019
			(0.0017)
One is female			0.0002
			(0.0006)
Both are female			0.0006
			(0.0009)
Same hospital			0.0030**
			(0.0006)
Both Longwood			0.0001
			(0.0006)
One imager + one clinician			0.0012
			(0.0007)
Both imagers			0.0022**
			(0.0008)
Same clinical area (SOI)			0.0026**
			(0.0006)
Previous coauthor			0.0079**
			(0.0016)
Night fixed effects	No	Yes	Yes
Pseudo R2	0.004	0.008	0.155
Nb. of Obs.	26,664	26,664	26,664

Notes: Dependent variable is *Collaboration*, an indicator variable for whether the pair appeared on any common pilot grant or concept award applications. The main variable of interest is *Same room*, which was randomized across pairs attending on the same night. All estimation is by probit, reporting marginal effects. Robust standard errors in parentheses. $^{+}$ *p < 0.10,* * *p < 0.05,* ** *p < 0.01*

Table A3. Treatment and Interactions with Measures of Distance – Estimated with Probit

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Same room	0.0005	-0.0001	0.0013	0.0014*	0.0020	0.0000	0.0010+	-0.0002
	(0.0007)	(0.0008)	(0.0008)	(0.0007)	(0.0013)	(0.0007)	(0.0006)	(0.0018)
One postdoc	-0.0021*							-0.0015*
	(0.0009)							(0.0008)
Same rm X One postdoc	0.0020							0.0018
	(0.0013)							(0.0012)
Both postdocs	-0.0023							-0.0013
	(0.0018)							(0.0017)
Same rm X Both postdocs	n/a							n/a
	n/a							n/a
One is female		-0.0009						-0.0004
		(0.0007)						(0.0007)
Same rm X One female		0.0019						0.0020
		(0.0012)						(0.0013)
Both are female		-0.0009						-0.0003
		(0.0014)						(0.0013)
Same rm X Both female		0.0029						0.0021
		(0.0020)						(0.0019)
Same hospital			0.0037**					0.0030**
			(0.0008)					(0.0007)
Same rm X Same hospital			-0.0003					0.0003
			(0.0011)					(0.0011)
Both Longwood				0.0005				0.0005
				(0.0007)				(0.0007)
Same rm X Both Longwood				-0.0017				-0.0015
				(0.0014)				(0.0014)
One imager + one clinician					0.0017+			0.0016+
					(0.0010)			(0.0010)
Same rm X One imager					-0.0010			-0.0011
					(0.0015)			(0.0015)
Both imagers					0.0034**			0.0029**
					(0.0011)			(0.0011)
Same rm X Both imager					-0.0014			-0.0019
					(0.0016)			(0.0016)
Same clinical area (SOI)						0.0021**		0.0016*
						(0.0008)		(0.0008)
Same rm X Same clin area						0.0026*		0.0028*
						(0.0012)		(0.0012)
Previous coauthor							0.0096**	0.0071**
							(0.0019)	(0.0017)
Same rm X Prev coauthor							0.0038	0.0043
							(0.0035)	(0.0032)
Night fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R2	0.027	0.014	0.069	0.001	0.001	0.045	0.061	0.175
Nb. of Obs.	26,184	26,664	26,664	26,664	26,664	26,664	26,664	26,184

Notes: Dependent variable is *Collaboration*, an indicator variable for whether the pair appeared on any common pilot grant or concept award applications. The main variable of interest is *Same room*, which was randomized across pairs attending on the same night. All estimation is by probit, reporting marginal effects. Robust standard errors in parentheses. $^{+}$ *p < 0.10,* * *p < 0.05,* ** *p < 0.01*

Table A4. Application quality for within-rooms and across-rooms proposals

Sample Means	*Collaborations Within rooms*	*Collaborations Across rooms*	*Difference*
Application quality (scores from the peer review)	3.91	4.00	-0.0912
Observations	17	30	

Notes: Stars indicate the results of t-tests for equality of means. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Figure A1. Examples of Participant Posters

[Name]

Massachusetts General Hospital
[Email]

Can we leverage the benefits of PET/MRI (e.g. non-ionizing radiation) with novel PET-labeled small molecule drugs for predicting and/or serially monitoring response in ovarian or endometrial cancers?

Why does it matter?
Reliable non-invasive prediction and monitoring of tumor response would obviate repeated biopsies and pave the way for rationale treatment selection and enrollment into appropriate clinical trials -- vast implications for treating providers and drug companies.

What do you need for your research to succeed?
Individuals with experience translating novel imaging agents into clinically viable research tools. By extension, expertise developing and interpreting first-in-human imaging studies using radio tracers.

Room: 205, Group 1, Participant ID: 01_504

[Name]

Children's Hospital Boston
[Email]

What is the relationship between white matter microstructure, tuber quantification measures and neurological phenotype (autism, seizures, cogition) in patients with Tuberous Sclerosis Complex?

Why does it matter?
Specifically, an advanced imaging biomarker would enable early identification of patients at high risk for autism, and allow for early therapeutic intervention to improve neurological outcome. Also, such physiologic imaging techniques will provide more insight into the biological basis of autism.

What do you need for your research to succeed?
A reliable computer algorithm capable of independent, automated segmentation of tubers and characterization of white matter tracts based on advanced imaging. To study the images of our unprecedented large sample in an objective and consistent manner, a processing pipeline would also be needed.

Room: 206, Group 1, Participant ID: 01_533

Notes: Names and emails have been removed to maintain the privacy of participants.

Figure A1. Examples of Participant Posters (cont'd)

[Name]

Beth Israel Deaconess Medical Center
[Email]

Microcirculatory dysfunction has been shown to be an indicator of patient outcomes in septic shock. Is microcirculatory dysfunction a reliable predictor of other cardiovascular and immunological disorders?

Why does it matter?
Hypoperfusion to various tissues may occur due to a maldistribution of vascular flow. Bedside assessment of tissue hypoperfusion may improve identification of severity of illness and outcomes, particularly in subtle cases. Sublingual mucosa provide non-invasive optical access to microcirculation.

What do you need for your research to succeed?
We use sublingual Sidestream Darkfield (SDF) imaging to measure clinical microcirculatory blood flow rates and perfused vessel density in septic patients. We are looking for a partner who has expertise in clinical studies of cardiovascular and/or immunological disorders.

Room: 206, Group 2, Participant ID: 01_555

[Name]

Brigham and Women's Hospital
[Email]

Following partial lung loss from injury or surgery, does the remaining lung expand through hyperinflation or through new growth and regeneration of functional gas exchange surfaces? Is this tissue response mediated through mechanical stretch associated with breathing?

Why does it matter?
If new growth of functional lung is possible in adult humans (which has yet to be demonstrated save for one patient examined by our group), and if the humoral and mechanical factors promoting such growth could be identified, this would lead to a transformation in therapy for patients with lung loss.

What do you need for your research to succeed?
We have identified a cohort of patients scheduled for pneumonectomy, and a validated protocol for assessing lung microstructure and function through innovative 129Xe MRI. To catalyze this research effort, we need funding for magnet time and supplies necessary for this noninvasive imaging modality.

Room: 204, Group 1, Participant ID: 01_567

Notes: Names and emails have been removed to maintain the privacy of participants.