LIKE STARS: HOW FIRMS LEARN AT SCIENTIFIC CONFERENCES

*

Stefano Baruffaldi[ac⋆] Felix Poege[bc⋆]

[a] *School of Management, Polytechnic University of Milan, 20133 Milan, Italy*
[b] *Department of Management and Technology and ICRIOS, Bocconi University, 20100 Milan, Italy*
[c] *Max Planck Institute for Innovation and Competition, 80539 Munich, Germany*
[⋆] *Corresponding authors*

24 May 2024

Forthcoming in Management Science

ABSTRACT

Scientific conferences are an underexplored channel by which firms can learn from science. We provide empirical evidence that firms learn from scientific conferences in which they participate but also that this is conditional on intense participation. Using data from conference papers in computer science since the 1990s, we show that corporate investments in participation are both frequent and highly skewed, with some firms contributing to a given conference scientifically, some as sponsors, and some doing both. We use direct flights as an instrumental variable for the probability that other scientists participate in the same conference as a firm, altering the knowledge set to which the firm is exposed. We find that a firm's use of scientists' knowledge increases when they participate in the same conferences. Greater participation efforts, where the firm seeks the spotlight by both sponsoring the conference and contributing to its scientific discourse, foretell research collaborations and a stronger learning effect. Such learning is disproportionately concentrated among the most prominent firms and scientists rather than benefitting those without alternative interaction channels. Therefore, on average, firms learn from scientists that they encounter at conferences, but the substantial heterogeneity of the effect reflects the influence of reputation mechanisms in social interactions.

KEYWORDS: Science, Corporate Science, Conferences, Learning, Innovation

*Contact: stefano.baruffaldi@polimi.it, https://orcid.org/0000-0003-4022-7691 (SB); felix.poege@unibocconi.it, https://orcid.org/0000-0002-9826-6581 (FP).
The authors are grateful for helpful comments to Christian Fons-Rosen, Fabian Gaessler, Dietmar Harhoff, Matt Marx, Rossella Salandra, Ammon Salter, Tim Simcoe, Florenta Teodoridis, Anne Ter Wal, Stefan Wagner, and Martin Watzinger; special thanks go to Markus Simeth for his help since the project's early stages. The authors are also grateful to editor Ashish Arora, the associate editor, and four anonymous referees for insightful and constructive comments that greatly improved the quality of this work. The authors then thank the participants of the REER, AOM, DRUID, EPIP, SMYE, and Innopat conferences and seminars at LMU and TU Munich, CREI-Bath, SKEMA, University of Bath, Boston University, Tilburg University, Polytechnic University of Milan, Queens Mary University, and Imperial College London for their comments.

1 Introduction

Operating at the scientific frontier can foster innovation but entails costs that are less and less sustainable for the average firm (Nelson, 1982; Ahmadpoor and Jones, 2017; Rosenberg, 1990; Hicks, 1995; Andrews et al., 2015). Research investments and scientific engagement are necessary to build absorptive capacity (Cohen and Levinthal, 1990), explore the scientific landscape (Fleming and Sorenson, 2004; Arora and Gambardella, 1994), and gain a first-mover advantage on new findings (Arora et al., 2023). As firms increasingly rely on extramural scientific knowledge (Arora et al., 2018), they also need to deploy diverse search strategies (Laursen and Salter, 2006). These include, for instance, hiring (Audretsch and Stephan, 1996), licensing (Dechenaux et al., 2008), collaboration (Cockburn and Henderson, 1998), geographic location (Alcacer and Chung, 2007; Belenzon and Schankerman, 2013), and selective attention (Bikard and Marx, 2020). Conferences are a landmark of modern science (Crane, 1974; Mokyr, 2011) and may provide another bridge toward the knowledge frontier. Examples show that some firms make large participation investments in order to feature prominently with scientific contributions and as conference sponsors.[1] Yet, this phenomenon has received little attention.

There are reasons to expect that scientific conferences nurture the innovation activities of participating firms, but empirical evidence is limited, and it is unclear which firms gain the best learning opportunities. In surveys, firms report meetings and conferences as critical for accessing public R&D (Cohen et al., 2002; ZEW, 2018). However, most evidence on learning within scientific conferences exists for the case of academic scientists (Chai and Freeman, 2019; Lopez de Leon and McQuillin, 2020; Boudreau et al., 2017).[2] Some studies depict conferences as markets for ideas that leverage temporary colocation to reduce search costs or as information repositories that facilitate spillovers, especially for participants lacking alternative channels (Maskell, 2014; Hansen and Pedersen, 2018; Vlasov et al., 2016). Others imply that knowledge sharing within scientific communities adheres to specific institutional norms and may remain foreclosed to most firms or put them at risk of unintended spillovers (Dasgupta and David, 1994; Gittelman and Kogut, 2003; Haeussler, 2011; Arora et al., 2021). Finally, participation investments may serve firm objectives other than corporate learning, such as obtaining marketplace certification (Polidoro and Theeke, 2012) or satisfying their scientists' "taste for science" (Stern, 2004).

We provide evidence that the presentation of a paper at a conference where a firm participates increases the probability that the firm utilizes related knowledge. However, this is contingent on intense firm participation at the conference and is more likely when both the participating firm and the presenting scientists are prominent in their field. These findings support

[1] For instance, at the 2017 Conference on Neural Information Processing Systems (NeurIPS, previously NIPS) – a leading academic machine-learning conference – Google featured as a sponsor and was the most represented affiliation among the papers, with 75 to its name; firms such as Microsoft, IBM, and Tencent followed close behind (see: https://nips.cc/Conferences/2017). In the 2016 Community Innovation Survey, conferences were rated as 'highly important' information sources by 13% of innovative German firms (ZEW, 2018).

[2] One notable exception is Vlasov et al. (2016), who relate small and medium firms' patenting to the diversity of conferences they attend. Others study non-academic events, such as an app developers' conference (Foerderer, 2020) and open standards community meetings (Waguespack and Fleming, 2009). Perkmann et al. (2021) find only a few studies in the academic engagement literature that mention scientific conferences.

an understanding of scientific conferences as social spaces governed by reputation mechanisms (Dasgupta and David, 1994; Cetina, 1999). While proximity facilitates face-to-face encounters, learning requires social interactions within scientific communities (Cockburn and Henderson, 1998). Under this perspective, "starring" firms with a strong presence and reputation access the best learning opportunities. To this end, scientific contributions generate interaction occasions, signal quality, and demonstrate compliance with scientific norms; conference sponsorship is a complementary symbolic contribution and provides access to physical space, increasing legitimacy and visibility (Fosfuri et al., 2015; Rao, 1994; Rindova et al., 2005). By the same token, knowledge sharing is concentrated among prominent firms and scientists.

To test these arguments, we employ a quasi-experimental empirical framework to isolate exogenous variation in the knowledge that firms encounter at the conferences in which they participate. We then estimate the effect of such exposure on their use of this knowledge and study how different participation modes and the characteristics of firms and scientists moderate this effect. The empirical challenge is to observe how firms participate in multiple conferences and identify a causal effect of the exposure to the associated knowledge. To this end, we leverage computer science (CS) proceedings data as a "paper trail" of conferences and their participants. The data we assemble allow us to compare papers at conferences where firms participate with papers from other conferences of similar quality and subject. We then use the availability of direct flights as a source of exogenous variation in the probability that these papers are presented at a given conference.

In addition to its economic significance, CS provides an ideal setting because its conference papers are well covered in the available data; this allows us to capture – on a large scale – firms' conference participation as well as their use of the associated knowledge (Franceschet, 2010). Our data cover all relevant CS conferences from 1996 to 2010, for a total of 6,846 conferences from 982 conference series, with 4,800 participating firms. We derive the modes of a firm's participation from the number of conference papers affiliated with the firm and from information on conference sponsors.[3] We use patent references to papers as evidence of firms using associated knowledge (Roach and Cohen, 2013). Further, we add insights from interviews at two important conferences to inform our understanding of corporate participation.[4]

Descriptive evidence demonstrates that although many firms invest in numerous conferences, their modes of participation vary substantially. Worldwide, 26.7% of conferences are sponsored by firms, and 89.7% have firm authors. However, these participation investments are highly skewed. Roughly 10% of firms account for 90% of the papers and sponsorships associated with firms. Moreover, firms favor conferences of the highest scientific quality, and the most active firms often contribute several papers to a conference, where they may also be sponsors. Interestingly, firms' papers are often highly cited by other scientists. Finally, we find prima facie evidence for the relevance of conferences to innovation. Approximately half of all firms' patent citations of the CS literature refer to conference papers, and in turn, almost a third of these citations link to papers presented at conferences in which the patenting firm participated.

[3]We observe active but not passive participation. We discuss this limitation in Sections 3.1 and 7.

[4]We visited the 2018 ECCV conference in Munich and the 2019 NeurIPS conference in Vancouver.

We implement our empirical strategy in two steps. First, we build an estimation sample based on a case-control approach wherein we match papers presented at a conference where a firm participated with a set of papers that could have been presented at the same conference but were presented elsewhere. Specifically, this sample includes firms paired both with papers from conferences in which they participated and papers of other conferences of similar topic, quality rank, and year (cf. Jaffe et al., 1993). We then estimate the difference in probability of citation by the firm in an ordinary least-squares (OLS) regression framework with fine-grained fixed effects (FEs) and covariate controls (Singh and Marx, 2013). Second, we introduce direct flights as an instrumental variable (IV) to address unobserved characteristics of the papers that may explain firms' citations. We take a firm's participation choice as a given but consider that transportation costs may exogenously influence the participation choices of the authors of other papers (Catalini et al., 2020; Fleming and Waguespack, 2007). Accordingly, we use the presence of direct flights from scientists' locations to conference venues as an IV.

Our results demonstrate that exposure to knowledge at a conference greatly increases the probability that a firm utilizes it, and the effect varies strongly according to the mode of the firm's participation. First, according to both OLS and IV analyses, a firm is more likely to cite a paper presented at a conference in which it participated. The IV estimates are larger in magnitude than those using OLS. We discuss possible explanations, including considerations of the local average treatment effect (LATE), potential sources of negative bias, and measurement-error bias in the OLS analysis. Second, increased firm participation – a higher number of papers affiliated with a firm at a given conference and sponsorship of the conference – enhances the baseline effect. Firms with minimal scientific contributions and no sponsorship or with sponsorship but without scientific contributions do not experience the effect.

Other results further suggest that social interactions and reputation are key for learning. We also find positive effects on firms' scientific citations, on citations of prior work by conference papers' authors, on collaborations, and on hiring. These effects are all either stronger for firms with intense participation or even confined to such firms. Second, the effects are stronger for firms that are better positioned in the scientific landscape: those already more engaged in science, specialized in research similar to the papers, and located in scientific hubs. Third, the effects are stronger if the papers' authors are more productive and male and hence, more visible and socially advantaged in an environment where most researchers are male. In summary, these results suggest that the effect we observe on corporate learning is concentrated between prominent firms and scientists rather than between participants without alternative interaction channels. Qualitative insights also support the role of social interactions and reputation.

Our study makes two main contributions. First, our findings have implications for R&D management decisions on participation in scientific and, more generally, open innovation communities (Waguespack and Fleming, 2009; Maskell, 2014; Foerderer, 2020). In these contexts, the roles of social interactions and reputation imply that focused and significant efforts, rather than dispersed and limited investments, are necessary. Second, we contribute to the ongoing debate on corporate science and learning (Simeth and Raffo, 2013; Arora et al., 2018, 2023).

In this respect, we establish scientific conferences as a channel to access scientific knowledge. Importantly, the results suggest that knowledge access remains tied to reputational mechanisms and defies the notion that knowledge spillovers at conferences are free and evenly distributed. This adds to concerns that firms may face steep investment curves in accessing frontier knowledge, resulting in divergence and concentration, rather than convergence, in innovation and productivity (Andrews et al., 2015; Autor et al., 2020b; Akcigit and Ates, 2023).

2 Conference participation and corporate learning

2.1 Learning at scientific conferences

The existing literature emphasizes two main mechanisms that facilitate learning and their implications for firms participating in scientific conferences. On the one hand, temporary colocation compensates for geographic distance and knowledge dispersion, enabling firms to navigate scientific results more effectively. While scientific knowledge is published, face-to-face conversations remain essential for conveying tacit knowledge and gaining access to information that has not yet been made public. At the same time, because science is cosmopolitan and distributed across diverse organizations (Merton, 1973), physical barriers and the ever-increasing burden of knowledge make it more difficult to identify and access external ideas (Jones, 2009; Wagner et al., 2014; Jaffe et al., 1993; Singh and Marx, 2013). In this context, conferences constitute repositories of information otherwise scattered across dispersed and possibly unknown locations (Maskell, 2014; Bathelt and Cohendet, 2014). As such, they serve as temporary markets for ideas that reduce search costs, while proximity facilitates knowledge spillovers (Chai and Freeman, 2019; Boudreau et al., 2017; Lopez de Leon and McQuillin, 2020).

On the other hand, learning requires social interactions and a favorable reputation within communities dedicated to knowledge production. Scientific conferences bring individuals together based on common interests, foster systematic knowledge sharing, and shape the development of networks of influence and collaboration (Crane, 1969; Cetina, 1999). These networks are the loci of knowledge production, and learning derives from active participation therein (Powell et al., 1996; Lee and Cole, 2003; Fleming and Waguespack, 2007).[5] In fact, scientific communities operate in the absence of pure market mechanisms, with a reputation-based system governing knowledge sharing in social interactions (Stephan, 1996; Dasgupta and David, 1994). Reputation serves as a readily accessible quality cue in the search for interlocutors, which affects the probability of interactions (Merton, 1973; Azoulay et al., 2014). At the same time, it shapes expectations for reciprocity, thereby influencing knowledge sharing (Mukherjee and Stern, 2009; Haeussler, 2011), future collaborations, and mobility, which are all essential avenues for learning (Zucker et al., 2002; Almeida and Kogut, 1999; Breschi and Lissoni, 2009).

Both of these mechanisms provide conference participants with an advantage in accessing related knowledge. However, the relevance of social interactions and reputation mechanisms implies that specific efforts toward active participation are necessary. Scientific conferences are

[5] The notion that knowledge production resides in distributed collaboration networks is common to the literature on both scientific and open innovation communities: "The institution of science might be considered the first open innovation community" (Dalle and David, 2003; Fleming and Waguespack, 2007, p. 166).

social spaces rather than exclusively markets for ideas or information repositories. Learning opportunities at them are facilitated by co-location but ultimately dependent on effective interactions. To paraphrase Cockburn and Henderson (1998, p. 165, emphasis in original), "it is not simply [attending conferences] that reshapes private sector problem solving but *participation* in the intellectual life of the wider scientific community, [hence] a focus on interactions is particularly important."[6] Accordingly, access to the most valuable knowledge is unlikely to derive from spillovers "in the air." Rather, investments in active participation and reputation are necessary to increase the quantity and quality of social interactions and bidirectional knowledge sharing.

2.2 Modes of participation and learning

Because of the learning mechanisms specific to scientific conferences, we anticipate that corporate learning at them is conditional on investments in different modes of participation to gain reputation and maximize social interactions. First, scientific contributions increase opportunities for social interaction. Each contribution is an occasion for debate between firm-affiliated and other scientists on common topics. Contributions also strengthen reputation. In economic terms, establishing a reputation requires the provision of costly signals that create positive expectations as to quality and future behavior (Fombrun and Shanley, 1990; Rao, 1994). In this sense, scientific contributions broadcast specific competence and compliance with scientific norms (Dasgupta and David, 1994; Hicks, 1995). This elicits reciprocity, renders personal interactions more fruitful, and creates positive conditions for future collaboration (Alexy et al., 2013; Haeussler et al., 2014; Haas and Park, 2010; Tartari et al., 2014).

Second, conference sponsorship helps increase legitimacy and visibility (Rao, 1994; Rindova et al., 2005). Legitimacy, defined as the "assumption that an organization is desirable, proper and appropriate within a widely shared system of norms and values" (Rao, 1994, p. 30), derives from the use of symbols rather than signals. Sponsors support the conference organization financially and typically appear in iconic categories stratified according to their contributions (e.g., Silver, Gold, Diamond).[7] Sponsorship creates a symbolically valuable association between a sponsor's image and the objectives and values of the organization they support (Fosfuri et al., 2015). In addition, sponsoring firms are granted physical spaces at the event to disseminate information, promote their activities, and interact with other participants (e.g., permanent booths or tutorial sessions). Such activities yield additional interaction opportunities and increase a firm's visibility (Rindova et al., 2005).

Although authorship and sponsorship are different processes, we expect them to reinforce each other and yield increasing returns on investment. They differ to the extent that authorship

[6]In the original formulation the authors refer to "publication" rather than "attending conferences." They convey the idea that solely publishing scientific papers is insufficient for enabling corporate learning, and active participation in external knowledge networks is necessary.

[7]Fees range from a few thousand US dollars to as much as $80,000, depending on the conference and sponsorship category. For instance, the 2017 NeurIPS conference offered five sponsorship categories: Diamond ($80,000), Platinum ($40,000), Gold ($20,000), Silver ($10,000), and Bronze ($5,000). In 2017, the conference attracted 84 sponsors with contributions totaling $1.76 million, an increase of 31.5% from the previous year, in which 64 sponsors contributed to a total of $840,000. Source: https://medium.com/syncedreview/a-statistical-tour-of-nips-2017-438201fb6c8a

may reflect the interest and focus of firms' research units, whereas sponsorship likely involves marketing and human resources (HR) (Fosfuri et al., 2015; Polidoro and Theeke, 2012). However, first of all, signals and symbols are complementary in the quest to gain reputation (Rao, 1994): A lack of scientific contributions may void the symbolic component of sponsorship for scientific communities, whereas scientifically active firms can use it to highlight their contributions. In turn, improved legitimacy from sponsorship can spill over to affiliated scientists, especially if they are active at the conference (Dahlander and Wallin, 2006). Second, reputation mechanisms in science lead to status-based social stratification (Cole and Cole, 1973; Merton, 1973; Sorenson, 2014). Starring organizations gain greater influence and a privileged position that allows them to increase both the quantity and quality of their social interactions, leading to a higher likelihood of collaborations and hiring. For these reasons, corporate participation is likely more effective if associated with significant and coordinated investments that attract collaborators.

2.3 Participants' heterogeneity and learning concentration

A view of conferences as social spaces also implies that learning may be concentrated among prominent firms and scientists rather than significantly extending to those lacking alternative ways of accessing knowledge. While previous studies have isolated specific determinants of access to knowledge, we seek to uncover a general pattern of how these factors interact with participation in the same conferences. In principle, factors that facilitate the use of knowledge outside of conferences, such as proximity (Belenzon and Schankerman, 2013) or attention drivers (Bikard and Marx, 2020; Bikard, 2018), may be substitutes for conferences. In other words, conferences may be less relevant for geographically well-positioned participants and for accessing content that already attracts attention. Conversely, these same characteristics may complement conference participation if they facilitate social relations and future collaborations. Conferences may stage interactions that are influenced by reputation mechanisms and complementarities and that are, therefore, fostered by the same factors that aid knowledge access in general.

Specifically, we posit that firm and scientist characteristics that help attract attention and build reputation will positively moderate the effect of knowledge exposure at a conference. From the scientists' perspective, firms better placed in the scientific landscape, those that strongly invest in research or are located in research hubs, will arouse more interest. Furthermore, geographic and research proximity can constitute a common ground for turning serendipitous encounters into productive interactions (Lane et al., 2021). From the firms' perspective, scientists that enjoy greater visibility – e.g., owing to their productivity or affiliation (Bikard and Marx, 2020; Bikard, 2018) – are more likely to attract attention, at conferences as well as elsewhere. Finally, several studies note that female researchers are overlooked in commercialization and engagement with industry (Bikard and Fernandez-Mateo, 2022; Koffi and Marx, 2022; Tartari and Salter, 2015); because informal social spaces also favor masculine appearance in science (Hansen and Pedersen, 2018), this tendency may manifest at conferences, too.

3 Data

3.1 Proceedings as conference paper trails

CS is an ideal setting for studying the participation of firms in scientific conferences for two main reasons. First, it is a field of great technological and economic relevance (Brynjolfsson and Hitt, 2003; Newell et al., 1967). CS publications are frequently cited in patents (Ahmadpoor and Jones, 2017), and the success of machine learning in recent years has significantly increased industry involvement (Hartmann and Henkel, 2020). Second, CS conference proceedings – collections of conference papers – are well covered in bibliographic data, enabling large-scale observation of key conference information, the papers presented, and their authors. This is because in CS, conferences represent a primary outlet for scientists. Acceptance of submissions is based on peer review and, for the most prestigious conferences, is highly competitive (Franceschet, 2010). Conference papers are original contributions and authors are very likely to also be conference participants. In fact, most conference organizers make paper publication conditional on the physical participation of at least one author.[8] Thus, these data serve as a paper trail for scientist participation. In addition, some databases also attempt to trace conference sponsors, which we use to identify corporate sponsorships.

Nevertheless, some limitations are noteworthy. The coverage of sponsors is not exhaustive and the related descriptive numbers remain underestimated.[9] More importantly, besides what can be gleaned from authorship, the data do not include actual lists of participants, which creates two main issues. First, we cannot know exactly which author(s) of a given paper participated in the conference where it was presented, although we can be confident that at least one did. The second issue is that we cannot observe *passive* attendance, which means that firms may encounter scientists we do not observe, who attend but are not authors on any accepted paper. Alternatively, the firms themselves may attend passively, neither presenting papers nor sponsoring. We return to these issues more extensively in our robustness analyses.

Finally, while the CS setting is particularly advantageous for our analysis, the results can also be informative for other fields. First, conferences can be just as crucial for scientists in other fields (Hauss, 2021). Second, Cohen et al. (2002) report survey results in which conferences score highly in importance across several industries, besides computing and semiconductors. We used data on these other fields to verify that the rate of corporate participation at their conferences is similar to what we observe in CS.[10]

[8] In manually collected data for a subsample of 1,021 conferences, we found average acceptance rates of 21% and 36%, respectively, at $A^\star$ and other conferences. The IEEE guidelines provide an example of policies to guarantee participation: www.ieee.org/conferences/organizers/handling-nonpresented-papers.html

[9] Of the conferences in our final sample, 80% list sponsorship information; completeness is heterogeneous, and our descriptive analysis of sponsorship remains conservative.

[10] In Scopus, the share of conference papers with corporate affiliations is 7.5% in CS, which is lower than for physics (9.7%), engineering (11.3%), or chemistry (15%), and slightly higher than for biochemistry/genetics (5.2%). Note that the more limited coverage of conferences in other fields impedes precise comparisons. However, consistent with the Scopus data, Cohen et al. (2002) found that 37.9% of respondents in computing and 48% in semiconductors indicated conferences to be important. The rates are similar in several other industries, e.g., 45.5% in machine tools, and even higher in petroleum (50%), drugs (64%), steel (54.6%), and aerospace (51%).

3.2 Data sources and matching

We obtain information on conferences, the conference series they belong to, papers listed in conference proceedings, and other scientific publications such as journal articles, from the dblp computer science bibliography (dblp).[11] We complement dblp data with Web of Science (WoS) and Scopus data on authors' affiliations, conference sponsors, citations, and abstracts. We incorporate information on conference series quality and CS research subfields curated by the Computing Research and Education Association of Australasia (CORE). Based on expert assessment, CORE classifies all relevant CS conference series into subfields and the quality-rank levels $A^{\star}$, A, B, and C. The coverage of Scopus data available to us (starting in 1996) and the need to observe citations after conferences confines the period of observation for conferences to between 1996 and 2010.

We measure innovative activity through patent data from the European Patent Office's (EPO's) PATSTAT database and trace patent-to-paper references to conference papers. To this end, we rely on a match between front-page citations in U.S. Patent and Trademark Office (USPTO) and EPO patent (applications) and World Intellectual Property Organization (WIPO) applications and science described in Knaus and Palzenberger (2018). Moreover, recent data contributions allow us to consider references appearing in the full text of patent technology descriptions ("in-text"). Among these, we use PatCit by Verluise and De Rassenfosse (2021) to complement the data with in-text references from USPTO patents, but our results are also robust to their exclusion. We consolidate the USPTO/EPO/WIPO patent data at the DOCDB family level and allocate them over time according to the priority year, maintaining families with priority years 1996 through 2015. Because of the lag between when patents are filed and when they are granted, we limit the sample of conferences to the period ending in 2010 in order to observe patents in the five years following a conference.

We capture firms active at CS conferences, starting from a comprehensive list of firm names, compatible with computational feasibility. We rely on lists of the world's top R&D performers (from the EU's innovation scoreboards), firms known to participate in science (from the Global Research Identifier Database, GRID), and firms known to be active in technology development (firms in Bureau van Dijk's Orbis database that are linked to patents). To further broaden our coverage, we also include any other US- or Germany-based firms for which Orbis has financial information, also without patents.[12] We then match firms to affiliations, conference sponsors, and patent applicants using a supervised machine-learning algorithm. This method combines string-similarity with similarity in the list of websites obtained by searching names online (see Autor et al., 2020a). With the same algorithm and manual inspection, we aggregate subsidiaries

[11]See dblp.org. For clarity, henceforth, we use "conference series" to indicate collections of conferences. For example, the Annual Conference on Neural Information Processing Systems (NeurIPS, previously NIPS) conference series takes place each year at varying locations. Conference events held in a specific location on a specific date (e.g., 7–10 December, 2009, in Vancouver, CAN) we refer to simply as "conferences." Each conference is, via conference proceedings, linked to a set of published conference papers, which we refer to simply as "papers." We contrast this with "articles" published in journals.

[12]These firms account for less than 5% of the final estimation sample, which increases our confidence that we capture the vast majority of corporate participation at CS scientific conferences. Their inclusion does not significantly bias our estimates. See also Appendix D.2.

at the level of the corporate group. Finally, we conduct substantial manual post-processing to further clean the data. A manual validation provides confidence in the quality of the automated matching.[13] To ensure that noise in the data does not bias the results, we verified that all the results presented hold when we focus only on companies whose main location is in the US; US-based companies are on average larger, and the data processing is most accurate for them.

To obtain information on direct flights, we purchased data on airports and flight connections from the International Civil Aviation Organization (ICAO). Because ICAO only covers international flights, we add data for domestic US flights from the Airline Origin and Destination Survey of the US Bureau of Transportation Statistics (BTS). We geolocalize all conference venues, scientist affiliations, and firm information, and map them to about 1,000 airport regions. In cases where multiple airports are near one another, we select the busiest. For papers, we focus on the affiliations of the first author to assign each paper to one airport. Our results prove robust to the use of all affiliations associated with a paper.

Finally, we gather a host of qualitative insights that we use to describe corporate participation activities in greater detail, and to inform our empirical analysis of the associated mechanisms. Thus, we attended two major conferences: the 2018 European Conference on Computer Vision (ECCV18) in Munich, Germany, and the 2019 Conference on Neural Information Processing Systems (NeurIPS19) in Vancouver, Canada. We interviewed more than 50 individuals in total, both scientists and other representatives from more than 20 firms, and a further 20 or so academic scientists. We provide additional details in Appendix C.

3.3 Measuring corporate participation in scientific conferences

Central to our approach is the use of bibliometric data to capture the participation of firms in conferences and to proxy for different modes of participation. We use affiliation information on conference papers to link scientists to firms and count the number of papers at a conference authored by a focal firm's scientists as a measure of its scientific contribution. Sponsorship information derives directly from the list of conference sponsors. Our interviews (at ECCV and NeurIPS) served to validate our indicators and aid the understanding of the different modes of participation. As noted, we cannot exclude the possibility that only co-authors who are academic scientists participated in a conference or that additional firms participated passively (i.e., without sponsorship or paper authorship). However, at the conferences we attended, the presence of firms was strong and appeared to accurately reflect information in the bibliometric data: More papers and higher levels of sponsorship aligned with a stronger presence at the event. Corporate participation, as we observe it in the data, is associated with highly heterogeneous

[13] We drew 500 random conference papers: in these, 80 unique affiliations were private organizations, of which 11 were very small and local and not on our list of firms; of the remaining 69, 61 were correctly matched while 8 were not matched at all.

and, in some cases, large investments.[14]

Qualitative evidence also provides initial confirmation that scientific contributions and sponsorship differ in their rationales and organizational processes but share synergies and a common learning objective. In terms of contributions, corporate scientists presented their work and routinely interacted with their peers. Most scientists we interviewed said they enjoyed the freedom to participate in conferences and implied that authorship was the outcome of seemingly autonomous decisions within research units. Scientists from smaller firms described similar freedoms as those from larger firms, albeit with tighter budget constraints. In contrast, sponsorship involves HR and marketing personnel, who coordinate corporate branding on the conference premises and website and disseminate information, especially about hiring. However, most firms noted a strong integration of the two processes. In particular, sponsorship decisions are often based on what events firm scientists see as most relevant. At the conferences, sponsorship also serves to promote a firm's science and, in some cases, sponsors organize workshops and social events. We provide further considerations in Section 6.2.

3.4 Measuring the use of scientific knowledge

In our main analyses, we focus on patent citations of papers, which we interpret as an indication of the use of related knowledge in pursuit of commercial value. In this, we follow a growing literature that has deployed patent-to-paper references to proxy for the contribution of scientific knowledge to new technologies (Roach and Cohen, 2013; Ahmadpoor and Jones, 2017; Bikard and Marx, 2020). Patent-to-paper citations correspond well with managers' reporting on the use of public R&D (Roach and Cohen, 2013) and reflect inventors' awareness of academic literature (Bikard and Marx, 2020), which distances them from the typical concern that the addition of citations by attorneys and patent examiners would confound the measure (Alcacer and Gittelman, 2006). Likewise, although patents are imperfect proxies of innovation and vary greatly in value, patents that cite science are consistently found to be of high value (Ahmadpoor and Jones, 2017; Poege et al., 2019; Schnitzer and Watzinger, 2019). However, in later sections we also employ other variables to capture the use of scientific knowledge in the firms' own research, or specific channels of knowledge access such as scientific collaborations and hiring.

4 Descriptive statistics

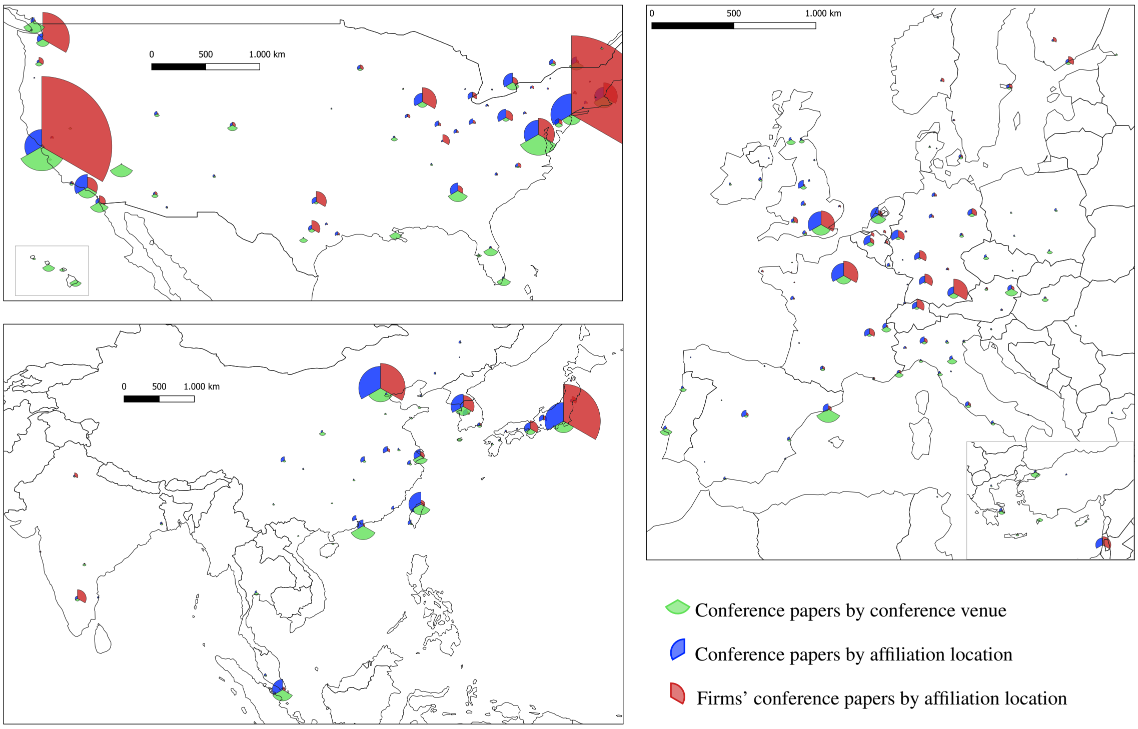
4.1 Main descriptives and geographic distribution

Our data are representative of the most relevant worldwide CS conference series in the period considered, between 1996 and 2010. The dataset covers 982 conference series, 6,846 conferences, and 559,127 conference papers. A total of 4,800 firms participated in at least one conference. Firms, conference venues, and scientists are distributed around the globe. As illustrated in Fig-

[14] As an example, consider the participation of Google at NeurIPS 2017 in Long Beach, California. It figured as a second-tier sponsor ($40,000) and was affiliated with 75 conference papers authored by 86 distinct scientists. Taking conservative assumptions of 43 scientists and 5 HR employees participating, travel costs of $130 per person, accommodation and expenses of $200 per person for each of the six days of the conference, and an opportunity cost per employee proportional to an average yearly wage of $120,000 for 260 working days, this level of participation amounts to an investment of approximately $260,000.

ure 1, almost 50% of all conference papers with corporate affiliation were produced by scientists located in North America, with 29% in Europe and 21% in Asia. Interestingly, conference locations and scientist distributions follow different patterns. Similarly, while corporate locations are disproportionately concentrated, scientists are more evenly distributed around the world. Furthermore, in North America and Europe, conferences tend to be located in more southern, touristically attractive locations.

Figure 1: Geographic distribution



Notes: Shows the geographic distribution of the share of papers by conference venues (lower circle segment), by paper affiliations (left circle segment), and of affiliations on firm's conference papers (right circle segment). The radii of the circle segments are proportional to the shares and are scaled to be comparable relative to each other and across map segments. The locations displayed correspond to the airport regions assigned to conference venues and affiliations. For example, San Francisco has values of about 3.3% (venue), 2.3% (paper affiliations), and 9.6% (firm paper affiliations).

The average conference event in our sample involves 81.67 conference papers authored by 209.40 unique scientists, with variation according to conference-series rankings: We count 68 $A^\star$ (top-ranked) conference series, 193 of rank A, 334 of rank B, and 387 of rank C. High-ranked conference series are fewer, but are longer-lived and larger, accounting for higher proportions of conference events and papers. On average, papers receive 3.47 scientific citations over a period of five years, varying from 1.53 (C-ranked) up to 9.85 ($A^\star$-ranked). On average, a paper is cited by 0.33 patents within five years. Patent citation rates are likewise higher for $A^\star$ papers (0.85). Around 9.5% of papers are ever cited in a patent by a firm in our sample. For $A^\star$ conferences, this proportion rises to 19.1%. Hence, as per Poege et al. (2019), we witness a positive association between scientific quality and technological impact.

4.2 Corporate participation

The data depict significant participation by firms in conferences, both in terms of scientific contribution and sponsorship. There is at least one corporate participant in 89.7% of the conferences in the dataset, and each conference has an average of 5.75 corporate participants. Corporate participation increases with the quality ranking of a conference; in $A^\star$ conferences, 18.4% of contributions are by firms, compared to a conference average of 11.2%. Papers with corporate affiliations receive more scientific citations than those without at all levels of conference ranking; for example, at $A^\star$-rated conferences, they receive, on average, 39.9% more citations than the average paper. This suggests that contributions by firms have scientific impact and receive attention. In terms of sponsorship, 26.7% of conferences have at least one corporate sponsor, with a conference average of 5.1 firms providing sponsorship (note that this estimate is conservative). Again, sponsorship is concentrated in conferences that are more highly ranked: 41.0% of $A^\star$ conferences are sponsored by at least one firm, dropping to 22.8% for rank C. The level of corporate participation is fairly constant throughout our observation period.

Table 1: Firm information

Firms	Conference participations	Papers per conference			Sponsoring with (%)	
		Mean	90^{th} pct.	99^{th} pct.	0 papers	≥1 paper
Top 10 (Mean)	1173.4	2.2	4.6	14.6	4.4	5.6
IBM	2603	2.9	6.0	19.0	3.7	9.0
Microsoft	2037	3.0	7.0	27.0	8.6	13.6
Intel	1147	2.0	4.0	10.0	4.5	9.3
Hewlett-Packard	1130	1.6	3.0	8.0	7.4	5.3
Alcatel-Lucent	954	2.3	5.0	18.0	2.2	2.3
Siemens	879	2.0	4.0	15.0	3.9	3.9
AT&T	843	2.1	5.0	12.0	1.1	3.1
NEC	785	1.6	3.0	7.0	3.2	2.3
NTT	709	1.9	4.0	10.0	0.7	1.3
Nokia	647	2.4	5.0	20.0	9.1	6.3
Top 50 (Mean)	460.7	1.6	3.2	8.9	5.8	2.7
All 4800 (Mean)	8.8	1.0	1.1	1.3	10.7	0.4

Notes: Firms are ranked by the number of conferences in which they participate. The last two columns show the share of conference participation that involves sponsorship, either exclusively (0 papers) or while also contributing with papers.

Corporate investments in conference participation are highly skewed, with a few firms accounting for a large proportion of such investments. On average, each firm in our data has participated in 8.8 conferences and has sponsored 0.9.[15] However, these averages conceal substantial heterogeneity. Table 1 provides information for different groups of firms, ranked by the number of conferences in which they participated. On average, the top 50 firms (roughly the top 1%) participated in 460.7 conferences, accounting for 55.6% of corporate participation with scientific contributions and 62.5% of sponsorships. Participation investments are also skewed at the level of the individual firm, reflecting a variety of participation modes. To demonstrate

[15] Interestingly, many conference papers associated with firms are co-authored with academic scientists (53.0%, on average), while collaborations with other firms are rarer (10.1%). Nevertheless, a significant share of corporate papers (36.9%) are authored exclusively by scientists affiliated with a single firm.

this, Table 1 shows the average number of papers per conference, together with the 90^{th} and 99^{th} percentiles, and the percentage of sponsored conferences, with and without scientific contributions. The top 1% of firms average fewer than two papers per conference, but contribute a remarkable number of papers to some conferences. As sponsors, this top 1% features in 8.5% of attended conferences and participates by simultaneously sponsoring and contributing scientifically in 2.7%.

Finally, descriptive statistics based on patent citations suggest that knowledge presented at conferences is important, especially for participating firms, and anticipate the findings in our econometric analysis. In our sample, firms filed, on average, 14.8 patents that cited one or more conference papers. In fact, 53.0% of all firms' patent citations of CS literature referenced conference papers. After excluding self-citations, 30.0% of these citations link to papers from conferences in which the patenting firm participated. This is a significant share, considering that the average firm participates in a relatively small proportion of conferences. The following analysis serves to better substantiate a causal relation between participation of firms and scientists in the same conferences and the use of associated knowledge.

5 Econometric strategy

5.1 Case-control estimation sample

To perform our econometric analysis, we build an estimation sample starting from all pairings between a firm and the papers it encounters at conferences in which it participates. Table 2 provides an example of the estimation sample based on the participation of Facebook in NeurIPS '09. We face the challenge that we only observe realized participation, resulting from the attendance choice made by firms and by the authors of other papers. To solve this, we adopt a case-control approach with the objective of expanding the sample to include a set of potential but unrealized participations from a choice set of similar conferences. This is analogous to studies in the geography of innovation that model inventors' "choice" of patent citations (Jaffe et al., 1993; Singh and Marx, 2013). In our context, we start by focusing on the conference choices made by the authors of papers to which a firm could have been exposed.

Given the list of papers that a firm encounters at a focal conference ("same-conference papers," represented by NeurIPS '09 in Table 2), we expand the sample to papers from other, similar conferences. We consider that any conference is within an unobserved choice set and authors from other conferences in the set could have presented their papers at the focal conference with positive probability. We then assume that conferences in the same year, subfield, size group, rank, and citations-based quality group belong to this set and randomly select up to two additional conferences within these strata ("matched conferences," represented by IJCAI/ICCV in Table 2).[16] We discard cases where the stratum only contains the focal conference, which reduces the number of focal conferences by 8.4%. Finally, we expand the sample by pairing

[16] This allows us to limit the size of the dataset and avoids inflating the weight of large fields, although results are robust to different selection criteria. We exclude conferences located very close to the focal conference (1.2%) because for these, exposure to direct flights would be identical between case and control. The average observable difference between paired conferences cannot be distinguished from zero.

each firm at the focal conference with the papers from these matched conferences "as if" they had been presented at the focal conference ("matched-conference papers").[17]

Table 2: Exemplary data setup

Firm	Focal conference	Matched conference	Conference paper	Paper relation to focal conference	Follow-on citations
Facebook	NeurIPS '09 *Vancouver*	NeurIPS '09	Miller et al. *San Francisco*	Same conf.: 1 Direct flight: 1	Patent: 0 Science: 1
Facebook	NeurIPS '09 *Vancouver*	IJCAI '09	Hamadi et al. *London*	Same conf.: 0 Direct flight: 1	Patent: 0 Science: 0
Facebook	NeurIPS '09 *Vancouver*	ICCV '09	Dong et al. *Xiamen*	Same conf.: 0 Direct flight: 0	Patent: 0 Science: 0

Notes: Shows example entries from the estimation dataset. Facebook participated in NeurIPS '09, and two matched conferences – IJCAI '09 and ICCV '09 – were selected. As an example, we show one paper from each conference, each paired with all firms (here, only Facebook is shown). Same conf. (in the regressions "Same conference") is then an indicator variable depicting whether the paper was presented at NeurIPS. Direct flight is an indicator of whether a direct flight connection from the location of the paper's authors to the venue of NeurIPS, Vancouver, existed in 2009. Citations are indicator variables for whether the firm cited paper p within five years in patents or science. See Appendix A.1 for detailed references.

This procedure generates variation between the same-conference papers to which a firm was exposed and matched-conference papers from conferences in which the firm did not participate. Firms can participate in multiple conferences in a choice set, but this is relatively rare, and we choose to drop these cases, which represents 8% of the overall sample.[18] We also drop both the matched and the same-conference firm-paper pairs if the firm participates in all matched conferences, which corresponds to 4% of the sample. The results are robust to these choices.

The resulting estimation sample has 7,611,107 observations, at the firm-conference-paper level. Owing to the observations dropped, this sample differs from that described in Section 3 and comprises 3,826 firms that participate in 5,163 conferences in 932 conference series with 345,446 unique papers. Within this sample, on average, each firm participates in 7.1 conferences and each conference event is associated with 106.7 papers. The average firm is paired, in total, with 870.4 same-conference papers. To these we add papers from, on average, 11.6 matched conferences in which the firm does not participate, for a total of 1,118.9 matched-conference papers per firm. Due to singletons in high-dimensional FEs, the number of observations can deviate slightly in the regression analyses. We discuss the construction of the choice set in greater detail in Appendix B. Table 3 shows summary statistics for the estimation sample.

[17] To exclude self-citations, we disregard same-conference or matched-conference papers authored by the focal firm. We maintain author-level self-citations, which may be the result of hiring, but results are robust to their exclusion.

[18] Within each choice set, the average firm participates in 6.5% of conferences. For our analysis, a high rate of multiple attendances by firms would invalidate our setup because there would be no variation in the variable of interest. At the same time, too few multiple attendances would suggest that the conferences in the choice set are too dissimilar. The composition of our sample suggests a reasonable balance between the two extremes.

Table 3: Summary statistics of the estimation sample

Variable	Mean	Std Dev.	Min	P50	P75	Max
Main variables (% points)						
Same conference	43.8	49.6	0	0	100	100
Direct flight	30.6	46.1	0	0	100	100
Dependent variables (% points)						
Patent citation	0.2	4.1	0	0	0	100
Science citation	0.5	6.8	0	0	0	100
Patent citation to past	5.3	22.3	0	0	0	100
Science citation to past	8.8	28.3	0	0	0	100
Collaboration	2.7	16.1	0	0	0	100
Hiring	0.6	7.7	0	0	0	100
Conference-level variables						
Conference size	387.5	343.4	10	281	616	1246
Year	2006.5	3.0	1996	2007	2009	2010
A* conference	0.1	0.3	0	0	0	1
A conference	0.3	0.5	0	0	1	1
B conference	0.4	0.5	0	0	1	1
C conference	0.2	0.4	0	0	0	1
Firm-related variables						
Geo. distance to paper	5424.9	4339.3	0	5519	8826	19 964
Research similarity	0.1	0.1	0	0	0	1
Science cit to past$_{pre-conf}$	0.0	0.2	0	0	0	1
Patent cit to past$_{pre-conf}$	0.0	0.1	0	0	0	1
Firm scientists n.	192.5	417.7	0	26	191	2784
Firm papers at conference	1.4	1.7	0	1	1	38
Firm contributes paper(s)	0.9	0.2	0	1	1	1
Firm is sponsor	0.1	0.3	0	0	0	1
Firm is sponsor and contributes	0.0	0.1	0	0	0	1
Firm in science hub	0.6	0.5	0	1	1	1
Paper-related variables						
Geo. distance from paper to conf.	6581.6	4486.5	0	7195	9562	19 957
Paper team female	0.2	0.3	0	0	0	1
Paper team productivity	2.4	5.6	0	0	2	94
Paper by other firm	0.1	0.3	0	0	0	1

Notes: Summary statistics for the estimation sample, N = 7,611,107. Variables about follow-on activity are in percentage points, indicating the share of observations in which the activity occurred. For example, a "Patent citation", where at least one patent of the focal firm cited the paper within five years of the conference year, occurred in 0.2% of observations. Other variables are as is; see Section D.4 for details. Geographic distance is measured in kilometers. Variables that are used in log transformations in the analyses are displayed here without transformation.

5.2 Regression analysis and instrumental variable approach

Regression framework. Having set up the estimation sample of actual and potential participation of papers in a given conference, we follow examples in the literature on the probability of patent citations (Singh and Marx, 2013) and study the probability that a participating firm cites a paper in a regression framework, as follows:

$$\text{Patent cit}_{fcp} = \beta_1 \text{Same conference}_{fcp} + \beta_2 X_{fcp} + \epsilon_{fcp} \tag{5.1}$$

Here, the dependent variable *Patent cit* is assigned a value of 1 if firm f cites a conference paper p in patents in the five years following the conference, and 0 otherwise. The variable of interest *Same conference* is assigned a value of 1 if the paper p is a "same-conference paper," that is, it was presented at the same conference c in which firm f participated. Conversely, *Same conference* is assigned a value of 0 if p is a "matched-conference paper" presented at a conference in which the firm did not participate, i.e., firm f was not exposed to it at conference c. The matrix X_{fcp} comprises a set of control variables and FEs. For the estimation, we use a Linear Probability Model (LPM).[19]

Instrumental variable. We propose an IV approach to address the endogeneity in the knowledge firms are exposed to at the conferences where they participate. Conditional on participating in a conference within a field, scientists and firms choose which conference to attend. This can result in a positive bias if these choices reflect pre-existing and unobserved connections and preferences or a negative bias if they reflect the desire to explore new knowledge, which would be otherwise inaccessible. More generally, unobserved characteristics of the papers or their authors may explain a firm's citations rather than its exposure to them via the conference. Our matching procedure, covariates, and FEs may only partially address this issue.

To alleviate these concerns, we introduce exogenous variation in the probability of participation of papers' authors while taking firm participation as given. Specifically, we rely on an IV affecting the probability that paper p's authors choose to participate in conference c, where firm f participates.[20] As an instrument, we use the availability of direct flights in the conference year from the locations of the authors of paper p to the venue of the conference c. This likely affects the probability that paper p's authors choose to participate in conference c because direct flights tend to reduce costs and travel time (Giroud, 2013; Catalini et al., 2020).

We argue that conditional on appropriate controls, direct flights are a valid IV to address concerns about the endogeneity of firms' exposure to papers. First, the locations of conferences and participating firms most often differ. Thus, direct flights to conference c that are relevant for paper p's authors are independent of the connectivity between firm f and paper p's authors, which increases confidence in the exclusion restriction. Second, direct flights depend on the airlines' and conference organizers' decisions, which are likely independent of the citation

[19] This is to ease the interpretation of the coefficients and because non-linear probability and count models are barely applicable owing to our sample size and the use of FEs.

[20] In principle, the authors of a paper can be authors on additional papers and present them to other conferences in a choice set. This may render the instrument weak. However, because it only occurs for 3.3 % of the sample, it is not a concern in practice. The results are robust to discarding these observations (Table A-6).

probability between f and p. Airline routes are determined largely by broader market demand and regulations (Campante and Yanagizawa-Drott, 2017); conference locations depend on the general attractiveness of venues and are often scheduled years in advance.[21]

Taking the firms' participation as given has two consequences. First, some firms never participate in conferences, and our results do not generalize to them. Second, although we address concerns about unobserved characteristics of the papers or their authors, unobserved firm characteristics related to conference choice may also play a role. For instance, we may observe a causal effect of the presentation of a paper at the conference because the firm pays attention to that conference independently from its participation. However, we limit this concern by comparing conferences of the same quality and topic and using relevant covariates and FEs, discussed below. Further, in section 6.2, we study heterogeneity in the firm's observable levels of engagement (participation) with the conference. We find no effects when firms make minimal participation efforts, which suggests that attention alone would also be insufficient. In summary, we study corporate learning at conferences as a response to exposure for those firms that have decided to participate.

First-stage regression. With the IV, we augment our regression into a two-stage regression model. The second stage is represented by Equation 5.1 and the first stage is the following LPM:

$$\text{Same conference}_{fcp} = \gamma_1 \text{Direct flight}_{cp} + \gamma_2 X_{fcp} + u_{fcp} \tag{5.2}$$

This equation models the probability of *Same conference* as a function of the existence of a *Direct flight* between the location of paper p's authors and conference c's venue. The matrix X_{fcp} comprises the same control variables and FEs as in Equation 5.1, which we detail below.

Fixed effects. We use FEs to absorb variance that could threaten the validity of the instrument, following Giroud (2013). For instance, direct flights may be more frequent from research-intensive regions and toward high-quality conferences in which more innovative firms participate. Unobserved time trends in the increase in connectivity and citations may cause a spurious correlation. Accordingly, we include FEs at the levels of the firm, the region of origin of paper p's authors, the conference series to which conference c belongs, and the year of the conference. Because the research quality of regions varies by field, we interact the respective FEs with the CS subfields of the conference c.

We use pair-level FEs to address additional concerns. Most importantly, direct flights from the regions of paper p's authors to the conference may still correlate with its direct connectivity with the firm. This can happen, for instance, if both authors with direct flights and firms are closer to the venue, and, hence, closer to each other. To rule out this type of concern, we include pair FEs between firms and the region of paper p's authors, which controls more generally for any type of proximity (e.g., geographic, cultural, linguistic). Finally, we control for time-specific shocks. For instance, trends in productivity could be location- or firm-specific, and might correlate with direct flights via improvements in infrastructure. This argument applies analogously both at the level of the firm and that of the region of paper p's authors; therefore,

[21] Overall, 79.6% of conferences take place at a location different from the previous year.

we add pair FEs between years and both the firm and the authors' region. The results hold for alternative FE specifications that we discuss in Section 7.

Control variables. Although we work under the assumption that, conditional on FEs, the instrument is valid, we add pair-level covariates to additionally control for the specific relevance of the knowledge embedded in paper p for the firm. We do so by first capturing any revealed attention by the firm to paper p's authors prior to the conference; specifically, *Science cit to* $past_{pre-conf}$ captures citations from any of the firm's publications in the five years preceding the year of the conference to publications of the authors concerned in the same five years. *Patent cit to* $past_{pre-conf}$ captures similar citations but from any previous *patent* of the firm. This variable is analogous to a lagged dependent variable. Second, we directly capture the similarity between the firm's research firm and paper p using the variable *Research similarity*, which measures the text-similarity of paper p to the firm's conference papers in the same CS subfield in the year before the conference. Other unobserved confounding factors, correlated with both the IV and the dependent variable, may exist at this level of analysis; this remains an empirical limitation. However, FEs capture much of the variation at which potential confounding factors may operate, and the overall stability of the estimates to the inclusion of these controls is reassuring. In robustness analyses, we also consider additional controls (see Section 7).

Clustering of standard errors. Finally, we cluster standard errors at the level of the paper authors' regions. This follows the recommendation by Abadie et al. (2023) to cluster standard errors at the level at which treatment (direct flight) is assigned or at a higher level. Clustering at the level of authors' regions takes this aspect into account and allows for correlated residuals within regions. These clusters also nest author-level repeated observations, specifically in cases where scientists present multiple papers at a conference. Results are robust to route and other clustering levels.

6 Results

6.1 The effect on patent citations

We start by presenting results about the effect of the presentation of a paper in the same conference where a firm participates on the likelihood that the firm cites the paper in patents. In Columns 1 to 4 of Table 4, we first show the OLS coefficient estimates. The first column exclusively includes level FEs for conference series, authors' locations and subfield pairs, firms, and years. Column 2 incorporates control variables for the expected relevance of the paper to the firm, and Column 3 incorporates firms' and paper authors' locations pair FEs. Finally, Column 4 adds pair FEs for year and firms and for authors' locations, respectively.

We find positive and significant estimates. The coefficients on covariates in Columns 2 to 4 are also, as expected, positive and significant. Participation in the same conference is associated with about a 0.12 percentage-point (pp) increase in the patent citation probability, with minimal variation across specifications. Patent citations are naturally sparse, leading to small absolute estimates. Relative to the dependent variable sample mean (0.17 pp), the coefficient size amounts to a 68% increase. Comparison with the covariate coefficients also serves

Table 4: OLS and IV regression results on the effect of *Same conference* on patent citations

	OLS: Patent citation				First stage: Same conference				Second stage: Patent citation			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Same conference	0.142***	0.103***	0.118***	0.118***					0.246***	0.210***	0.402***	0.304***
	(0.012)	(0.008)	(0.010)	(0.010)					(0.065)	(0.064)	(0.114)	(0.118)
Direct flight					8.734***	8.567***	6.718***	6.901***				
					(0.910)	(0.883)	(0.742)	(0.754)				
Science cit to past$_{pre-conf}$		0.687***	0.572***	0.617***		14.581***	12.993***	12.497***		0.672***	0.535***	0.593***
		(0.052)	(0.043)	(0.047)		(0.255)	(0.255)	(0.254)		(0.052)	(0.042)	(0.046)
Patent cit to past$_{pre-conf}$		0.957***	0.826***	0.843***		7.464***	7.012***	6.694***		0.949***	0.806***	0.830***
		(0.122)	(0.112)	(0.112)		(0.393)	(0.391)	(0.396)		(0.122)	(0.111)	(0.112)
Research similarity		1.001***	1.058***	1.187***		109.798***	101.368***	122.595***		0.883***	0.770***	0.959***
		(0.095)	(0.102)	(0.112)		(1.460)	(1.561)	(1.703)		(0.114)	(0.135)	(0.168)
Firm FE	Yes	Yes			Yes	Yes			Yes	Yes		
Year FE	Yes	Yes	Yes		Yes	Yes	Yes		Yes	Yes	Yes	
Conf Ser FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Origin × Field FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Origin × Firm FE			Yes	Yes			Yes	Yes			Yes	Yes
Year × Origin FE				Yes				Yes				Yes
Year × Firm FE				Yes				Yes				Yes
Observations	7611039	7611039	7344378	7344254	7611039	7611039	7344378	7344254	7611039	7611039	7344378	7344254
Number clusters	1283	1283	1097	1088	1283	1283	1097	1088	1283	1283	1097	1088
DV mean (% points)	0.17	0.17	0.17	0.17	43.756	43.756	43.756	43.756	0.17	0.17	0.17	0.17
F (excl. Inst.)									92.1	94.1	81.9	83.8

Notes: OLS and IV estimates. $^{*}p < .1$, $^{**}p < .05$, $^{***}p < .01$. Standard errors in parentheses, clustered at the level of the paper authors' origin region. In Columns 5-8, the dependent variable is *Same conference*. *Direct flight* is an indicator variable of whether a direct flight connected the region of paper p with the conference c. All variables are in percentage points. Data is at the firm-conference-paper level. *Same conference* is assigned 1 for papers presented at the conference in which the firm participates and 0 for papers presented at a matched conference. We instrument *Same conference* by the direct flight availability between the paper authors' location and the conference venue. Firm-paper controls include indicators for patent or science citations from the firm to the authors in the years before the conference and how similar the firm's publications in the previous year were to the focal paper. F (excl. Inst.) is the first-stage Kleibergen-Paap F-statistic.

as a benchmark. For instance, the estimate on *Same conference* is about 14% of the increase associated with the authors' work being cited in the firm's patents prior to the conference.

Results from the IV analysis confirm the positive effect of exposure to a paper at a conference on the probability of patent citations. We present these in Columns 5 to 12 of Table 4, following the same specifications presented for the OLS analysis. Columns 5 to 8 show results for the first-stage regression. The instrument *Direct flight* has a strongly significant and positive effect on the probability that a paper is actually presented at the conference in which the firm participates. The F-test on the excluded instrument is between 80 and 95, well above common weak-instrument thresholds. The magnitude of the effect is economically meaningful, demonstrating an increase, in the most complete specifications, of about 6.9 pp, relative to a sample mean of 43.8%.[22] Columns 9 to 12 show results for the second-stage estimates for exposure on patent citations; the magnitude of the coefficient varies between 0.2 and 0.4 pp, and is 0.3 pp in the most complete specification, in Column 12.

Notably, the IV estimates are larger than the OLS, which could have multiple explanations. First, the IV identifies a LATE on the population most sensitive to the instrument: Scientists sensitive to travel costs may be harder to meet in general, hence meeting at a conference may yield larger effects at the margin.[23] Second, as mentioned in Section 5, the OLS estimates may be negatively biased; for instance, in comparing similar conferences, scientists and firms may choose to participate in the one that provides access to otherwise inaccessible interaction opportunities. Finally, the fact that for a given paper we observe participation at the extensive margin (of at least one author) and not at the intensive margin (the number of participating authors per paper) could be regarded as a measurement error in the endogenous variable.[24]

6.2 The role of different modes of corporate participation

We analyze whether different modes of participation on the part of the firm moderate the effect of exposure to knowledge at a conference. To this end, we augment our model with interaction variables reflecting the main corporate participation modes. Table 5 presents results based on the complete IV specification. For comparison, Column 1 reports the results of the main analysis. In Column 2, we focus on the level of scientific contribution, distinguishing between firms with one (weak contribution), two to four (medium), and five or more papers (strong). Stronger scientific contributions are associated with ever-increasing effects. Next, in

[22]The sample mean results from the sample construction (Section 5.1): for a set of papers in a conference in which a firm participates, we select up to two sets of papers from two conferences from the choice set of similar conferences. By construction, the sample mean would be 33% if all choice sets contained two or more conferences, and if all conferences were of equal size. However, some choice sets contain only one conference, which leads to a higher sample mean. Variance in conference size plays a role, too.

[23]To explore the heterogeneity of the IV estimate across characteristics of the paper's team and the conference, we present an extension of the first-stage results in Appendix A.3 using variables introduced in Section 6.4. Although the instrument remains significant across groups, we find that direct flights matter less for teams with authors of higher productivity and male authors, which is compatible, for instance, with the idea that they may be less resource-constrained. The estimates are also larger for medium- and low-ranked conference series, about which alternative authors might potentially be more indifferent.

[24]In other words, the allocation of papers to conferences is precise and implies the participation of at least one author. However, authors may participate in smaller or greater number, increasing chances for interactions and subsequent citations. The use of direct flights as IV would overcome this form of measurement error.

Column 3, we add sponsorship information, separating firms that only sponsor and those that both sponsor and contribute scientifically; here, variables on scientific contributions are specific to firms that are not sponsors. We find that firms that both contribute scientifically and as sponsors exhibit larger effects. Interestingly, the same holds for firms that make several scientific contributions (two or more) but do not provide sponsorship. However, firms that act only as sponsors experience no significant effect. This supports the idea that scientific contributions are necessary, whereas sponsorship is complementary.[25]

Table 5: Heterogeneity by modes of corporate participation

	Patent citation			
	(1)	(2)	(3)	(4)
Same conference	0.304***			
	(0.118)			
× Weak contribution (<2 p.)		0.080	0.012	0.139
		(0.111)	(0.106)	(0.182)
× Medium contribution (2-4 p.)		0.975***	0.874***	1.220***
		(0.267)	(0.248)	(0.380)
× Strong contribution (5+ p.)		1.928***	1.971***	2.415***
		(0.737)	(0.747)	(0.854)
× Sponsor only			0.363	0.920
			(0.346)	(0.632)
× Any contribution ($\geq$ 1 p.) + Sponsor			2.979***	3.225***
			(0.874)	(0.927)
Firm sample	All	All	All	Top 50
Paper-level controls	Yes	Yes	Yes	Yes
Standard FE	Yes	Yes	Yes	Yes
Observations	7344254	7344254	7344254	2963241
Number clusters	1088	1088	1088	1083
DV mean (% points)	0.17	0.17	0.17	0.389
F (excl. Inst.)	83.8	27.2	18.0	18.4

Notes: IV estimates. $^{*}p < .1$, $^{**}p < .05$, $^{***}p < .01$. Standard errors in parentheses, clustered at the level of the paper authors' origin region. Column 4 restricts the sample to the yearly top 50 firms by scientist count. All variables are in percentage points. Data is at the firm-conference-paper level. *Same conference* is assigned 1 for papers presented at the conference in which the firm participates and 0 for papers presented at a matched conference. We instrument *Same conference* by the direct flight availability between the paper authors' location and the conference venue. Firm-paper controls include indicators for patent or science citations from the firm to the authors in the years before the conference and how similar the firm's publications in the previous year were to the focal paper. F (excl. Inst.) is the first-stage Kleibergen-Paap F-statistic. Standard FEs include conference series, researcher origin×field, origin×firm, year×origin and year×firm FEs.

Importantly, even for the top firms, the effect of exposure at conferences where they participate with a weak contribution or only as a sponsor is not significant, implying that heterogeneity among firms does not fully explain the results. In Column 4, we show that the results hold for the top 50 firms by scientist count in a given year (firm-year FEs remain in all specifica-

[25] These group categories are also associated with a larger mean of the dependent variable. However, the effect sizes grow more than proportionally relative to them. So, the average citation probability for firms with only weak scientific contribution is 0.1%, compared with 0.9% in the category of firms that both sponsor and contribute scientifically. The difference between the coefficient estimates is substantially larger.

tions). The analysis primarily exploits variation in a firm's participation investments across conferences. Hence, participation investments seem to have a specific relevance in the context in which they are deployed (Fosfuri et al., 2015). Because the mode of participation of the firm is not exogenous in our empirical setup, this observation also helps rule out the main alternative explanations. In particular, even firms with a superior absorptive capacity and legitimacy seem incapable of leveraging minimal participation efforts.[26] However, we cannot entirely rule out that absorptive capacity is specific to the conferences where a firm chooses to invest more.

We can partially address other alternative explanations based on qualitative insights. First, one specific concern is that sponsorship may, from the outset, have motivations other than learning. For instance, some firms may sponsor conference events for commercial purposes because science is their target market. However, commercial objectives were rarely mentioned in our interviews. Corporate participation reflected an interest in the conference's research field, and sponsorship shared a learning objective with scientific activities. Most significantly, HR employees explicitly referred to synergies between scientific contributions and sponsorship, citing active participation and the presence of scientists as favoring interactions and increasing the engagement of other participants with the firm's promotional spaces. Many reported that sponsorship without active participation could result from inexperience or lack of available scientists or coordination, and that it was less effective. However, firms' motivations remain unobserved, and we cannot control for them empirically.

Finally, the role of specific micro-mechanisms remains tentative, but the evidence at least suggests a role for reputation. If only other mechanisms were involved, such as proximity and salience, minimal participation efforts would likely also yield significant results, and larger investments would face decreasing marginal returns: Interactions would gradually become redundant. However, minimal efforts are associated with small and insignificant coefficients, and coefficients are highest for top participants, which is compatible with a role for reputation. This aspect also emerged qualitatively: Firms reported using internal peer review to guarantee that work presented was of above-average quality, precisely to safeguard reputation. They also declared motives of contributing to the community and showing reciprocity for what they learned from it. Accordingly, firms often link participation in a conference with corporate statements that emphasize the symbolism of sponsorship, the number and quality of their contributions, and their openness to interaction and collaboration (we report a list in Appendix C).[27]

6.3 Conference participation and other learning channels

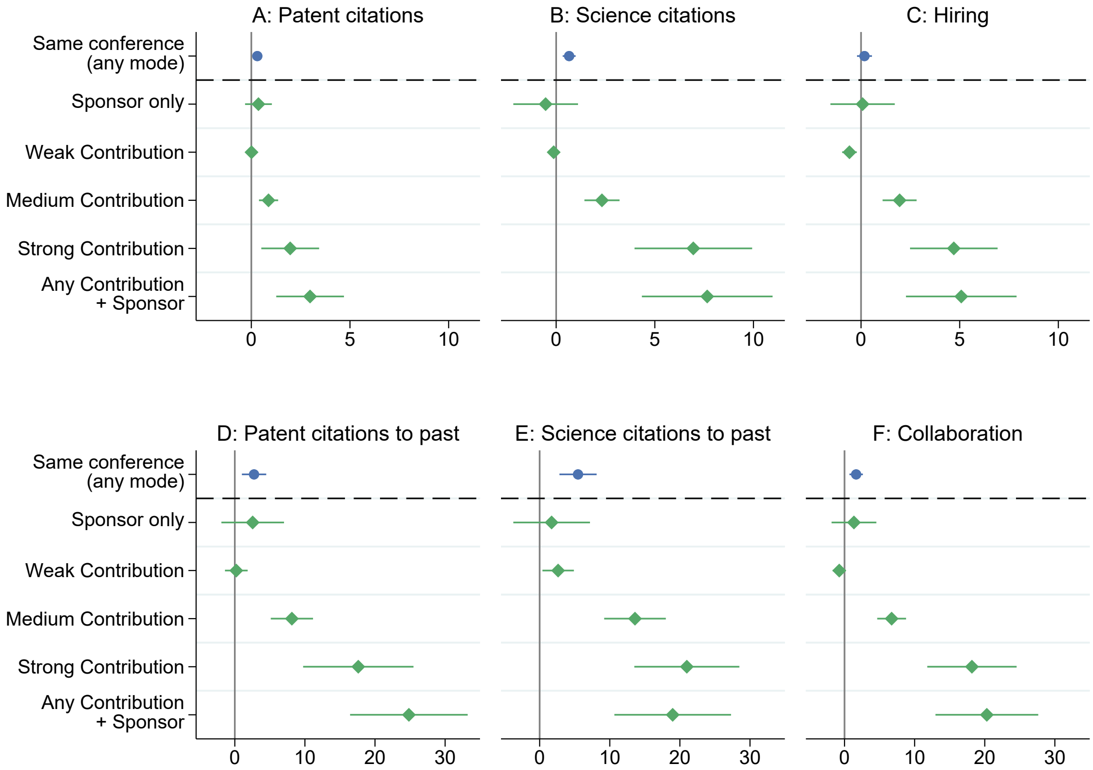
Next, we study the effect of exposure to knowledge at a conference on additional outcomes as evidence for the role of personal connections as learning channels. In Figure 2, we show

[26] In this respect, one notable insight from our interviews at ECCV is the response of a scientist affiliated with one of the leading companies in the computer industry, who – while attending a small sponsorship booth together with another colleague but without any corporate scientific contributions to the conference – expressed dissatisfaction with the limited commitment of their company to the conference, which the scientist deemed important.

[27] One of the most exemplary statements runs: "Facebook is thrilled to be a Champion Sponsor of CVPR 2021. This year, we will present over 52 publications [...] and participate in 28 workshops and tutorials. We seek to advance the state-of-the-art in [...] open collaboration with CVPR's dynamic scientific community."

average estimates for each outcome at the top of the six panels, and below the estimates by mode of corporate participation. The estimates are based on our preferred IV specification; for comparison, Panel A shows the estimates for patent citations (cf. Table 5).

Figure 2: Learning channels and heterogeneity by mode of corporate participation



Notes: Each panel shows the result from two IV regressions. At the top, (circles) we report the average IV coefficient for the respective dependent variable. The full specification is reported in Online Appendix Table A.1 and the OLS/IV comparisons in Online Appendix Table A.2. Below this, (diamonds) we follow Table 5, column 3. For each interaction of Same conference and firm participation mode, we use as instrument the interaction between Direct flight and the corresponding participation mode. Full estimation results are available upon request.

First, Panel B shows that participation in the same conference also affects citations from follow-on science, with increasing magnitude for increasing levels of firm participation, as witnessed for patent citations. The results suggest that learning from conferences, as measured by patent citations, does not replace corporate engagement with the science base. In a robustness analysis, we use a text-similarity measure as an alternative to science citations (see Appendix Table A-6). The results are equivalent, suggesting that the material content of the research subsequently carried out by the firm also changes. This alleviates the concern that changes in patent or science citations may only reflect the salience of knowledge. Overall, the results suggest that exposure to knowledge at scientific conferences leads to actual learning and enables follow-on corporate scientific activity.

Second, we consider citations to previous research of the papers' authors. We do so in the same way for patent citations (Panel D) and science citations (Panel E), labeled as *[Patent/Science] citations to past*. The large effects we encounter for this type of indicator suggest that interactions with conference participants may support learning in relation to research not presented at the conference that, in principle, was already in the public domain. Conversely, a larger effect on citations of the focal paper might have suggested that participation serves mainly as a means to screen information and gain a time advantage in its exploitation.

Finally, we study the effect on indicators more closely associated with personal connections. Figure 2(c), presents results on the *Hiring* of scientists, which we measure via evidence from affiliation information indicating that at least one of a paper's authors moved to the focal firm in the five years following the conference. Figure 2(f), presents results on *Collaboration*, captured by future scientific output in which the firm and at least one of a paper's authors appear together. We find large effects for both, especially – again – for firms that both contribute scientifically and sponsor. The effect on collaborations is more prominent than that on hiring, which is not significant in the full sample. These results suggest that intense participation permits access to knowledge channels that are rival in nature (Zucker et al., 2002, 2008; Breschi and Lissoni, 2001). In fact, in science production, citations may reflect use of knowledge in the form of a public good, freely available and actionable by multiple actors. By contrast, collaboration and hiring imply active engagement on the part of scientists who are resource- and time-constrained.

6.4 Heterogeneity at the firm and paper levels

We study how characteristics of firms and papers moderate the effect, in particular to assess whether learning is concentrated among specific participants. On the one hand, conferences may provide a level playing field, especially for participants lacking alternative learning channels. On the other hand, selective social interactions and concentrated learning may prevail. In this heterogeneity analysis, this distinction does not depend on a causal interpretation of the moderation coefficients but helps to explain under what circumstances participation in the same conference may be more important. The estimates follow the preferred IV specification.

In Table 6, we first explore the role of a firm's position relative to a paper. In Column 2, we consider the interaction between participation and the similarity of the firm's research to the paper. We find that the effects are stronger for firms with greater research similarity. In Column 3, we test the interaction with geographic distance according to the first author's location. While conferences are often framed as a means of overcoming geographic distance, some degree of proximity may be necessary to follow up effectively on conference encounters. We find a negative interaction coefficient, although it is only weakly significant or not significant when considering other interactions jointly. Extending this analysis to other outcomes (see Table A-4), we find that the interactions are negative and significant for those most closely related to personal connections, such as collaborations. This could suggest that the importance of following up with more interactions predominates. However, the F-test on excluded instruments reveals a potential weak-instrument problem, which imposes caution on the interpretation.

In Columns 4 and 5 of Table 6, we focus on corporate characteristics. First, we consider whether a firm has a significant R&D presence in a science hub. We proxy this with an indicator variable for whether scientists affiliated with the firm are present in a top-10 region according to field-specific yearly CS output. We find that the effect is stronger for such firms, although the significance of the coefficient fades when we consider different interactions simultaneously (Column 6). Second, we study the interaction with the size of a firm's investments in research, which we proxy using the number of active scientists it employs according to affiliations listed in publications. We find that the effect increases strongly for firms with high research investments.

Table 6: Heterogeneity by corporate characteristics

	Patent citation					
	(1)	(2)	(3)	(4)	(5)	(6)
Same conference	0.304^{***}	-0.657^{***}	1.401^{**}	-0.097	-1.167^{***}	-1.221
	(0.118)	(0.156)	(0.630)	(0.081)	(0.259)	(0.841)
× Research similarity		11.190^{***}				10.130^{***}
		(2.382)				(2.456)
× Geo. distance to paper			-0.139^{*}			-0.043
			(0.077)			(0.079)
× Firm in science hub				0.615^{***}		0.304
				(0.206)		(0.185)
× Firm scientists n.					0.329^{***}	0.178^{***}
					(0.075)	(0.061)
Paper-level controls	Yes	Yes	Yes	Yes	Yes	Yes
Standard FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7344254	7344254	7325947	7344254	7344254	7325947
Number clusters	1088	1088	1088	1088	1088	1088
DV mean (% points)	0.17	0.17	0.17	0.17	0.17	0.17
F (excl. Inst.)	83.8	39.8	5.9	43.8	26.2	2.7

Notes: IV estimates. $^{*}p < .1$, $^{**}p < .05$, $^{***}p < .01$. Standard errors in parentheses, clustered at the level of the paper authors' origin region. *Research similarity* is the average similarity of firm publications in the previous year to the focal paper. *Geo. distance to paper* is the log distance between the closest firm R&D location in the previous year to the region of the paper's first author. *Firm in science hub* indicates whether the firm has an R&D location in a hub of the conference field. *Firm scientists n.* is the size of the scientific workforce of a firm, based on CS publication information, lagged by one year. All regressions also include the interaction variables separately, although the coefficients are not reported. All variables are in percentage points. Data is at the firm-conference-paper level. *Same conference* is assigned 1 for papers presented at the conference in which the firm participates and 0 for papers presented at a matched conference. We instrument *Same conference* by the direct flight availability between the paper authors' location and the conference venue. Firm-paper controls include indicators for patent or science citations from the firm to the authors in the years before the conference and how similar the firm's publications in the previous year were to the focal paper. F (excl. Inst.) is the first-stage Kleibergen-Paap F-statistic. Standard FEs include conference series, researcher origin×field, origin×firm, year×origin and year×firm FEs. Regression results for alternative dependent variables are available in Table A-4.

In Table 7, we show that the effect of exposure to knowledge at a conference is related to the characteristics of a paper's authors. In Column 2, we consider the team's productivity, using as a proxy the (log) highest number of $A^{\star}$ publications in the last five years among the authors (plus one), which is also associated with scientific prestige or visibility. The effect is stronger for teams with higher productivity. In Column 2, we consider gender, looking at the proportion of female authors.[28] Some literature suggests that women are at a disadvantage because the social spaces at conferences favor masculine appearance in interpersonal interactions (Hansen and Pedersen, 2018), and this may be especially true in CS where the proportion of women is low (16% in

[28] Gender information is not available in publications data and is impossible to collect on a large scale. We infer gender from the first name of authors using services from Gender-API.com. We have searched for the predicted gender for the roughly 2 million unique author names in the dblp CS bibliography, finding information for 87% of the sample.

our sample). Consistent with this argument, we find that the effect is attenuated for female teams, also controlling for the role of productivity. Finally, in Column 4, we differentiate *Paper by other firm* as publications associated with other firms. This dimension could be informative of a possible role for conferences in knowledge spillovers between rivals. Competition might imply that learning between firms is less likely (Arora et al., 2021) or, conversely, research by rival firms may receive special attention (Bikard, 2018). In this respect, our results are mixed. The interaction coefficient is positive but only weakly significant, and the instrument is weaker. However, interestingly, in extending this analysis to collaboration outcomes (see Table A-5), we find that this interaction coefficient turns significantly negative.

Table 7: Heterogeneity by characteristics of the conference paper's authors

	Patent citation				
	(1)	(2)	(3)	(4)	(5)
Same conference	0.304***	−0.016	0.062	0.248**	0.013
	(0.118)	(0.088)	(0.087)	(0.105)	(0.087)
× Paper team productivity		0.474***	0.480***		0.478***
		(0.146)	(0.145)		(0.146)
× Paper team female			−0.475***		−0.460***
			(0.168)		(0.173)
× Paper by other firm				0.691*	0.572
				(0.409)	(0.353)
Paper-level controls	Yes	Yes	Yes	Yes	Yes
Standard FE	Yes	Yes	Yes	Yes	Yes
Observations	7344254	7344254	7249463	7344254	7249463
Number clusters	1088	1088	1085	1088	1085
DV mean (% points)	0.17	0.17	0.17	0.17	0.17
F (excl. Inst.)	83.8	44.6	29.4	9.9	9.4

Notes: IV estimates. $^{*}p < .1$, $^{**}p < .05$, $^{***}p < .01$. Standard errors in parentheses, clustered at the level of the paper authors' origin region. *Paper team productivity* is the maximum log(1+) number of $A^{\star}$ papers published by an author of the focal paper in the last five years. *Paper team female* is the average proportion of female authors of the focal paper. *Paper by other firm* is an indicator of whether the paper is affiliated with a firm. All regressions also include the interaction variables separately, although the coefficients are not reported. All variables are in percentage points. Data is at the firm-conference-paper level. *Same conference* is assigned 1 for papers presented at the conference in which the firm participates and 0 for papers presented at a matched conference. We instrument *Same conference* by the direct flight availability between the paper authors' location and the conference venue. Firm-paper controls include indicators for patent or science citations from the firm to the authors in the years before the conference and how similar the firm's publications in the previous year were to the focal paper. F (excl. Inst.) is the first-stage Kleibergen-Paap F-statistic. Standard FEs include conference series, researcher origin×field, origin×firm, year×origin and year×firm FEs. Regression results for alternative dependent variables are available in Table A-5.

Taken together, these results suggest more relevance for social and reputation mechanisms and support the interpretation that learning interactions at conferences, although facilitated for all participants, are concentrated in accord with social and reputational mechanisms. Arguably, if the reduction of search costs due to proximity were the sole explanation and conferences were most relevant for organizations lacking alternative learning channels, conferences would matter more in the exploration of distant knowledge and matter less for firms better positioned in the scientific landscape. However, the results suggest that such a more favorable position may be complementary to proximity at a conference. Similarly, because more prestigious (or productive) teams are more visible, we might expect that conferences matter less when it comes to identification and use of their knowledge. On the other hand, a conference may represent a unique opportunity to interact and establish a connection with such teams, which would also explain the observed corporate participation efforts as necessary to attract their attention. Overall, the evidence suggests that participation is more important in relation to more valuable knowledge but may be less effective for participants who are disadvantaged in the social context.

7 Robustness

We conduct a series of robustness analyses. First, we test different FEs and clustering of standard errors. We consider conference location × year FEs, which controls, for instance, for unobserved amenities at the conference venue. It is possible to include FEs for the full interaction of firm, authors' region, and year. This specification leverages the fact that direct flights to conference venues do not coincide with direct flights to firm regions and excludes the possibility that contemporaneous shocks between the firm and scientists' region pairs affect the results. We also verify that the results hold with reduced FEs, for instance, only for firms and years. In this case, estimates are inflated, which encourages us to believe that the rationale for the inclusion of these FEs is well-founded. We cluster standard errors at different levels, including the paper level, firm level, authors' region × conference location level (corresponding to the level of observation of the instrument), and their combination in two-way clustering. We report the results from these and other robustness analyses in Tables A-3 and A-6.

Second, we show robustness to additional controls. We test the results' robustness by adding distance controls, either between paper authors and the conference or between matched conferences, or by excluding observations characterized by long distance or Asian conferences.[29] In Table A-7, we also add controls for the appliedness of papers on the basis of title and abstract (Boyack et al., 2014), or that of conferences, based on the Journal Commercial Impact Factor (JCIF; Bikard and Marx, 2020). This reduces concerns that a correlation between flight connectivity and the composition of conferences drives the results.

Third, we consider that our data does not reveal which exact author(s) of a paper attend a conference. This may be a minor concern: If some of a given firm's scientists were not at the conference, it would bias our estimates downwards. Furthermore, from the firm's perspective, it is the extensive margin of participation – at least one author presenting against none – that is most relevant. Knowledge flows may even occur through subsequent referrals to other co-authors. However, we need a criterion to allocate papers to regions when a team of authors has multiple affiliations. In the analyses reported, we focus on the first author to determine paper regions and direct flights, since first authors are the most likely participants at conferences,[30] but the results are robust to the alternative of using all authors' affiliations to determine paper regions (Table A-6). Relatedly, when papers are co-authored by a firm and other academic

[29] We include the distance between the researcher region and the conference and whether the two were in the same location (2.1% of the estimation sample).

[30] Typically, in CS, the first author is the project leader associated with a paper, and they are most likely to present it. For one conference we attended, we had access to the detailed program with participants and designated presenters, who were first authors 79% of the time (7% last author, 14% another position).

scientists, we may attribute participation and learning by the academic co-authors to the firm. For example, it may be these co-authors, rather than the firm's scientists, who participate and add citations to future collaborative papers of which the corporate scientists are unaware. We verify that results are similar when we restrict citations from future corporate science to publications in which a corporate scientist is the first author (Table A-6). In this case, it is implausible for the corporate scientist to be unaware of the papers cited.

Finally, firms may passively attend conferences by only buying entrance tickets, which is unobservable to us. Due to this limitation, our conclusions are largely confined to active participation, but we can bound the extent of the issue and whether it constitutes a concern for our study. Importantly, if passive attendance was effective or strengthened the effect for sponsors, we might expect an effect when firms make only minimal contributions or just sponsor. This, however, is not the case even for the largest firms (see Section 6.2). We design an empirical test that isolates cases in which passive attendance is more likely. We consider firms local to the conference venue, who face low costs to attend the conference passively, and find reduced effect sizes for such firms (Table A-8).[31] Relatedly, a firm that actively participates may also send additional scientists to passively participate. We test this by differentiating future science citations from publications authored by conference participants. In this case, compared to citations from unrelated authors, we find a much larger effect (Table A-6). Finding citations from these latter authors is plausible, given that post-conference knowledge-sharing activities within firms emerged as a prominent theme from our interviews. Second, we bound the significance of the phenomenon with manually collected data on the number of conference attendees. We compare the number of paper authors in the bibliometric data with the number of reported attendees, and find that passive attendance appears to be concentrated in a limited number of top conferences and is limited in scale.[32]

8 Conclusion

We provide evidence as to the relevance of scientific conferences to the innovation of participating firms and find that learning remains contingent on corporate participation investments and reputation mechanisms. Our key finding is that participation in the same conference where a firm participates increases the firm's use of a scientist's knowledge, especially when the firm

[31] For this test, we collapse the data at the focal conference $\times$ (matched) paper level to maintain the analysis feasibility despite the large number of local companies. We run analogous regressions for attending firms and find similar results to those in the main analysis. These regressions also show that the results are not driven by conferences with many attending firms, which otherwise receive larger weight in our estimation data.

[32] Focusing on $A^{\star}$ conferences, we obtain attendance counts for 188 conferences. In most cases, attendance was directly proportional to the number of paper authors. About ten conference series (e.g., NeurIPS, SIGGRAPH) were special outliers. For example, NeurIPS 2019 has seen unprecedented attendance but, for this case, we have information about the composition of attendees. The discrepancy with the number of paper authors was, in large part, explained by active participation in workshops, tutorials, and other events peripheral to the scientific program that may not reflect passive attendance by firms. About 25% of tickets were sold by lottery to the general public. Although we cannot ascertain the composition of this group, we expect most to be students, academic scientists, scientists from firms active at the conference, or the interested general public, such as journalists (or one co-author of this manuscript). It is likely that only a subgroup reflects the participation of firms without any other active role. Our own experience as passive attendees supports this assessment. Most importantly, extreme cases such as NeurIPS 2019 seem rare.

contributes scientifically and sponsors the conference. The effect is also seen in future interactions with scientists, such as collaborations and hiring. Finally, we find the effect to be at its greatest among firms and scientists that are already prominent rather than among those lacking alternative opportunities for interaction.

This evidence supports the understanding of scientific conferences as social spaces that facilitate learning but demand specific efforts, with implications for corporate strategies. It builds upon and extends the literature that highlights how knowledge is tied to personal interactions and remains naturally exclusive (Zucker et al., 2002), making specific investments in external learning channels necessary (e.g., Cockburn and Henderson, 1998; Powell et al., 1996; Laursen and Salter, 2006; Gittelman, 2007). Scientific conferences favor face-to-face interactions on a similar basis to geographic proximity, but learning appears unlikely to emerge via knowledge spillovers from temporary colocation alone. Rather, it seems that compliance with social norms and the establishment of effective relations are needed. Accordingly, R&D managers should consider scientific conferences as a valid channel by which to access external knowledge but should focus corporate investments where their firms can achieve sufficient presence and perhaps stand out. Conferences do not appear viable for purely exploratory search. Rather, they provide opportunities to leverage and expand pre-existing research programs.[33]

Finally, the paper contributes to the literature on corporate science (Rosenberg, 1990; Hicks, 1995; Arora et al., 2018; Simeth and Cincera, 2016) and relates to debates on the heterogeneous access of firms to frontier knowledge (Andrews et al., 2015; Autor et al., 2020b; Akcigit and Ates, 2023). The findings suggest that firms participating to scientific conferences access related scientific knowledge without fully internalizing its production. In fact, in light of the decline of corporate investment in internal research (Arora et al., 2021), firms active at conferences may enjoy a more favorable position within external scientific communities, allowing them to better absorb external knowledge. In this sense, firms capable of establishing a reputation and building connections within science-focused social spaces enjoy a competitive advantage relative to others. Although this perspective justifies conference participation, it also implies that effective participation may be costly and include the risk of unintended knowledge spillovers to rivals. Such endeavors may be affordable for only a small number of firms, leading to concentration rather than diffusion in the use of scientific knowledge.

This study has a number of limitations. First, its results could be explained in part by a shift, as opposed to an increase, in the use of scientific knowledge, although this would still be evidence of corporate learning and of an important role of conferences for innovation trajectories. The stronger effect seen for knowledge that is already more salient also speaks against this concern. Second, we introduce exogenous variation in papers' authors' participation but not in firms' participation; interaction coefficients reveal an important pattern of heterogeneity with respect to firms' participation efforts but do not represent causal effects. Resolving this would require

[33]Descriptively, we observe cases of dispersed, minimal, and most likely inexpensive investments in conference participation. These may have low opportunity cost, satisfy the taste for science of corporate scientists, or support objectives other than knowledge acquisition, such as advertising scientific accomplishments, but appear inconsequential for corporate learning.

exogenous variation in each moderator (e.g., sponsorship). Relatedly, we have not explicitly observed passive attendance as a possible mode of participation, and our conclusions on its role remain tentative. Lastly, our results may be generalized to similar contexts but not necessarily all. By virtue of related evidence, we would expect similar results in other scientific fields – e.g., chemistry, pharmaceutics, biotechnology, medicine, engineering (Cohen et al., 2002) – and, beyond these, in events of other communities focused on knowledge production (e.g., Waguespack and Fleming, 2009). On the other hand, our conclusions set scientific conferences apart from more exploratory contexts, such as trade fairs (e.g., Maskell, 2014). We hope our results will inspire future research to overcome such limitations and further improve the understanding of interactions between science and industry.

References

Abadie, A., S. Athey, G. W. Imbens, and J. M. Wooldridge (2023). "When Should You Adjust Standard Errors for Clustering?*." In: *The Quarterly Journal of Economics* 138.1, pp. 1–35.

Ahmadpoor, M. and B. F. Jones (2017). "The dual frontier: Patented inventions and prior scientific advance." In: *Science* 357.6351, pp. 583–587.

Akcigit, U. and S. T. Ates (2023). "What Happened to US Business Dynamism?" In: *Journal of Political Economy* 131.8, pp. 2059–2124.

Alcacer, J. and W. Chung (2007). "Location Strategies and Knowledge Spillovers." In: *Management Science* 53.5, pp. 760–776.

Alcacer, J. and M. Gittelman (2006). "Patent citations as a measure of knowledge flows: The influence of examiner citations." In: *Review of Economics and Statistics* 88.4, pp. 774–779.

Alexy, O., G. George, and A. J. Salter (2013). "Cui Bono? The Selective Revealing of Knowledge and Its Implications for Innovative Activity." In: *Academy of Management Review* 38.2. Publisher: Academy of Management, pp. 270–291.

Almeida, P. and B. Kogut (1999). "Localization of Knowledge and the Mobility of Engineers in Regional Networks." In: *Management Science* 45.7, pp. 905–917.

Andrews, D., C. Criscuolo, and P. N. Gal (2015). "Frontier Firms, Technology Diffusion and Public Policy: Micro Evidence from OECD Countries." In: *OECD Productivity Working Papers* 2, p. 39.

Arora, A., S. Belenzon, and B. Dionisi (2023). "First-Mover Advantage and the Private Value of Public Science." In: *Research Policy* 52.9, p. 104867.

Arora, A., S. Belenzon, and A. Patacconi (2018). "The decline of science in corporate R&D." In: *Strategic Management Journal* 39.1, pp. 3–32.

Arora, A., S. Belenzon, and L. Sheer (2021). "Knowledge Spillovers and Corporate Investment in Scientific Research." In: *American Economic Review* 111.3. ZSCC: 0000043, pp. 871–898.

Arora, A. and A. Gambardella (1994). "Evaluating Technological Information and Utilizing It. Scientific Knowledge, Technological Capability, and External Linkages in Biotechnology." In: *Journal of Economic Behavior and Organization* 24.1, pp. 91–114.

Audretsch, D. B. and P. E. Stephan (1996). "Company-Scientist Locational Links : The Case of Biotechnology." In: *American Economic Review* 86.3, pp. 641–652.

Autor, D., D. Dorn, G. H. Hanson, G. Pisano, and P. Shu (2020a). "Foreign Competition and Domestic Innovation: Evidence from US Patents." In: *American Economic Review: Insights* 2.3, pp. 357–374.

Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. V. Reenen (2020b). "The fall of the labor share and the rise of superstar firms." In: *Quarterly Journal of Economics* 135.2, pp. 645–709.

Azoulay, P., T. Stuart, and Y. Wang (2014). "Matthew: Effect or Fable?" In: *Management Science* 60.1, pp. 92–109.

Bathelt, H. and P. Cohendet (2014). "The creation of knowledge: local building, global accessing and economic development–toward an agenda." In: *Journal of Economic Geography* 14.5, pp. 869–882.

Belenzon, S. and M. Schankerman (2013). "Spreading the Word: Geography, Policy, and Knowledge Spillovers." In: *Review of Economics and Statistics* 95.3, pp. 884–903.

Bikard, M. (2018). "Made in Academia: The Effect of Institutional Origin on Inventors' Attention to Science." In: *Organization Science* 29.5, pp. 818–836.

Bikard, M. and M. Marx (2020). "Bridging Academia and Industry: How Geographic Hubs Connect University Science and Corporate Technology." In: *Management Science* 66.8, pp. 3425–3443.

Bikard, M. A. and I. Fernandez-Mateo (2022). "Standing on the Shoulders of (Male) Giants: Gender Inequality and the Technological Impact of Scientific Ideas." In: *SSRN Electronic Journal*.

Boudreau, K. J., T. Brady, I. Ganguli, P. Gaule, E. Guinan, A. Hollenberg, and K. R. Lakhani (2017). "A field experiment on search costs and the formation of scientific collaborations." In: *Review of Economics and Statistics* 99.4, pp. 565–576.

Boyack, K. W., M. Patek, L. H. Ungar, P. Yoon, and R. Klavans (2014). "Classification of individual articles from all of science by research level." In: *Journal of Informetrics* 8.1, pp. 1–12.

Breschi, S. and F. Lissoni (2001). "Knowledge spillovers and local innovation systems: a critical survey." In: *Industrial and Corporate Change* 10.4, p. 975.

— (2009). "Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows." In: *Journal of Economic Geography* 9.4, pp. 439–468.

Brynjolfsson, E. and L. M. Hitt (2003). "Computing productivity: Firm-level evidence." In: *Review of Economics and Statistics* 85.4, pp. 793–808.

Campante, F. and D. Yanagizawa-Drott (2017). "Long-Range Growth: Economic Development in the Global Network of Air Links." In: *The Quarterly Journal of Economics*.

Catalini, C., C. Fons-Rosen, and P. Gaulé (2020). "How Do Travel Costs Shape Collaboration?" In: *Management Science*, pp. 1–21.

Cetina, K. K. (1999). *Epistemic cultures: how the sciences make knowledge*. Ed. by Harvard University Press. Vol. 53. 9. Harvard University Press, pp. 1689–1699.

Chai, S. and R. B. Freeman (2019). "Temporary colocation and collaborative discovery: Who confers at conferences." In: *Strategic Management Journal* 40.13, pp. 2138–2164.

Cockburn, I. M. and R. Henderson (1998). "Absorptive capcity, coauthoring behavior, and the organization of research in drug discovery." In: *The Journal of Industrial Economics* 46.2, pp. 157–182.

Cohen, W. M. and D. Levinthal (1990). "Absorptive Capacity: A New Perspective on Learning and Innovation." In: *Administrative Science Quarterly* 35.1. ISBN: 9780750672238, pp. 128–152.

Cohen, W. M., R. R. Nelson, and J. P. Walsh (2002). "Links and impacts: The influence of public research on industrial R&D." In: *Management Science* 48.1, pp. 1–23.

Cole, J. R. and S. Cole (1973). *Social Stratification in Science*. Chicago, IL: Chicago University Press.

Crane, D. (1969). "Social Structure in a Group of Scientists: A Test of the "Invisible College" Hypothesis." In: *American Sociological Review* 34.3, pp. 335–352.

— (1974). *Invisible Colleges: Diffusion of Knowledge in Scientific Communities*. Vol. 15. 1, p. 139.

Dahlander, L. and M. W. Wallin (2006). "A Man on the inside: Unlocking Communities as Complementary Assets." In: *Research Policy* 35.8, pp. 1243–1259.

Dalle, J.-m. and P. A. David (2003). "The Allocation of Software Development Resources in 'Open Source' Production Mode." In: Working paper, Stanford University, Stanford, CA.

Dasgupta, P. and P. A. David (1994). "Toward a new economics of science." In: *Research Policy* 23.5, pp. 487–521.

Dechenaux, E., B. Goldfarb, S. Shane, and M. Thursby (2008). "Appropriability and commercialization: Evidence from MIT inventions." In: *Management Science* 54.5, pp. 893–906.

Fleming, L. and O. Sorenson (2004). "Science as a map in technological search." In: *Strategic Management Journal* 25.89, pp. 909–928.

Fleming, L. and D. M. Waguespack (2007). "Brokerage, Boundary Spanning, and Leadership in Open Innovation Communities." In: *Organization Science* 18.2, pp. 165–180.

Foerderer, J. (2020). "Interfirm exchange and innovation in platform ecosystems: Evidence from apple's worldwide developers conference." In: *Management Science* 66.10, pp. 4772–4787.

Fombrun, C. and M. Shanley (1990). "What's in a Name? Reputation Building and Corporate Strategy." In: *The Academy of Management Journal* 33.2, pp. 233–258.

Fosfuri, A., M. S. Giarratana, and E. Roca (2015). "Walking a Slippery Line: Investments in Social Values and Product Longevity." In: *Strategic Management Journal* 36.11, pp. 1750–1760.

Franceschet, M. (2010). "The role of conference publications in CS." In: *Communications of the ACM* 53.12, pp. 129–132.

Giroud, X. (2013). "Proximity and investment: Evidence from plant-level data." In: *Quarterly Journal of Economics* 128.2, pp. 861–915.

Gittelman, M. (2007). "Does geography matter for science-based firms? Epistemic communities and the geography of research and patenting in biotechnology." In: *Organization Science* 18.4, pp. 724–741.

Gittelman, M. and B. Kogut (2003). "Does good science lead to valuable knowledge? Biotechnology firms and the evolutionary logic of citation patterns." In: *Management Science* 49.4, pp. 366–382.

Haas, M. R. and S. Park (2010). "To Share or Not to Share? Professional Norms, Reference Groups, and Information Withholding Among Life Scientists." In: *Organization Science* 21.4, pp. 873–891.

Haeussler, C. (2011). "Information-sharing in academia and the industry: A comparative study." In: *Research Policy* 40.1, pp. 105–122.

Haeussler, C., L. Jiang, J. Thursby, and M. Thursby (2014). "Specific and general information sharing among competing academic researchers." In: *Research Policy* 43.3, pp. 465–475.

Hansen, T. T. and B. D. Pedersen (2018). "The impact of academic events—A literature review." In: *Research Evaluation* 27.4, pp. 358–366.

Hartmann, P. and J. Henkel (2020). "The Rise of Corporate Science in AI: Data as a Strategic Resource." In: *Academy of Management Discoveries.*

Hauss, K. (2021). "What are the social and scientific benefits of participating at academic conferences? Insights from a survey among doctoral students and postdocs in Germany." In: *Research Evaluation* 30.1, pp. 1–12.

Hicks, D. (1995). "Published papers, tacit competencies and corporate management of the public/private character of knowledge." In: *Industrial and Corporate Change* 4.2, pp. 401–424.

Jaffe, A. B., M. Trajtenberg, and R. Henderson (1993). "Geographic localization of knowledge spillovers as evidenced by patent citations." In: *Quarterly Journal of Economics* 108.3, pp. 577–598.

Jones, B. F. (2009). "The Burden of Knowledge and the Death of the Renaissance Man: Is Innovation Getting Harder?" In: *Review of Economic Studies* 76.1, pp. 283–317.

Knaus, J. and M. Palzenberger (2018). "PARMA. A Full Text Search Based Method for Matching Non-Patent Literature Citations with Scientiic Reference Databases. A Pilot Study." In: *Technical Report by the Max Planck Digital Library, Big Data Analytics Group.*

Koffi, M. and M. Marx (2022). "Cassatts in the Attic." In: *NBER Summer Institute, Science of Science Funding.*

Lane, J. N., I. Ganguli, P. Gaule, E. Guinan, and K. R. Lakhani (2021). "Engineering serendipity: When does knowledge sharing lead to knowledge production?" In: *Strategic Management Journal* 42.6, pp. 1215–1244.

Laursen, K. and A. J. Salter (2006). "Open for innovation: The role of openness in explaining innovation performance among U.K. manufacturing firms." In: *Strategic Management Journal* 27.2, pp. 131–150.

Lee, G. K. and R. E. Cole (2003). "From a Firm-Based to a Community-Based Model of Knowledge Creation: The Case of the Linux Kernel Development." In: *Organization Science* 14.6. Publisher: INFORMS: Institute for Operations Research, pp. 633–649.

Lopez de Leon, F. L. and B. McQuillin (2020). "The Role of Conferences on the Pathway to Academic Impact." In: *Journal of Human Resources* 55.1, pp. 164–193.

Maskell, P. (2014). "Accessing remote knowledge-the roles of trade fairs, pipelines, crowdsourcing and listening posts." In: *Journal of Economic Geography* 14.5, pp. 883–902.

Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. Chicago: University of Chicago Press.

Mokyr, J. (2011). *The gifts of athena: Historical origins of the knowledge economy*. Princeton University Press, p. 359.

Mukherjee, A. and S. Stern (2009). "Disclosure or secrecy? The dynamics of Open Science." In: *International Journal of Industrial Organization* 27.3, pp. 449–462.

Nelson, R. R. (1982). "The Role of Knowledge in R&D Efficiency." In: *The Quarterly Journal of Economics* 97.3, pp. 453–470.

Newell, A., A. J. Perlis, and H. A. Simon (1967). "Computer science." In: *Science* 157.3795, pp. 1373–1374.

Perkmann, M., R. Salandra, V. Tartari, M. McKelvey, and A. Hughes (2021). "Academic Engagement: A Review of the Literature 2011-2019." In: *Research Policy* 50.1.

Poege, F., D. Harhoff, F. Gaessler, and S. Baruffaldi (2019). "Science quality and the value of inventions." In: *Science Advances* 5.12, eaay7323.

Polidoro, F. and M. Theeke (2012). "Getting Competition Down to a Science: The Effects of Technological Competition on Firms' Scientific Publications." In: *Organization Science* 23.4, pp. 1135–1153.

Powell, W. W., K. W. Koput, and L. Smith-Doerr (1996). "Interorganizational Collaboration and the Locus of Innovation: Networks of Learning in Biotechnology." In: *Administrative Science Quarterly* 41.1, p. 116.

Rao, H. (1994). "The Social Construction of Reputation: Certification Contests, Legitimation, and the Survival of Organizations in the American Automobile Industry: 1895–1912." In: *Strategic Management Journal* 15.S1, pp. 29–44.

Rindova, V. P., I. O. Williamson, A. P. Petkova, and J. M. Sever (2005). "Being Good or Being Known: An Empirical Examination of the Dimensions, Antecedents, and Consequences of Organizational Reputation." In: *The Academy of Management Journal* 48.6, pp. 1033–1049.

Roach, M. and W. M. Cohen (2013). "Lens or Prism? Patent Citations as a Measure of Knowledge Flows from Public Research." In: *Management Science* 59.2, pp. 504–525.

Rosenberg, N. (1990). "Why do firms do basic research (with their own money)?" In: *Research Policy* 19.2, pp. 165–174.

Schnitzer, M. and M. Watzinger (2019). "Standing on the Shoulders of Science." In: *CEPR Discussion Paper* No. DP1376.

Simeth, M. and M. Cincera (2016). "Corporate Science, Innovation, and Firm Value." In: *Management Science* 62.7, pp. 1970–1981.

Simeth, M. and J. D. Raffo (2013). "What makes companies pursue an Open Science strategy?" In: *Research Policy* 42.9, pp. 1531–1543.

Singh, J. and M. Marx (2013). "Geographic Constraints on Knowledge Spillovers: Political Borders vs. Spatial Proximity." In: *Management Science* 59.9, pp. 2056–2078.

Sorenson, O. (2014). "Status and Reputation: Synonyms or Separate Concepts?" In: *Strategic Organization* 12.1, pp. 62–69.

Stephan, P. E. (1996). "The economics of science." In: *Journal of Economic literature* 34.3, pp. 1199–1235.

Stern, S. (2004). “Do Scientists Pay to Be Scientists?” In: *Management Science* 50.6, pp. 835–853.

Tartari, V., M. Perkmann, and A. Salter (2014). “In Good Company: The Influence of Peers on Industry Engagement by Academic Scientists.” In: *Research Policy* 43.7, pp. 1189–1203.

Tartari, V. and A. Salter (2015). “The engagement gap:: Exploring gender differences in University – Industry collaboration activities.” In: *Research Policy* 44.6. ZSCC: 0000121, pp. 1176–1191.

Verluise, C. and G. De Rassenfosse (2021). *patCit*. https://github.com/cverluise/PatCit.

Vlasov, S. A., M. D. Bahlmann, and J. Knoben (2016). “A study of how diversity in conference participation relates to SMEs’ innovative performance.” In: *Journal of Economic Geography*.

Wagner, S., K. Hoisl, and G. Thoma (2014). “Overcoming localization of knowledge - the role of professional service firms.” In: *Strategic Management Journal* 35.11, pp. 1671–1688.

Waguespack, D. M. and L. Fleming (2009). “Scanning the Commons? Evidence on the Benefits to Startups Participating in Open Standards Development.” In: *Management Science* 55.2. Publisher: INFORMS, pp. 210–223.

ZEW (Centre for European Economic Research) (2018). *Results of CIS 2016 for Germany*. Tech. rep. https://www.zew.de/en/publications/zew-expertises-research-reports/research-reports/innovations/community-innovation-survey-cis.

Zucker, L. G., M. R. Darby, and M. B. Brewer (2008). “Intellectual Human Capital and the Birthh of U . S . Biotechnology Enterprises.” In: *American Economic Review* 88.1, pp. 290–306.

Zucker, L. G., M. R. Darby, and J. S. Armstrong (2002). “Commercializing Knowledge: University Science, Knowledge Capture, and Firm Performance in Biotechnology.” In: *Management Science* 48.1, pp. 138–153.

ONLINE APPENDIX

Contents

A Supplementary results

A.1 References for Table 2

Conference (proceedings):

- NeurIPS '09: 23rd Annual Conference on Neural Information Processing Systems 2009. Proceedings of a meeting held 7-10 December 2009, Vancouver, British Columbia, Canada. (Note that at the time, the acronym NIPS was used. For consistency, we use the current name.)
- IJCAI 2009: Proceedings of the 21st International Joint Conference on Artificial Intelligence, Pasadena, California, USA, July 11-17, 2009
- ICCV '09: IEEE 12th International Conference on Computer Vision, ICCV 2009, Kyoto, Japan, September 27 - October 4, 2009

Conference papers:

- Kurt T. Miller, Thomas L. Griffiths, Michael I. Jordan: Nonparametric Latent Feature Models for Link Prediction. NeurIPS 2009: 1276-1284
- Youssef Hamadi, Saïd Jabbour, Lakhdar Sais: Control-Based Clause Sharing in Parallel SAT Solving. IJCAI 2009: 499-504
- Zilong Dong, Guofeng Zhang, Jiaya Jia, Hujun Bao: Keyframe-based real-time camera tracking. ICCV 2009: 1538-1545

A.2 Second stage results

Table A-1: The effect of *Same conference* for all dependent variables

	(1) Patent citation	(2) Science citation	(3) Patent cit to past	(4) Science cit to past	(5) Collaboration	(6) Hiring
Same conference	0.304***	0.659***	2.748***	5.476***	1.664***	0.175
	(0.118)	(0.167)	(0.885)	(1.352)	(0.488)	(0.193)
Science cit to past$_{pre-conf}$	0.593***	2.497***	21.598***	42.387***	14.572***	3.497***
	(0.046)	(0.133)	(0.366)	(0.350)	(0.259)	(0.121)
Patent cit to past$_{pre-conf}$	0.830***	1.430***	23.185***	19.331***	8.585***	1.724***
	(0.112)	(0.160)	(0.435)	(0.410)	(0.443)	(0.172)
Research similarity	0.959***	2.556***	13.705***	20.637***	7.969***	1.977***
	(0.168)	(0.270)	(1.171)	(2.016)	(0.721)	(0.291)
Method	IV	IV	IV	IV	IV	IV
Conf Ser FE	Yes	Yes	Yes	Yes	Yes	Yes
Origin × Field FE	Yes	Yes	Yes	Yes	Yes	Yes
Origin × Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year × Origin FE	Yes	Yes	Yes	Yes	Yes	Yes
Year × Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7344254	7344254	7344254	7344254	7344254	7344254
Number clusters	1088	1088	1088	1088	1088	1088
DV mean (% points)	0.170	0.463	5.258	8.779	2.661	0.598
F (excl. Inst.)	83.8	83.8	83.8	83.8	83.8	83.8

Notes: IV estimates. $^{*}p < .1$, $^{**}p < .05$, $^{***}p < .01$. Standard errors in parentheses, clustered at the level of the paper authors' origin region. Columns 1 and 2 analyze the probability of a citation from firm patents and firm science, respectively, in the subsequent five years towards the focal paper. Column 3 and 4 consider citations toward CS publications by the paper authors in the five years before the conference. Column 5 considers whether a paper authors subsequently jointly authored a publication with the firm. Column 6 considers whether any paper author becomes a researcher of the firm. All variables are in percentage points. Data is at the firm-conference-paper level. *Same conference* is assigned 1 for papers presented at the conference in which the firm participates and 0 for papers presented at a matched conference. We instrument *Same conference* by the direct flight availability between the paper authors' location and the conference venue. Firm-paper controls include indicators for patent or science citations from the firm to the authors in the years before the conference and how similar the firm's publications in the previous year were to the focal paper. F (excl. Inst.) is the first-stage Kleibergen-Paap F-statistic.

Table A-2: OLS/IV results for all outcome variables

Regression	Patent citation	Science citation	Patent cit to past	Science cit to past	Collaboration	Hiring	
	Coefficient (Std. Err.)	Coefficient (Std. Err.)	Coefficient (Std. Err.)	Coefficient (Std. Err.)	Coefficient (Std. Err.)	Coefficient (Std. Err.)	F (excl. Inst.) (N clusters)
Firm and Year FE only: OLS	0.136***	0.455***	2.659***	4.965***	1.872***	0.434***	.
	(0.012)	(0.035)	(0.121)	(0.218)	(0.076)	(0.028)	(1288)
Firm and Year FE only: IV	1.222***	3.511***	21.651***	40.562***	10.962***	2.573***	102.7
	(0.187)	(0.476)	(2.811)	(4.881)	(1.110)	(0.380)	(1288)
Baseline without controls and yearly FE: OLS	0.156***	0.509***	2.696***	4.824***	1.749***	0.414***	.
	(0.013)	(0.037)	(0.106)	(0.202)	(0.079)	(0.029)	(1097)
Baseline without controls and yearly FE: IV	0.422***	0.759***	4.203***	7.248***	2.360***	0.481**	79.0
	(0.116)	(0.176)	(1.081)	(1.654)	(0.586)	(0.217)	(1097)
Baseline without controls: OLS	0.159***	0.511***	2.708***	4.790***	1.741***	0.417***	.
	(0.013)	(0.037)	(0.109)	(0.202)	(0.080)	(0.029)	(1088)
Baseline without controls: IV	0.342***	0.779***	3.750***	7.151***	2.267***	0.318	81.7
	(0.119)	(0.170)	(1.013)	(1.631)	(0.543)	(0.194)	(1088)
Baseline: OLS	0.118***	0.388***	1.805***	3.208***	1.183***	0.289***	.
	(0.010)	(0.027)	(0.068)	(0.131)	(0.056)	(0.022)	(1088)
Baseline: IV	0.304***	0.659***	2.748***	5.476***	1.664***	0.175	83.8
	(0.118)	(0.167)	(0.885)	(1.352)	(0.488)	(0.193)	(1088)

Notes: OLS and IV estimates. $^{*}p < .1$, $^{**}p < .05$, $^{***}p < .01$. Standard errors in parentheses, clustered at the level of the paper authors' origin region. The baseline contains fixed effects for conference, origin×field, origin×firm, year×origin and year×firm, as well as additional control variables. All variables are in percentage points. Data is at the firm-conference-paper level. *Same conference* is assigned 1 for papers presented at the conference in which the firm participates and 0 for papers presented at a matched conference. We instrument *Same conference* by the direct flight availability between the paper authors' location and the conference venue. Firm-paper controls include indicators for patent or science citations from the firm to the authors in the years before the conference and how similar the firm's publications in the previous year were to the focal paper. F (excl. Inst.) is the first-stage Kleibergen-Paap F-statistic.

Table A-3: Robustness checks with additional fixed effects or different cluster levels

Regression	Patent citation	Science citation	Patent cit to past	Science cit to past	Collaboration	Hiring	
Same conference	Coefficient (Std. Err.)	Coefficient (Std. Err.)	Coefficient (Std. Err.)	Coefficient (Std. Err.)	Coefficient (Std. Err.)	Coefficient (Std. Err.)	F (excl. Inst.) (N clusters)
Baseline	0.304***	0.659***	2.748***	5.476***	1.664***	0.175	83.8
	(0.118)	(0.167)	(0.885)	(1.352)	(0.488)	(0.193)	(1088)
Baseline plus firm × researcher region × year FE	0.431**	0.740***	3.644***	6.730***	1.972***	0.201	89.3
	(0.178)	(0.224)	(1.230)	(1.618)	(0.692)	(0.278)	(1041)
Baseline plus conference location × year FE	0.270***	0.565***	2.075***	4.524***	1.388***	0.120	81.3
	(0.102)	(0.157)	(0.746)	(1.231)	(0.455)	(0.168)	(1088)
Baseline plus matched conference series FE	0.299**	0.584***	2.496***	4.693***	1.498***	0.156	96.6
	(0.121)	(0.172)	(0.895)	(1.353)	(0.507)	(0.198)	(1088)
Baseline, cluster at firm-level	0.304*	0.659**	2.748***	5.476***	1.664**	0.175	1535.0
	(0.166)	(0.271)	(0.805)	(0.784)	(0.663)	(0.204)	(3826)
Baseline, cluster at researcher region	0.304***	0.659***	2.748***	5.476***	1.664***	0.175	213.1
× conference region level	(0.106)	(0.198)	(0.723)	(0.964)	(0.464)	(0.190)	(103859)
Baseline, twoway cluster at researcher region	0.304**	0.659***	2.748**	5.476***	1.664***	0.175	70.1
and conference region level	(0.145)	(0.208)	(1.086)	(1.604)	(0.642)	(0.223)	(1088 × 507)
Baseline, twoway cluster at researcher region	0.304*	0.659**	2.748***	5.476***	1.664**	0.175	81.6
and firm level	(0.179)	(0.275)	(1.051)	(1.405)	(0.713)	(0.207)	(1088 × 3826)
Baseline, cluster at paper level	0.304***	0.659***	2.748***	5.476***	1.664***	0.175	1399.4
	(0.096)	(0.171)	(0.522)	(0.718)	(0.365)	(0.169)	(344325)

Notes: IV estimates. $^{*}p < .1$, $^{**}p < .05$, $^{***}p < .01$. Standard errors in parentheses, clustered at the level of the paper authors' origin region. Other cluster levels are noted in Column 1. All variables are in percentage points. Data is at the firm-conference-paper level. *Same conference* is assigned 1 for papers presented at the conference in which the firm participates and 0 for papers presented at a matched conference. We instrument *Same conference* by the direct flight availability between the paper authors' location and the conference venue. Firm-paper controls include indicators for patent or science citations from the firm to the authors in the years before the conference and how similar the firm's publications in the previous year were to the focal paper. F (excl. Inst.) is the first-stage Kleibergen-Paap F-statistic. The baseline contains fixed effects for conference series, origin × field, origin × firm, year × origin and year × firm, as well as additional control variables.

Table A-4: Alternative variables and firm-level heterogeneity

Regression	Patent citation	Science citation	Patent cit to past	Science cit to past	Collaboration	Hiring	
	Coefficient (Std. Err.)	Coefficient (Std. Err.)	Coefficient (Std. Err.)	Coefficient (Std. Err.)	Coefficient (Std. Err.)	Coefficient (Std. Err.)	F (excl. Inst.) (N clusters)
Same conference	-0.657^{***}	-1.313^{***}	-2.799^{***}	-2.261^{*}	-1.501^{**}	-0.830^{***}	39.8
	(0.156)	(0.300)	(0.780)	(1.256)	(0.618)	(0.260)	(1088)
× Research similarity	11.190^{***}	22.962^{***}	64.607^{***}	90.110^{***}	36.855^{***}	11.702^{***}	
	(2.382)	(3.892)	(11.603)	(16.940)	(6.907)	(2.747)	
Same conference	1.401^{**}	3.513^{*}	5.483	7.587	9.228^{**}	0.358	5.9
	(0.630)	(1.805)	(4.247)	(5.432)	(4.131)	(1.602)	(1088)
× Geo. distance to paper	-0.139^{*}	-0.360^{*}	-0.345	-0.266	-0.955^{*}	-0.023	
	(0.077)	(0.219)	(0.470)	(0.572)	(0.503)	(0.188)	
Same conference	-0.097	-0.275^{*}	1.382^{*}	3.717^{***}	-0.796	-0.459^{**}	43.8
	(0.081)	(0.141)	(0.790)	(1.329)	(0.552)	(0.198)	(1088)
× Firm in science hub	0.615^{***}	1.434^{***}	2.095^{**}	2.697^{**}	3.771^{***}	0.972^{***}	
	(0.206)	(0.242)	(1.012)	(1.165)	(0.754)	(0.371)	
Same conference	-1.167^{***}	-2.193^{***}	-8.776^{***}	-8.401^{***}	-3.818^{***}	-0.937^{*}	26.2
	(0.259)	(0.393)	(1.777)	(1.789)	(1.136)	(0.478)	(1088)
× Firm scientists n.	0.329^{***}	0.638^{***}	2.580^{***}	3.106^{***}	1.227^{***}	0.249^{*}	
	(0.075)	(0.106)	(0.503)	(0.478)	(0.298)	(0.132)	

Notes: This table extends Table 6 to alternative dependent variables. IV estimates. $^{*}p < .1$, $^{**}p < .05$, $^{***}p < .01$. Standard errors in parentheses, clustered at the level of the paper authors' origin region. All variables are in percentage points. Data is at the firm-conference-paper level. *Same conference* is assigned 1 for papers presented at the conference in which the firm participates and 0 for papers presented at a matched conference. We instrument *Same conference* by the direct flight availability between the paper authors' location and the conference venue. Firm-paper controls include indicators for patent or science citations from the firm to the authors in the years before the conference and how similar the firm's publications in the previous year were to the focal paper. F (excl. Inst.) is the first-stage Kleibergen-Paap F-statistic. The baseline contains fixed effects for conference, origin × field, origin × firm, year × origin and year × firm, as well as additional control variables.

Table A-5: Alternative variables and paper-level heterogeneity

Regression	Patent citation	Science citation	Patent cit to past	Science cit to past	Collaboration	Hiring	
	Coefficient (Std. Err.)	Coefficient (Std. Err.)	Coefficient (Std. Err.)	Coefficient (Std. Err.)	Coefficient (Std. Err.)	Coefficient (Std. Err.)	F (excl. Inst.) (N clusters)
Same conference	−0.016	−0.370*	−1.236	−1.817	0.495	0.105	44.6
	(0.088)	(0.200)	(1.061)	(1.557)	(0.494)	(0.205)	(1088)
× Paper team productivity	0.474***	1.520***	5.342***	9.770***	1.459*	0.041	
	(0.146)	(0.350)	(1.692)	(2.491)	(0.836)	(0.366)	
Same conference	0.062	−0.221	−0.795	−1.403	0.479	0.175	29.4
	(0.087)	(0.198)	(1.052)	(1.552)	(0.502)	(0.210)	(1085)
× Paper team productivity	0.480***	1.524***	5.226***	9.693***	1.417*	0.015	
	(0.145)	(0.352)	(1.696)	(2.499)	(0.834)	(0.366)	
× Paper team female	−0.475***	−0.932***	−2.558*	−2.556	0.115	−0.403	
	(0.168)	(0.279)	(1.472)	(1.620)	(0.708)	(0.261)	
Same conference	0.248**	0.581***	2.799***	5.499***	2.215***	0.258	9.9
	(0.105)	(0.165)	(0.915)	(1.405)	(0.491)	(0.188)	(1088)
× Paper by other firm	0.691*	0.947	−0.773	−0.478	−6.777**	−1.036	
	(0.409)	(0.737)	(3.151)	(4.743)	(2.946)	(0.903)	

Notes: This table extends Table 7 to alternative dependent variables. IV estimates. $^{*}p < .1$, $^{**}p < .05$, $^{***}p < .01$. Standard errors in parentheses, clustered at the level of the paper authors' origin region. All variables are in percentage points. Data is at the firm-conference-paper level. *Same conference* is assigned 1 for papers presented at the conference in which the firm participates and 0 for papers presented at a matched conference. We instrument *Same conference* by the direct flight availability between the paper authors' location and the conference venue. Firm-paper controls include indicators for patent or science citations from the firm to the authors in the years before the conference and how similar the firm's publications in the previous year were to the focal paper. F (excl. Inst.) is the first-stage Kleibergen-Paap F-statistic. The baseline contains fixed effects for conference, origin × field, origin × firm, year × origin and year × firm, as well as additional control variables.

Table A-6: Various robustness specifications

	Patent citation				
	(1) Baseline	(2) Front page	(3) US only	(4) EP/WIPO	(5) Applicant only
Same conference	0.304***	0.288**	0.271**	0.096**	0.236**
	(0.118)	(0.112)	(0.109)	(0.048)	(0.101)
Paper-level controls	Yes	Yes	Yes	Yes	Yes
Standard FE	Yes	Yes	Yes	Yes	Yes
Observations	7344254	7344254	7344254	7344254	7344254
Number clusters	1088	1088	1088	1088	1088
DV mean (% points)	0.170	0.162	0.150	0.040	0.125
F (excl. Inst.)	83.8	83.8	83.8	83.8	83.8
	Patent citation				
	(6) Author not at Counterf conf	(7) No long dist scientist-conf	(8) No long dist conf-conf	(9) Counterf conf not Asia-Pac	(10) Dist control
Same conference	0.275**	0.257*	0.459**	0.411***	0.733***
	(0.119)	(0.147)	(0.189)	(0.157)	(0.263)
Paper-level controls	Yes	Yes	Yes	Yes	Yes
Standard FE	Yes	Yes	Yes	Yes	Yes
Observations	7090426	5779858	5885910	4220053	7344254
Number clusters	1086	1027	1042	985	1088
DV mean (% points)	0.170	0.170	0.170	0.170	0.170
F (excl. Inst.)	82.7	62.4	52.4	89.2	37.3
	Any firm science citation to conference paper				
	(11) Baseline	(12) No self-cit	(13) First affil	(14) Author at conf	(15) Not at conf
Same conference	0.659***	0.538***	0.297**	0.544***	0.237**
	(0.167)	(0.151)	(0.125)	(0.124)	(0.119)
Paper-level controls	Yes	Yes	Yes	Yes	Yes
Standard FE	Yes	Yes	Yes	Yes	Yes
Observations	7344254	7344254	7344254	7344254	7344254
Number clusters	1088	1088	1088	1088	1088
DV mean (% points)	0.463	0.396	0.240	0.223	0.275
F (excl. Inst.)	83.8	83.8	83.8	83.8	83.8
	Science citation		Text similarity	IV: All authors	
	(16) Conf. papers	(17) Articles	(18) Mean, post	(19) Pat cit	(20) Sci cit
Same conference	0.516***	0.270**	1.858***	0.305**	0.616***
	(0.130)	(0.113)	(0.284)	(0.123)	(0.161)
Paper-level controls	Yes	Yes	Yes	Yes	Yes
Standard FE	Yes	Yes	Yes	Yes	Yes
Observations	7344254	7344254	7344254	7365968	7365968
Number clusters	1088	1088	1088	1190	1190
DV mean (% points)	0.322	0.201	10.020	0.170	0.463
F (excl. Inst.)	83.8	83.8	84.9	86.7	86.7

Notes: See next page.

Notes to Table A-6: IV estimates. $^{*}p < .1$, $^{**}p < .05$, $^{***}p < .01$. Standard errors in parentheses, clustered at the level of the paper authors' origin region. All variables are in percentage points. Data is at the firm-conference-paper level. *Same conference* is assigned 1 for papers presented at the conference in which the firm participates and 0 for papers presented at a matched conference. We instrument *Same conference* by the direct flight availability between the paper authors' location and the conference venue. Firm-paper controls include indicators for patent or science citations from the firm to the authors in the years before the conference and how similar the firm's publications in the previous year were to the focal paper. F (excl. Inst.) is the first-stage Kleibergen-Paap F-statistic.Column 1 reports the baseline for patent citations from Table A-1. Column 2 focuses on front-page patent citations only. Column 3 focuses on citations from USPTO patents only, whereas Column 4 restricts to citations from EPO/WIPO patent (applications). Column 5 focuses on citations generated by applicants (i.e., excluding citations by examiners or other sources). In Column 6, as a robustness check for the sample composition, we exclude all observations with authors that attended more than one conference in a choice set. In Columns 7-10, we focus on robustness to geographic restrictions and controls. In Column 7, we remove observations with distances between the author and the conference of more than 10,000 km. In Column 8, we remove matched conferences that are more than 10,000 km from the respective focal conference (we also remove the focal conference if no matched conference remains). In Column 9, we drop matched conferences from the Asia-Pacific region, which are also often associated with long travel distances. In Column 10, we add geographic control variables, including the distance between the paper's authors and the conference and an indicator for whether they are in the same state or region. Next, we return to regressions with science citations as the dependent variable. Column 11 reports the baseline from Table A-1. In Column 12, we exclude author-level self-citations. All regressions exclude firm-paper pairs that might involve firm-level self-citations. In Column 13, we restrict science citations to such from firm publications where the firm is the first affiliation. In Columns 14-15, we analyze whether science citations are more likely from firm authors that were part of the authors of the firm papers at the focal conference, i.e., more likely to have participated in the focal conference (Column 14). As a comparison, we analyze follow-on science citations from firm authors who were unlikely to attend the focal conference (Column 15). We restrict the CS publications that firms' science citations can come from to conference papers only (Column 16) and journal articles only (Column 17). Next, Column 18 studies the similarity of future firm papers and the focal conference proceeding. Finally, we show specifications with an alternative instrument based on all author locations (Columns 19 and 20).

Table A-7: Robustness checks with additional control variables

Regression	Patent citation	Science citation	Patent cit to past	Science cit to past	Collaboration	Hiring	
	Coefficient (Std. Err.)	Coefficient (Std. Err.)	Coefficient (Std. Err.)	Coefficient (Std. Err.)	Coefficient (Std. Err.)	Coefficient (Std. Err.)	F (excl. Inst.) (N clusters)
Same conference	0.320***	0.676***	2.980***	5.582***	1.621***	0.204	82.6
	(0.119)	(0.169)	(0.909)	(1.380)	(0.496)	(0.193)	(1086)
Appliedness (Abstract)	0.136***	0.160***	2.166***	2.055***	0.862***	0.199***	
	(0.015)	(0.025)	(0.138)	(0.181)	(0.076)	(0.029)	
Same conference	0.304***	0.629***	2.730***	5.428***	1.658***	0.182	83.7
	(0.117)	(0.164)	(0.886)	(1.350)	(0.488)	(0.194)	(1088)
Appliedness (JCIF)	−0.004	0.785***	0.478	1.241***	0.142	−0.178*	
	(0.078)	(0.126)	(0.365)	(0.366)	(0.196)	(0.105)	
Same conference	0.383***	1.132***	4.158***	8.439***	2.787***	0.471**	74.0
	(0.135)	(0.206)	(1.239)	(1.712)	(0.579)	(0.208)	(1088)
Appliedness (JCIF)	1.558***	3.772***	10.835***	13.783***	2.975***	1.024***	
(Conf. Series FEomitted)	(0.138)	(0.249)	(0.498)	(0.595)	(0.214)	(0.112)	

Notes: IV estimates. $^{*}p < .1$, $^{**}p < .05$, $^{***}p < .01$. Standard errors in parentheses, clustered at the level of the paper authors' origin region. Appliedness (Abstract)is an appliedness score (range 0-1) based on the paper title and abstract, following Boyack et al. (2014). Appliedness (JCIF)is the journal commercial impact factor of the focal conference series, calculated following Bikard and Marx (2020). As the JCIF is highly correlated with conference series FE, we omit these in the third specification. All variables are in percentage points. Data is at the firm-conference-paper level. *Same conference* is assigned 1 for papers presented at the conference in which the firm participates and 0 for papers presented at a matched conference. We instrument *Same conference* by the direct flight availability between the paper authors' location and the conference venue. Firm-paper controls include indicators for patent or science citations from the firm to the authors in the years before the conference and how similar the firm's publications in the previous year were to the focal paper. F (excl. Inst.) is the first-stage Kleibergen-Paap F-statistic. The baseline contains fixed effects for conference, origin × field, origin × firm, year × origin and year × firm, as well as additional control variables.

Table A-8: Robustness: Local and participating firms

	(1) Patent cit	(2) Science cit	(3) Patent cit to past	(4) Science cit to past	(5) Collaboration	(6) Hiring
	Activity by local non-attending firms at the conference venue					
Same conference	0.594*	1.399***	13.933***	19.247***	9.613***	2.561*
	(0.326)	(0.533)	(2.071)	(2.522)	(3.070)	(1.403)
DV mean (% points)	0.368	0.764	8.532	11.441	4.704	1.326
	Activity by firms attending the conference					
	(7) Patent cit	(8) Science cit	(9) Patent cit to past	(10) Science cit to past	(11) Collaboration	(12) Hiring
Same conference	2.124***	5.179***	8.035***	13.691***	16.194***	7.156***
	(0.699)	(0.808)	(1.642)	(2.049)	(2.009)	(1.178)
DV mean (% points)	2.513	5.522	33.454	39.055	22.773	7.546
Conf Ser FE	Yes	Yes	Yes	Yes	Yes	Yes
Origin × Field FE	Yes	Yes	Yes	Yes	Yes	Yes
Year × Origin FE	Yes	Yes	Yes	Yes	Yes	Yes
Year × Venue FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	781538	781538	781538	781538	781538	781538
Number clusters	1249	1249	1249	1249	1249	1249
F (excl. Inst.)	138.1	138.1	138.1	138.1	138.1	138.1

Notes: IV estimates. $^{*}p < .1$, $^{**}p < .05$, $^{***}p < .01$. Standard errors in parentheses, clustered at the level of the paper authors' origin region. The regressions study the probability of follow-on activity of firms located close to the conference but not attending (Columns 1-6) or attending the conference (Columns 7-12). In the latter case, the dataset is an aggregate version of the estimation dataset. Due to the large number of local, non-attending firms, a firm-level dataset is not feasible. Venue-year fixed effects take the place of firm-year fixed effects. Variable definitions otherwise remain as in the main regressions, as does the first stage. Fixed effects, observations and first stage in Columns 1-6 correspond to Columns 7-12.

A.3 First stage results

We investigate how the strength of the first stage varies along conference characteristics. We estimate regression A.1 with various sets of i heterogeneity dimensions h_i.

$$\text{Same conference}_{fcp} = \beta_0 \text{Direct flight}_{fcp} + \sum_i \beta_{1i} \left[\text{Direct flight}_{fcp} \times h_{pi} + \beta_{2i} h_{pi} \right] + \beta_4 X_{fcp} + u_{fcp} \tag{A.1}$$

In Table A-9, Column 1 shows that the effect of the instrument is unchanged for larger conferences. However, in Column 2, we find that direct flights matter more for less prominent researchers. In Column 3, we find that scientists from industry - with presumably larger budgets - are less sensitive to the presence of direct flights. The quality level of the conference matters as well; the coefficient size for $A^\star$ and A conferences is smaller. It is plausible for direct flights to matter more when the researchers are otherwise indifferent between similar conferences along with our matching criteria. For example, we expect researchers to spare less to attend $A^\star$ or A conferences. For lower-level conferences, the discomfort of traveling might start to play a stronger role, leading to more instrument relevance for low-ranked conferences.

Table A-9: First stage - Heterogeneity of the effect of *Direct flight* on *Same conference*

	Same conference				
	(1)	(2)	(3)	(4)	(5)
Direct flight	8.697***	8.042***	7.294***	11.079***	
	(1.693)	(1.038)	(0.746)	(1.608)	
× Conference size	−0.321			−0.553	
	(0.360)			(0.341)	
× Paper team female		1.950***		1.953***	
		(0.730)		(0.714)	
× Paper team productivity		−1.679***		−1.681***	
		(0.411)		(0.389)	
× Paper by other firm			−1.749**	−1.289*	
			(0.724)	(0.724)	
× $A^\star$ conference					3.884***
					(0.718)
× A conference					4.788***
					(0.899)
× B conference					8.984***
					(1.041)
× C conference					10.887***
					(1.169)
Standard FE	Yes	Yes	Yes	Yes	Yes
Cluster	Origin	Origin	Origin	Origin	Origin
Number clusters	1088	1085	1088	1085	1088
R^2	0.240	0.239	0.239	0.241	0.239
Observations	7344254	7249463	7344254	7249463	7344254

Notes: First stage estimates. $^*p < .1$, $^{**}p < .05$, $^{***}p < .01$. Standard errors in parenthesis, clustered at the level of the paper authors' region. The dataset, variable definitions, and FE follow the description of Table 7. We omit coefficient values for the interacted variables themselves, but they are included insofar as not colinear with the FE.

B Construction of the estimation dataset

This section extends the discussion in Section 5.1 of how the estimation dataset is created.

We start by drawing a set of matched conferences at which a given paper could have been presented as well. We will pair every conference with other conferences of the same year, rank, subfield, and within the same size and forward-citations count categories. We source the year information from dblp and the quality rank from CORE, considering the ranks A*, A, B, and C conferences. All restrictions are equivalent to those applied for the descriptive analysis, but we repeat here for completeness. We disregard uncategorized and regional conferences and disregard conferences that were not classified to any fields, as well as those not classified to CS or related engineering. We consider only conferences held between 1996 and 2010, for which at least ten papers listed in dblp could be matched to WoS or Scopus. For the field information drawn from CORE, Table B-1 shows the field definition used in the match. In the match, we consider all listed field information if more than one is available. We drop conferences that are uncategorized in CORE or belong to fields outside of CS. We consolidate small fields and fields listing miscellaneous conferences into Computer Science (general) and Engineering (general). The results remain robust to the exclusion of these categories. In the data entering the match, there are 6846 unique conference events, equivalent to the descriptive analysis. 4800 unique firms participated at these conferences. Note that of the initial set of conference events, only 6248 saw participation by a firm. We do not condition the set of potential conferences on firm participation, but this will reduce the number of focal conference events in the final estimation sample compared to the sample for descriptive analysis.

Table B-1: CORE fields.

Field	Conferences			Papers		
	(First)	(All)	(%)	(First)	(All)	(%)
Computer Science (general)	501	592	7.9	40933	44135	7.2
Artificial Intelligence and Image Processing	1320	1353	18.2	161985	172756	28.1
Computation Theory and Mathematics	1013	1029	13.8	45776	46458	7.6
Computer Software	1350	1399	18.8	56430	58340	9.5
Data Format	660	715	9.6	35509	38308	6.2
Distributed Computing	826	880	11.8	56397	58268	9.5
Information Systems	789	928	12.5	77330	84998	13.8
Library and Information Studies	0	72	1.0	0	7135	1.2
Engineering (general)	387	485	6.5	84767	103344	16.8
Total	6846	7453	100.0	559127	613742	100.0

Notes: Shows CORE fields at the conference event and paper-level. Each conference series is associated with up to three CORE fields. Counts are shown counting only conference events with the first or using all CORE fields, shares are for the latter. Shows 1996-2010 data.

Within year-rank-field strata, we create sub-categories by conference size and forward citation count. With this, we account for the potential variation in atmosphere across smaller and larger conferences and quality differences not captured by the ranks. We coarsen the conference size (Split at 25% and 50% of maximum size within the group) and 5-year forward citation counts (Median split for $A^{\star}$ and A conferences, quartiles for B/C conferences).

We then pair each conference with all conferences in the set of strata, including the conference itself, yielding 33,978 pairs. We exclude conference pairs that take place in the nearby airport region (separated by no more than 50km, 1.2% of pairs). We then drop conferences that found no pair (another 4.5% of pairs). At this point, 5,725 focal conferences are paired with on average 4.7 (standard deviation=3.4, min=1, max=17) conferences. 3,828 unique firms participate in the focal conferences that remain in the data at this point. In order to not further increase the weight of large fields, we retain up to two other conferences selected randomly within the same strata. After this restriction, each focal conference is paired with on average 1.8 conferences. The results are robust to different choices, e.g. five other conferences.

Table B-2: Covariate balancing table.

	Average value (Actual conf)	Difference (SE) (Matched-Actual)		p-value
Exact matching criteria				
Year	2005.22	0.000	(0.000)	1.00
Rank: A*	0.10	0.000	(0.000)	1.00
Rank: A	0.28	0.000	(0.000)	1.00
Rank: B	0.36	0.000	(0.000)	1.00
Rank: C	0.26	0.000	(0.000)	1.00
Field: General CompSci	0.08	−0.001	(0.003)	0.65
Field: General Engineering	0.05	0.001	(0.001)	0.54
Field: AI / Computer Vision	0.21	−0.001	(0.001)	0.39
Field: Computation Theory	0.16	0.000	(0.001)	0.83
Field: Computer Software	0.22	0.000	(0.001)	0.69
Field: Data Format	0.10	0.003	(0.002)	0.13
Field: Distributed Computing	0.13	0.000	(0.002)	1.00
Field: Information Systems	0.13	0.001	(0.002)	0.58
Coarsened matching criteria				
Size of the conference	70.38	0.242	(0.551)	0.66
Mean 5-year citations	4.17	−0.028	(0.051)	0.58
Untargeted matching criteria				
Conference series age	5.24	−0.015	(0.052)	0.77
Number of fields	1.08	0.004	(0.006)	0.46
Number of sponsors	1.35	−0.003	(0.018)	0.89
Number of firms	4.90	0.000	(0.057)	1.00
Appliedness (Text)	0.79	0.001	(0.001)	0.67
Appliedness (JCIF)	0.07	0.001	(0.002)	0.71
Observations	5725	10355		

Notes: Covariate balancing for retaining up to two matched conferences. Shows the average deviation of the matched conference from the actual conference.

We verified that the average difference for any observable between matched conferences cannot be distinguished from zero. Table B-2 shows the results of this analysis. Note that each conference in the sample is randomly paired with one or more other conferences within the same sample (and according to the matching strata). Consequently, this analysis is not a test of equivalence of paired conferences. It is purely aimed at excluding a malfunction of the matching algorithm and random selection.

C Appendix: Understanding firms' attendance of conferences

C.1 Interviews at ECCV18 and NeurIPS19

To better understand the nature of corporate participation in conferences and the associated learning mechanisms, we attended two important ($A^{\star}$) conferences: the European Conference on Computer Vision 2018 (ECCV, https://eccv2018.org/) in Munich, Germany, and the Conference on Neural Information Processing Systems 2019 (NeurIPS, https://nips.cc/Conferences/2019) in Vancouver, Canada. We interviewed more than 50 people in total, including scientists (mostly from research units and a few from product development units), HR representatives, engineers from more than 20 firms, and about 20 academic scientists. We talked to firms from several countries of various sizes and with different levels of participation. We investigated their activities at conferences and the processes taking place before, during, and after conferences. Falling short of a full qualitative study, we here report our general impressions from the interviews. The evidence discussed is also necessarily selective. Importantly, industry investments at ECCV18 and NeurIPS19, like other conference series in ML, have increased sharply in recent years. Nonetheless, we expect that the type of firm activities carried out at other conferences would be similar when observed in proceedings data. Scientific activities of firms at the conference were reflected in the participation of scientists, presenting their work, and interacting with their academic and corporate peers. Firm scientists reported having a high degree of autonomy in the conference choices and what to present. Firm-level processes, mostly unknown to academic scientists, play a role in the screening of presented work before and in activities after the conference. Interestingly, the screening of submitted papers concerns a selection based on quality: most firms have in place internal peer-review systems (and not hierarchical approval) to ensure presenting above-average scientific work. A screening by an IP staff unit is often also made for potential patent filing to avoid that presentation creating prior art that impedes future patenting, but no scientist declared this to have ever impeded their participation.[1]

After a conference, all firms appeared to have knowledge-sharing processes in place. These may take the form of informal activities, such as the sharing of references among colleagues (including those who did not participate in the conference). More often, researchers had to write more structured reports or prepare presentations about the content of the conference for internal meetings. In some cases, this was supported by an internal IT information system, which traced the participation of different individuals at different conferences and collected information on their feedback.

A second possible type of activity, directly connected with sponsorship, entailed promotion and recruiting activities, mostly carried out by personnel at the conference booths. HR personnel, in particular, advertise job opportunities, mostly for PhDs and young researchers, attend

[1] It also remains that most scientists presenting were not directly related to product development. Some said that the research related to product development is normally under trade secrets, and the few researchers from product development units present and that we interviewed would not present nor talk about their research.

the booth to all potential candidates, and schedule possible follow-up interviews. The HR units then prepare the material, define the main activities before the conference, and have follow-up meetings to discuss the outcomes and possible improvements after the conference.

Despite being distinct activities, scientific and promotion activities complemented each other. The personnel employed at the firm's booths also advertise the scientific activities of the firm's scientists at the conference and can offer organizational and logistic support to scientists. The promotion of research, especially focused on the specific contributions at the focal conference, was at least equally evident as the promotion of job positions. The sponsorship spaces, especially for large sponsors, are also used to create opportunities for divulging research results and to offer tutorials and workshops in addition to conference presentations.

Several HR representatives referred to having experimented with participating in conferences without the presence of scientists. However, this proved ineffective for hiring objectives due to the difficulty of engaging with other scientists. The presence of scientists at the booths is planned to facilitate conversations with potential job candidates, and informal interactions of scientists at the conference may also constitute a vehicle to reach and engage candidates. The promotion of research at the firm is complementary to this objective. Interestingly, most HR we interviewed declared that the decision of the conferences to sponsor often follows the choice of scientists of where to present their research. Conversely, scientists did not seem to base their participation choices on whether the firm sponsored or not the conference. The few sponsors we could talk to with a small booth or no parallel scientific activity demonstrated limited engagement with the conference participants, and their booths were poorly attended. One of these sponsors' representatives (from a large firm) explicitly expressed dissatisfaction with the lack of a more significant investment by the firm, in her/his own words, "to a community that I deem important for our research unit."

C.2 Selected sponsor statements from corporate webpages

We report a selection of statements from corporate web pages dedicated to participation in conferences to demonstrate that their declared motives align with our theoretical discussion. The selection is not representative and intentionally focused on the most active firms. The cases are recent because older ones are rarely available online, so this evidence is also influenced by the recent boom in the field of AI. We highlight the many elements that speak to scale and mode of participation, the intent to contribute and collaborate with the scientific community, and the search for reputation and status signals (e.g., references to quality and awards). This includes references to the symbolic use of sponsorship, which is often accompanied by emphatic expressions such as "proud" or "excited", and to the complementarity between sponsorship and scientific contributions, where firms frequently reference dedicated spaces for sponsors as an opportunity to interact with them and learn about research. All websites were accessed between February and April 2022. Emphases are added.

- NeurIPS, 2021. Google: "Google will have a **strong presence** with more than **170 accepted papers**, additionally **contributing to and learning from** the broader academic **research community** via talks, posters, workshops, and tutorials. You can learn more about our work being presented in the list below (Google affiliations highlighted in bold)." Source: https://ai.googleblog.com/2021/12/google-at-neurips-2021.html

- NeurIPS, 2021. Meta: "We're excited to share that Meta AI researchers will be **presenting 83 papers** at NeurIPS 2021, including **eight as spotlights** and five as orals and one paper received an **Outstanding Paper Award**. [...] Meta AI is a **proud sponsor** of two NeurIPS affinity group workshops [...]" Source: https://ai.facebook.com/blog/-meta-ai-research-at-neurips-2021-embodied-agents-unsupervised-speech-recognition-and-more/

- CVPR, 2021. Facebook: "Facebook is **thrilled** to be a **Champion Sponsor** of CVPR 2021. This year, we will **present over 52 publications** at the conference and **participate in 28 workshops and tutorials**. We seek to advance the state-of-the-art in computer vision through fundamental and applied research in **open collaboration with CVPR's dynamic scientific community**." Source: https://research.facebook.com/events/conference-on-computer-vision-and-pattern-recognition-cvpr-2021/

- CAV, 2021. Microsoft: "Microsoft is **excited** to be a **Platinum sponsor** of the CAV 2021 [. . .] Microsoft is proud to announce three of our very own researchers have received the **CAV Award** this year." Source: https://www.microsoft.com/en-us/research/event/cav-2021

- UIST, 2021. Adobe: "There are **five Adobe co-authored papers** at UIST this year. Adobe researchers have also **contributed to the conference in many other ways**, including serving on the program committee, serving on the organizing committee, and reviewing papers." Source: https://research.adobe.com/news/adobe-research-at-uist-2021/

- POPL, 2021. Extracts from Andy Gordon's page introduction video (Microsoft): "[...] To me, **POPL feels like home**. [...] We at **MSR learn a lot from POPL and I think we made some strong contributions** [...]" Source: https://www.microsoft.com/en-us/research/video/microsoft-research-sponsors-popl-2021/

- KDD, 2021. Toloka: "Toloka is **proud** to be a **Silver sponsor** of KDD (August 14 to August 18, 2021) alongside Microsoft, IBM, Google, Yahoo, and other renowned companies. [...] Come to the event to **chat with our experts**, see our latest offerings, and find out about **career opportunities** with Toloka." Source: https://toloka.ai/events/sponsor-kdd/?from=znatoki

- CVPR, 2020. Amazon: "Amazon is **proud** to be a **Platinum sponsor** of CVPR." The webpage also reports information on papers presented and other contributions. Source: https://www.amazon.science/conferences-and-events/cvpr-2020

- SIGGRAPH, 2020. Nvidia: "At the SIGGRAPH 2020 virtual conference, we **shared the latest innovations** in research and AI for graphics." Source: https://developer.nvidia.com/events/recordings/siggraph-2020"

- SIGGRAPH, 2020. Microsoft: "As a **bronze sponsor**, Microsoft will **join academic and industrial participants** as they present research and experience papers [...]. Microsoft is also a **proud co-sponsor** for the second ACM Student Research Competition (SRC) [...] **Stop by our table to chat with our experts, see demos of our latest research and find out about career opportunities** with Microsoft." Source: https://www.microsoft.com/en-us/research/event/sosp-2019/

- PODS, 2020. IBM Research: "IBM is a **proud Platinum Sponsor** at this year's conference. [...] You can also access our **virtual booth**, where you can **learn more about our work**, as well as see and hear about our latest **technology demos, and publications**." Source: https://www.ibm.com/blogs/research/2020/06/ibm-research-sigmod/

- VLDB, 2020. Amazon: "Amazon is **proud** to be a **Gold sponsor** of VLDB." The webpage then reports information on papers presented and other contributions. Source: https://www.amazon.science/conferences-and-events/vldb-2020

- IJCAI, 2020. Sony: Dedicated webpage on the firm website for participation in the conference, with information on sponsorship status and activities. Source: https://www.sony.com/en/SonyInfo/sony_ai/ijcai2020/

- OSDI, 2020. Amazon: Dedicated webpage on the firm website for participation in the conference, with information on one accepted paper. Source: https://www.amazon.science/conferences-and-events/osdi-2020

- KDD, 2020. IBM Research AI: "We will **present several demos, talks, tutorials, and papers** that explore a wide range of topics [...] we will showcase some of the work around healthcare resulting from **collaborations with Watson Health and Cornell University**. [...] We will also be presenting a series of papers and tutorials on automated machine learning [...]. IBM Research is a **gold sponsor** of KDD 2020. We hope you will **join us at our**

virtual booth to **chat with our researchers and recruiters** about our latest research, career opportunities, internships including the AI." Source: https://www.ibm.com/blogs/research/2020/08/ibm-research-ai-kdd-2020/

- SIGIR, 2020. Amazon: "Amazon is **proud** to be a **Platinum sponsor** of SIGIR."" The webpage then reports information on papers presented and other contributions. Source: https://www.amazon.science/conferences-and-events/sigir-2020

- WSDM, 2020. Amazon: "Amazon is **proud** to be a **Gold sponsor** of WSDM 2020. Amazon's research teams are **looking forward to meeting you** at WSDM 2020. Come and **visit us at the Amazon booth**, and read on for more information about **Amazon's involvement at WSDM, academic collaboration, and career opportunities**." Source: https://www.amazon.science/conferences-and-events/wsdm-2020

- NeurIPS, 2019. Microsoft: "Microsoft is a **proud Diamond sponsor** of the 33rd annual conference [...]. **Over 300 of our researchers are involved** in **spotlight sessions**, presentations, symposiums, posters, accepted papers, and workshops." Source: https://www.microsoft.com/en-us/research/event/neurips-2019/

- NeurIPS, 2019. Google: "As a **Diamond Sponsor** of NeurIPS 2019, Google will have a **strong presence** at NeurIPS 2019 with more than **500 Googlers attending in order to contribute to, and learn from**, the broader academic research community via talks, posters, workshops, competitions and tutorials. We will be presenting work that pushes the boundaries of.. [...], with Googlers **co-authoring more than 120 accepted papers**." Source: https://ai.googleblog.com/2019/12/google-at-neurips-2019.html

- UIST, 2019. Microsoft: "As a **Bronze sponsor**, Microsoft looks forward to **making UIST an ideal opportunity to exchange research results and ideas**." Source: https://www.microsoft.com/en-us/research/event/uist-2019/

- KDD, 2019. IBM Research AI: "IBM Research has been at the forefront of knowledge discovery and data mining research for more than a quarter century and is a **proud Platinum level sponsor** [...] At KDD 2019, IBM Research AI will **present technical papers** [...]. At KDD's opening ceremony, IBM Research Staff Member Charu Aggarwal received the **SIGKDD innovation award for lifetime achievements** in data mining. Congratulations to Charu! He is also presenting three of his papers at the conference (see below)." Source: https://www.ibm.com/blogs/research/2019/08/ibm-research-ai-kdd2019/

- SIGCOMM, 2018. Ericsson: "Continuing a multi-year tradition, Ericsson again **proudly sponsors** the upcoming ACM Sigcomm conference, the most **prestigious scientific conference** in the area of communication technologies." Source: https://www.ericsson.com/en/blog/2018/7/join-us-at-acm-sigcomm-2018

- EC, 2017. Microsoft: "Microsoft researchers will have a **significant presence** at the conference, **co-authoring many papers**, serving in **leadership roles**, giving an invited talk, and receiving an **award**. [. . .] Insights from **the economics and computation (EC) community** are impacting how the government raises [...], how publishers monetize [...], how rural farmers in Uganda sell produce [...]" Source: https://www.microsoft.com/en-us/research/blog/microsoft-intelligent-markets-acm-ec17/

D Appendix: Data

In this appendix, we extend the description of our data sources and data construction following Section 3 of the paper. Table D-1 summarizes the type of information obtained from each source and gives references to the source. The relationships between the data sources are visually documented in Figure D-1.

D.1 Detailed data description

Table D-1: Data sources.

Data source	Variables
dblp	Conference, conference series information including place, time and presented papers, author disambiguation http://dblp.uni-trier.de/
CORE	Conference series quality ranking, sub-fields classification http://www.core.edu.au/conference-portal
WoS, Scopus	Affiliation information, citations, scientific classifications of articles, sponsorship information https://www.webofscience.com/wos/ https://www.scopus.com/home.uri
Gender-API.com	Predicted gender of author names
NPL database	Patent citations to CS publications: - Front-page citations from Knaus and Palzenberger (2018) - In-text citations from Verluise and De Rassenfosse (2021)
PATSTAT	Patent information, applicant names and addresses https://www.epo.org/en/searching-for-patents/business/patstat
ICAO, BTS	Direct flight connections, airport regions https://www4.icao.int/newdataplus https://www.transtats.bts.gov/DatabaseInfo.asp?QO_-VQ=EFI&Yv0x=D
Firm database - Orbis - GRID - EU Scoreboards	Firm names, ownership structure, industry information https://www.bvdinfo.com/en-us/our-products/data/international/orbis https://www.grid.ac https://ec.europa.eu/growth/industry/innovation/facts-figures/scoreboards_en

dblp has a very broad coverage of conference proceedings and the linked conference papers as well as journal articles in CS (Cavacini, 2015). It also covers books, but the number is small. dblp does not index pre-print publications from ArXiv or other online archives. Compared to other sources, dblp contains more consistent conference and conference series information. Additionally, dblp supplies a high-quality author name disambiguation (Kim, 2018). dblp has the highest coverage rate among specialized databases. The data provide an identifier for the conference series. Conference event locations and dates are not available as indepen-

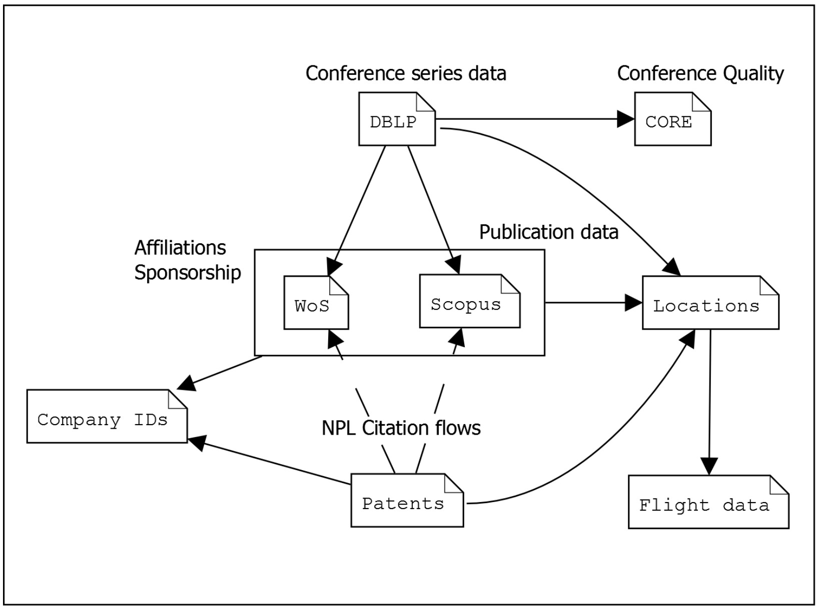


Figure D-1: Structure of the dataset

dent fields but can be easily parsed from conference volume titles. WoS and Scopus have a higher coverage rate due to the coverage of other fields, but the information in dblp is more consistent and representative for CS (Cavacini, 2015). Additional information on dblp is available at dblp.uni-trier.de. A recent discussion of the disambiguation procedures is available at blog.dblp.org/2020/01/08/corrections-in-dblp-2019/.

Other relevant bibliographic information is missing in dblp, which we obtain from Web of Science (WoS) and Scopus. Since Scopus is available to us since 1996, we focus our attention on those years. Both WoS and Scopus are widely used bibliometric databases with large coverage of different scientific fields but possibly with lower coverage of specific fields relative to specialized databases like dblp. The match between dblp and the complete WoS and Scopus is done using the DOI and the cleaned title. Matches are verified using page numbers, publication years, and author names, and only matches showing sufficient overlap are kept. Necessarily, we drop conferences and conference papers for which no match is found in WoS or Scopus. We can match up to 90% of the dblp entries with an item in WoS and/or Scopus. Combining both WoS and Scopus is necessary to maximize coverage (especially from Scopus) as well as to increase quality (affiliation information in WoS tends to be superior). Further, our patent data relies on the combination of WoS and Scopus. In the sample for descriptive and regression analysis, we will restrict to conferences with at least ten papers matched in WoS or Scopus.

We add information on conference series quality and CS research subfields from the Computing Research and Education (CORE) data curated by the Computing Research and Education Association of Australasia. The CORE data classify conference series into the quality-rank levels $A^\star$, A, B, and C (some remain unclassified or tagged as of regional interest only) and up to three subfields. We disregard CORE fields unrelated to CS or related engineering, dropping design. Unless otherwise specified, we use the first listed subfield. The CORE data constitutes an expert-based assessment of conference quality and subfields and is meant to cover a comprehensive set of all relevant conferences in CS. We match CORE to our data manually, partially supported by probabilistic string-matching algorithms. We retain conference series that match with CORE ranking information exclusively and drop conference series that are unclassified. We use the latest available version of CORE, which provides the broadest coverage. Consequently, our rank classification is time-invariant. However, by comparing different versions of CORE rankings (2008, 2010, 2013, 2014, 2017, 2018), it is evident that changes in ranks are rare and, in most cases, minimal.

Table D-2: Observation counts

	Observation counts			
	All	WoS/Scopus	With CORE	Sample
Dataset				
Conference papers	1617581	1424820	906681	559127
Conference Events	22104	20162	10583	6846
Conference Series	3660	3443	1039	982
Firms				
All Firms		9695	7089	4800
Participants		8923	6561	4418
Sponsors		2097	1380	906

Notes: Observation counts for different matching steps. Fourth column is the estimation sample. Third column from the right is relevant for the descriptive part. First column: All dblp items. Second column: dblp items found in WoS or Scopus. Third column: Also restricting to conference series matched with CORE. Last column: Restricting to the final sample, especially to 1996-2010.

Table D-2 provides an overview of the number of observations in our data. Merging dblp with WoS/Scopus and CORE inevitably reduces the number of available observations. The match with CORE data leads to a more substantial drop in the number of unique conference series and conference events originally covered. However, we verified that these are largely small and less relevant conferences, with few corresponding papers each. We still retain 906681 conference papers, corresponding to 70% of the initial total (the number of papers in dblp matched with WoS or Scopus). In the final data underlying both descriptive and regression analysis, we restrict to the years 1996-2010, to papers matched to CORE and WoS/Scopus, and to conference events with at least ten papers satisfying these criteria.

Most importantly, as noted in the paper, our data cover 75% of all conference series listed in CORE. Eighty percent of conference series listed in CORE and not in our data are of the lowest quality rank, C. This implies that the data covers almost the entirety of top and medium-ranked conferences in CORE. In general, the sample is biased against small conference events, short-lived conference series, and conference series of the lowest quality that are less likely covered in a generic bibliographic database such as WoS or Scopus and are less likely ranked in CORE.

We can claim that the data are largely representative of all relevant conference events in CS in our period of observation. Table D-2 also shows the difference between our estimation sample with years 1996-2010 and our full sample 1996-2015. Citation-based variables require time windows in which the citations can be observed. We choose five-year windows. This truncation issue forces us to limit to 1996-2010 the sample for econometric analyses. The full dataset (up to 2015) consists of a total of 10,583 conference events in the 1996-2015 period – pertaining to 1,039 conference series and more than one million papers. A total of 7,089 firms have participated in at least one conference event, either authoring at least one conference paper or sponsoring a conference event. The sample up to 2010 comprises instead 4,800 firms, 6,846 conference events pertaining to 982 conference series, and a total of 559,127 of conference papers. In the remainder of the paper, we also present descriptive statistics limited to this sample.

Scientist biographies For the collaboration and hiring variables and for self-citations and scientist counts, we rely on disambiguated scientist profiles and affiliation information combined with firm information. The scientist profiles are taken from dblp, but we rely on data from WOS and Scopus for the affiliation information. While for the majority of the data, only a paper-firm or patent-firm link is relevant, here, a paper-person-firm link is required. To achieve this, we, at the paper level, match individual authors from dblp to individual authors and corresponding affiliations in Scopus and, if unavailable, in WoS.

With this, we establish a person-year panel and compute the fractional association of individual scientists with firms or academia. Whenever a scientist is associated with a firm on a journal article or conference paper, that information is taken into account. If, in a given year, a scientist features different affiliations from one or several papers, fractional counts are used. In years where the scientist did not publish, linear interpolations from years before and after are used for the variable on firm size of research investments.

There is a small share of cases where the individual information of affiliation cannot be retrieved. A small part of this issue is due to missing affiliation information. The rest comes from a limitation of WoS, which does not provide a direct link between the author list and affiliation list; they are simply listed uniquely in their order of appearance. For this reason, when available, we prefer information from Scopus, which is essentially complete. We also mitigate this issue as far as possible in WoS: the first affiliation can always be assigned to the first person, or papers with only one affiliation can be assigned completely. Still, some cases remain where the information is missing.

Patent data Since no single data source links dblp to non-patent literature references in patents, we need to use secondary datasets as a detour. Primarily, we use the match between dblp, WoS, and Scopus, combined with the data by Knaus and Palzenberger (2018). These data cover information about non-patent literature (NPL) references from USPTO and EPO patent (applications) as well as WIPO publications towards all publications covered by WoS and Scopus during the period of interest for us (Around 80% of the citations stem from the US system, and EPO/WIPO are contained with roughly similar proportion. The US portion comprises almost exclusively grants, whereas, among references from the EPO, the grant rate was around 40%. The major patent offices conduct search reports for WIPO applications, most frequently the EPO). The NPL references were originally taken from DOCDB (in 2017), which maps to PATSTAT. For all supplementary patent information (priority years, applicant information, patent families) we rely on PATSTAT. However, this data only covers front-page citations. To rectify this, we supplement this data with in-text citations derived from PatCit (Verluise and De Rassenfosse, 2021), which we directly link to dblp via DOIs. These are citations made in the technological descriptions of US patents. We aggregate all patents at the family-level (docdb families) and use the earliest priority year within each family as the point in time the knowledge embedded in the patent family originated. When multiple publications with associated applicant affiliation information are available within a family, we prefer the first publication over grant publication over subsequent publications. For the firm match, we utilize the person_name field from PATSTAT for entities classified as a company, but in order to maximize coverage we expand the matches that were found based on the HAN_ID (and thus HAN_NAME) field.

Text similarity We calculate text similarity scores between conference papers using their abstract and title. Text similarity is a cosine similarity and so conceptually bounded between $[-1, 1]$ but in practice almost always on the interval $[0, 1]$. We calculate text similarity scores using the cosine similarity between reduced term frequency–inverse document frequency (tf-idf) values of the cleaned abstracts and titles. In the first step, we clean the abstract.[2] Of the so-cleaned abstract, we take the 50,000 most frequent tf-idf values of one, two, and three-grams. We exclude very frequent terms. We then use a truncated singular value decomposition (SVD) to reduce the dimensionality from 50,000 to 300. This approach is also called latent semantic analysis (LSA). The latter name hints at the purpose - finding dimensions that concisely describe the semantic content of an abstract. Multiple words can have the same meaning and the same word can have several meanings, depending on the context. All in all, this approach generates a procedure that maps an abstract into 300 dimensions. For the tf-idf measure as well as the SVD, it is necessary to take the full body of documents into account in a training stage. For this, we use all 2.6 million dblp items for which we can find abstracts. Once this training stage is completed, individual abstracts can be analyzed. Finally, the cosine similarity is calculated for pairs of transformed abstracts.

D.2 Firms

Matching firms We generate a list of firm entities that we use as candidates. Our goal is to provide global coverage of the - probably - most important firms that conduct scientific publishing. For this reason, we use the firm names from the EU scoreboards as well as the firms included in GRID. The Scoreboard lists the top companies worldwide by R&D expenditure. In the first year, 2003, the list contains 500 EU and 500 global companies. Over time, the length of the lists increased, so the 2017 Scoreboard lists the top 1,000 EU firms and the top 2,500 worldwide. The 2017 Scoreboard is the last included in our data. All in all, this adds

[2]Cleaning involves concatenating title and actual abstract, removing copyright statements, and replacing special keywords with character strings (2D becomes twod, L2 becomes eltwo, ...). Then, everything which is not a character is replaced with a whitespace. We employ stemming, which reduces flexed forms of words to their stem. We also remove stop words.

roughly 8,300 distinct firm name strings, of which often several refer to the same firm entity. For GRID, we use a snapshot from May 2018. GRID, as a curated dataset of research-active entities, is a prime candidate for adding firms likely involved in scientific activities. We only add entities labeled as companies, which adds another roughly 21,000 match candidates. We further wanted to complement this list with firms that possibly use information from conferences in their technological activities but do not publish frequently enough to be included in the curated GRID list. Therefore, we add all firm names for firms that in Orbis (2016 version) were found to be connected to at least one patent. Also, we added firms from the US and DE section of Orbis to try to capture smaller firms this way. From those, we drop entries with limited financial information (no financial data, with consolidation code 'LF,' or with only one year of financial data) and such that are fully dependent subsidiaries (e.g., more than 75% of ownership stake by another entity). The entries from the US/DE greatly expand the set of candidate firms while adding mostly irrelevant candidates, so we did not further expand to additional countries to retain computational feasibility.

Matching bibliometric information to firms is particularly hard, as little additional information exists besides the affiliation string or the sponsor name. Location information is often unavailable, and when it is available, it often refers not to the headquarters location but to the particular research lab. Therefore, we try to enrich the affiliation name with contextual knowledge from the Internet, following the approach by Autor et al. (2020). We search for the affiliation string in a search engine and retain the first ten results. We disregard very frequent occurrences where, for example, many firms are listed on a single website. We also use frequency weighting to put a higher weight on less common entries.

The match uses the software package Dedupe (https://github.com/dedupeio/dedupe/). Dedupe provides a probabilistic algorithm that, based on manually crafted training data, calculates weights for different input features. These input features are the web search-based similarities but also traditional string similarity measures. Dedupe also calculates a minimum similarity threshold for which matches are kept. This is done based on a comparison of precision and recall scores. The matching step returns for each affiliation string a set of candidate firm strings to which this affiliation string might belong.

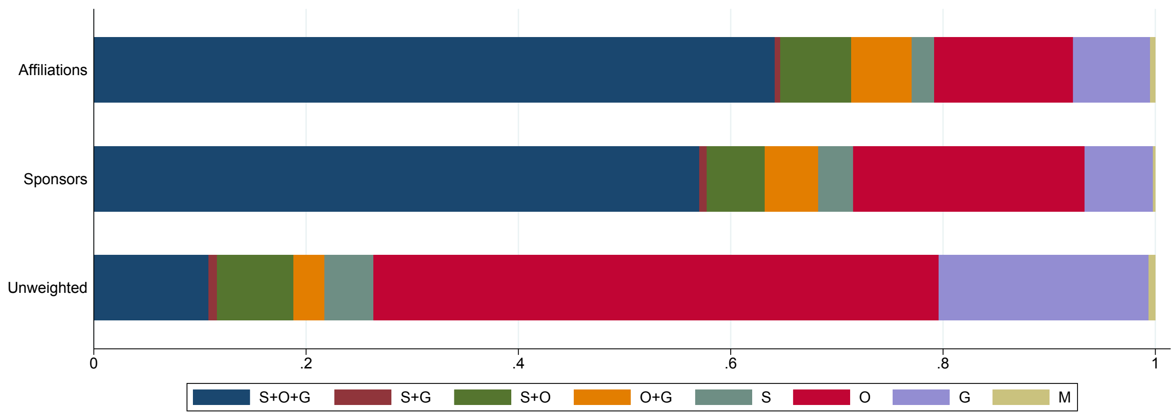
In the next step, we cluster the n:m match provided by Dedupe to group firm strings that belong to the same entity. In GRID, Orbis, and the EU Scoreboards, several possibilities for writing of the same firm name are possible. Additionally, firms may have been renamed, merged, acquired, etc. Incidentally, the web search-based algorithm is by itself quite good at picking up these name changes. However, this step required much manual refinement. Whenever multiple entity names were grouped, we validated these choices. When in doubt, the clustering implicitly and our validation explicitly clustered entities in larger groups. So, if two firms merged during a part of the sample time frame, we consider them to be the same entity for our full sample. Also, when the matching algorithm could not confidently distinguish subgroups of conglomerates, we grouped them into one entity. This happens with firms like Samsung or LG. The firm clusters yield our firm entities for this study.

Composition of the data Figure D-2 shows the sources of firm observations in the conference dataset. These give an overview of successful matches and the extent of clustering. The number of firms in each category is weighted by the number of papers authored (blue) and the number of conferences visited (red). Also, an unweighted count is provided. Most individual matches are from Orbis only, followed by GRID. However, the most important firms can be found in all three databases ("Orbis+Scoreboard+GRID").

We apply our match algorithm for a variety of data sources. At the core of this study is the match between affiliations for conference participants and sponsorship information found in WoS and Scopus. Further, we use the same matching strategy to match firm applicants from patents citing computer science papers or otherwise relevant for computer science (the technology main area 'Electrical Engineering'). Due to this broader match target set, there can be several firms that are matched to some affiliation or applicant string but never occur in the CS data.

We comment on firm locations in the next section.

Figure D-2: Sources of firm data



Notes: Shows the data sources of firms in the dataset. Firm data is taken from the EU Scoreboards (S), Orbis (O) as well as GRID (G). A small number of entries was added manually while clustering and cleaning (M). The number of firms in each category is weighted by the number of papers authored by the firm (top) and the number of conferences sponsored (middle). An unweighted count is provided as well (bottom).

D.3 Geolocalization and flights data

Airports and direct flights For the data on airport direct flights, we combine ICAO and BTS data. The BTS data is the Airline Origin and Destination Survey, a 10% sample of airline tickets, including origin, destination, and other details. Based on the DB1BMarket module, we construct passenger flows using direct flights between airports in the United States. If a ticket contains layovers, we count each leg of the journey as a direct flight connection instead of the (ultimate) origin-destination flow. The ICAO data is the Traffic by Flight Stage module of the ICAO. From this data, we derive the number of passengers from individual flight stages (the legs of a journey) of international connections. For connecting domestic flights, the ICAO data sometimes contains passenger information as well, but not systematically. Therefore, we disregard all ICAO information on domestic flights. We connect the BTS and ICAO data by assigning the airports given in each individual data to each other. On a pair-level of airports, the data are mutually exclusive. For calculating the number of passengers an airport serves, we add the pairwise connections given by BTS and ICAO, i.e., we sum up incoming and outgoing passengers on domestic and international flights.

The unavailability of domestic flights in the ICAO data is a concern. Especially for the US, domestic direct flight connections play an important role in the attendance of international conferences, which is why we added the complementary BTS dataset. In the estimation dataset – excluding cases where the airports of scientists and conferences coincide - about 4% of observations relate to non-US domestic airport pairs, and more than 13% refer to US domestic airport pairs. Outside of the US, attendance at international conferences is much less likely to be domestic, and for domestic transportation, flights are much less important, as alternative transportation is available. Still, including non-US domestic flights would be desirable, but unfortunately, no single dataset on domestic flights outside the US is available.

Geolocalization and airport assignment We assign geolocalized locations to airport regions based on proximity and airport activity. The assignment procedure is intended to be conservative in the sense of increasing the likelihood of finding direct flights when they were reasonably available while assigning a single airport for fixed effects. In the first step, for each geolocated information (conference venue, affiliation location, etc.), we select the five closest airports with at least one passenger in the combined ICAO/BTS data. If more than one airport is closer than 250 km, we assign the airport with the highest passenger volume in the relevant year for the affiliation, which is likely to feature the most direct flight connections and is most likely used by scientists. For example, consider an affiliation string of a company located in Menlo Park, CA, USA, in 2016. The closest airport in our data would be San Carlos Airport, which is a reliever airport for San Francisco International Airport and does not feature any scheduled connections in our data. Therefore, we exclude this airport from consideration. Among all airports with passenger volume, the closest one was San Jose, CA (around 30km, 620,000 passengers), but San Francisco International Airport (around 40km, 6.8 million passengers) is the airport with the highest traffic volume. For example, for conference venues, this procedure selects the closest airport in 71% of cases, in 95% when disregarding small airports.

For papers with multiple authors, we focus on the affiliations of the first author. Due to norms in CS, the first author is the one most likely to present the paper. On the practical side, the amount of choices between alternative regions is reduced. For papers with multiple geolocated information, by default, we consider the airport with the overall highest traffic volume. On average, the distance between WoS/Scopus first author affiliation locations and the selected airport is 31 km (median 11 km). In some descriptive analyses, we use equal weighting of all airport locations associated with a paper. In robustness checks, we use the airport assigned to all (author) affiliations and find similar results.

Firm locations We derive firm locations from various sources, preferring bibliometric data (WoS, Scopus) over firm-level information (Orbis, GRID). We geocode firm locations and assign them to the nearest airport region. If an analysis requires a unique (main) location, we focus on the airport location with most associated conference papers and journal publications. For 2.4% of firms that participate in at least one conference, geographic assignment fails. In many cases, these are firms that only sponsor, so the originating data contains no geographic information.

D.4 Detailed variable definitions

In all definitions, we employ the notation of firm f which participated in conference c and was exposed to paper p. Conference c is paired with conferences c', where paper p' was presented. The variables typically characterize relations between f and p or p', respectively.

Patent citations is a binary variable equal to 1 if paper p (p') presented at conference c (c'), and authored by a researcher external to firm f, will be cited by a patent applied for by firm f within five years from c. To accommodate the 5-year window, the descriptive and estimation sample includes years up to 2010. This may still be insufficient: citations are added to patent families over time in subsequent publications of the patent (e.g., grant publication and international filings). Because lags of several years are possible, many citations may remain unobserved. We ensured this did not affect our analyses by considering a sample up to 2008, obtaining equivalent results. Patent citations include both front-page and in-text non-patent literature citations linked to CS publications. In robustness checks, we exclude in-text citations derived from a separate database.

Science citations is a binary variable equal to 1 if paper p (p') presented at conference c (c'), and authored by a researcher external to firm f, will be cited by a CS publication by firm f within five years from c. We include follow-on citations from any CS publication covered by dblp. In robustness checks, we limit to citations from only papers or only articles.

Patent cit to past, Science cit to past We capture follow-on citations from patents or scientific publications of the firm in the five years after the conference to previous scientific work of the authors of a paper p. The variables measure whether any such citation was made towards previous CS publications by any of the authors of paper p published in the five years before the conference. Formally, this is a binary variable equal to 1 if any CS publication p^- published up to five years before the conference by authors of paper p (p' - authors external to firm f) will be cited by a patent (CS publication) by firm f within five years from c.

Patent cit to past$_{pre-conf}$, Science cit to past$_{pre-conf}$ We capture the pre-existing engagement of the firm with the scientific work of the authors of a paper, either in the domain of science or technology. For this, we construct a variable capturing whether any citation exists from firm patents (science, respectively) published in the five years before the conference (*pre-conference*) towards previous CS publications by any of the authors of paper p published in the five years before the conference. Formally, this is a binary variable equal to 1 if any CS publication p^- published up to five years prior to the conference by authors of paper p (p' - authors external to firm f) was previously cited by a patent (CS publication) by firm f within five years prior to c. These variables are conceptually related to "Patent/Science cit to past" but focus on a different set of *citing* publications.

Collaboration We describe whether follow-on collaborations occurred between any scientist in the author team in a paper and a firm. A collaboration occurs if we observe any CS publication, article, paper, or otherwise covered in dblp, with the scientist as the author and any affiliation assigned to the firm. The scientist may become affiliated with the firm - which resembles the 'Hiring' variable - or join another firm scientist in coauthoring a publication. Collaborations also rely on the construction of scientist biographies as described above. Notably, the scientists' biographies do not consider information from patents, and we thus exclude collaboration on technology development.

Hiring We consider hiring to have occurred when a scientist who, at the time of the conference, was not affiliated with a firm subsequently became affiliated with that firm. Note that since we only track scientist biographies via publications, our hiring variable is closely related to the 'Collaboration' variable but restricted to cases where the affiliation of the focal scientist also switches.

Firm in science hub We consider a firm to have access to a hub if a significant proportion of scientists affiliated with the firm are located in a region with world-leading CS output. First, we calculate the scientific strength of an airport region by aggregating the number of papers assigned to this region by year and field. For each field, we annually identify the top ten research-active regions globally. We assume that firms have access to a region if at least 10% of the scientific output associated with this firm is produced by scientists located in that airport region. If no information about the regions of firm scientists is available, which is the case for companies that only sponsor or where geographic information is missing on bibliometric entries, we use the headquarters location. Results in the paper are presented for field-specific hubs, but results are robust to using hubs according to a definition based on the overall rather than field-specific scientific strength of regions.

Research similarity We measure pre-existing research similarity using the text-similarity of the focal conference paper p to the firms' papers in the year before the conference in the same CS sub-field. Appendix D.1 provides a detailed description of text similarity. If a firm had multiple papers, we take the average similarity. If a firm had no paper, we set the similarity to zero. Results remain robust to an alternative variable indicating whether the firm had any conference papers in the same field in the prior year.

Paper team female In the main analysis, we calculate the gender of the team using the share of female authors in the team. Results are robust to using the gender of the first author. Gender information is not available in publications data and is impossible to collect on a large scale. For the gender of individual authors, we rely on the gender prediction of their first names using services from Gender-API.com. We have searched for the predicted gender for the roughly 2 million unique author names in dblp, finding information for 87% of the sample. For each name, we retain the most likely gender. For authors with multiple first names, we focus on the first, ignoring initials.

Paper team productivity For this variable, we consider productivity in the team as the maximum number of $A^\star$ publications within the last five years between the authors. Publications at $A^\star$ conferences are considered the most prestigious, so the variable also captures aspects of scientific prestige and visibility.

Paper by other firm Firms may learn from the work of academic scientists but also benefit from knowledge spillovers from competitors. To capture this distinction, we introduce a variable of whether a conference paper was authored by at least one author affiliated with a firm.

Conference size We proxy the size of a conference by the log number of papers at a conference. We expect this number to correlate well with the total number of participants, although, as discussed in Sections 3.1 and 3.3, we cannot detect the number of scientists who physically attended the conference in our data.

Firm scientists n. We proxy the overall research investment of firms by the number of scientists they employ. For this, we rely on author biographies constructed according to Section D.1. In short, author biographies are based on all CS publications and impute between years of publication.

Geo. distance to paper We measure the distance between the firm and a conference paper by the distance in km from the nearest R&D location the firm had in the prior year. We transform distance using the $log(1 + x)$ transformation. For R&D locations of a firm, we consider all airport regions where at least 5% of the firm's science or at least 20 papers of that year were produced. If neither was true of any location, we consider the most frequent location. If the firm did not produce any science in a given year, we consider the firm's time-invariant main location. Results using the distance from the firm's time-invariant main location are comparable.

Geo. distance from paper to conf., Same geographic region as conference, Same airport as conference The distance between the region of a conference paper's authors and the conference venue, in km, is an indicator as to whether the two are with the same country or US state, as well as an indicator of whether they are the same.

Paper author attended paired conf. This variable is an indicator as to whether any author of conference paper p from the actual or the matched conference was also an author of a paper presented at another actual or matched conference paired with the conference of p.

IV: All authors The alternative instruments follow the same logic as the default one and instrument the participation of paper authors and firms to the same conference. The default instrument considers the locations of the first author of conference papers for direct flight connections to the conference. Alternatively, we construct an instrument focusing only on the location of all authors. In this case, we select the busiest airport located in the vicinity of any author. In this logic, if another author is located near a busy airport, the probability of a direct flight, and consequently them attending, is higher, too. For comparison, see Section D.3. Region-fixed effects are then based on the region of the author whose airport was selected.

Appliedness We construct appliedness indicators at the article level and the conference level. Our article-level appliedness indicator relies on the title and abstract text and implements the trained models from Boyack et al. (2014). They provide pre-trained multinomial logit models that sort each article into four levels of appliedness, from most (1) to least (4). For our article-level appliedness indicator, we use the sum of the probability of falling into category 1 or 2. The conference-level indicator follows the logic of the Journal Commercial Impact Factor (JCIF) of Bikard and Marx (2020). For this indicator, we rely on our patent-paper citation links. For a given conference and year, we calculate the number of patent families with priority in that year that cite papers from the same conference series published in the prior two years, divided by the number of such articles. Due to the nature of this indicator, it is highly correlated with conference series fixed effects. We report results in Table A-7.

References

Autor, D., D. Dorn, G. H. Hanson, G. Pisano, and P. Shu (2020). "Foreign Competition and Domestic Innovation: Evidence from US Patents." In: *American Economic Review: Insights* 2.3, pp. 357–374.

Bikard, M. and M. Marx (2020). "Bridging Academia and Industry: How Geographic Hubs Connect University Science and Corporate Technology." In: *Management Science* 66.8, pp. 3425–3443.

Boyack, K. W., M. Patek, L. H. Ungar, P. Yoon, and R. Klavans (2014). "Classification of individual articles from all of science by research level." In: *Journal of Informetrics* 8.1, pp. 1–12.

Cavacini, A. (2015). "What is the best database for computer science journal articles?" In: *Scientometrics* 102.3, pp. 2059–2071.

Kim, J. (2018). "Evaluating author name disambiguation for digital libraries : a case of DBLP." In: *Scientometrics*.

Knaus, J. and M. Palzenberger (2018). "PARMA. A Full Text Search Based Method for Matching Non-Patent Literature Citations with Scientiic Reference Databases. A Pilot Study." In: *Technical Report by the Max Planck Digital Library, Big Data Analytics Group*.

Verluise, C. and G. De Rassenfosse (2021). *patCit*. https://github.com/cverluise/PatCit.