Paper to be presented at
the DRUID16 20th Anniversary Conference
Copenhagen, June 13-15, 2016

At the origins of learning: Absorbing knowledge flows from within or outside the team?

Charles Ayoubi
EPFL
CDM
charles.ayoubi@epfl.ch

Charles Ayoubi
EPFL
CDM
charles.ayoubi@epfl.ch

Michele Pezzoni
University of Nice
GREDEG
michele.pezzoni@unibocconi.it

Abstract

Empirical studies document an increasing trend towards more collaboration in the scientific community and a positive effect of collaboration on team productivity. The most common explanation to these productivity benefits is that teamwork stimulates individuals' learning through knowledge sharing. However, little has been done to assess how knowledge flows between team members. We use a novel dataset based on a Swiss funding program promoting team collaboration to (1) define a team and (2) to measure knowledge flows within its boundaries. We find that the probability of the focal individual learning is correlated with individual characteristics and with team characteristics. Moreover, team characteristics play a role in determining the origin of learning, namely if team members share their knowledge or not. Specifically, we find a negative effect of the social distance of team members on their probability of learning from within the team. We find no effect of the geographical distance. Finally, we find an inverted U-shape effect of the cognitive distance of the team members on the probability of learning from within the team.

Jelcodes:O30,-

At the origins of learning: Absorbing knowledge flows from within the team

Charles Ayoubi
École Polytechnique Fédérale de Lausanne, Chair in Economics and Management of Innovation
ayoubi.charles@epfl.ch

Michele Pezzoni
University of Nice and GREDEG; Bocconi University, CRIOS
michele.pezzoni@unice.fr

Fabiana Visentin*
École Polytechnique Fédérale de Lausanne, Chair in Economics and Management of Innovation
fabiana.visentin@epfl.ch

Abstract: Empirical studies document a positive effect of collaboration on team productivity. However, little has been done to assess how knowledge flows among team members. Our study addresses this issue by exploring uniquely rich data on a Swiss funding program promoting research team collaboration. We find that being involved in an established collaboration, the presence of a female scientist, and the size of the team foster the probability of an individual learning from the other team members. We find also that young team members and team members with a limited experience are more likely to learn from senior and more experienced peers. Moreover, there is an inverted U-shaped effect of the cognitive distance on the probability of learning.

Keywords: research team; learning process; knowledge flows; cognitive distance; social distance; geographical distance.

Acknowledgments
We are indebted to Dominique Foray, Francesco Lissoni and Jacques Mairesse for their invaluable advice. We also thank Stefano Bianchini, Michele Cincera, Paul David, Cesar A. Hidalgo, Patrick Llerena, Marco Guerzoni, Brian Uzzi, Reinhilde Veugelers, Bruce Weinberg, and the seminar participants at the University of Bordeaux, University of Strasbourg, International Symposium on Science of Science 2016, GREDEG, the workshop on Complex Evolving System Approach in Economics at the University of Nice Sophia Antipolis, the Leuven University and the 10th Workshop on The Organization, Economics and Policy of Scientific Research, BRICK, Collegio Carlo Alberto. We also thank the SNSF for the continuous support.

***Corresponding author:** Fabiana Visentin, Chair of Economics and Management of Innovation, École Polytechnique Fédérale de Lausanne, ODY 1.18 Station 5, Lausanne 1015, Switzerland. Email: fabiana.visentin@epfl.ch

1. Introduction

This paper assesses the characteristics of a research team that foster the probability of learning from teammates. We add to the Science of Team Science (SciTS) literature by investigating research teams under the point of view of team members' learning, an aspect often neglected in favor of the team productivity analysis.

Over the past century, the process of scientific knowledge production has fundamentally changed. Nowadays the teamwork model of conducting science has mainly replaced the single scientist model (Jones et al., 2008; Wuchty et al., 2007). Several reasons explain this trend. First, the cost of scientific instrumentation leads scientists to organize in teams in order to share resources and to avoid duplication of costs. Second, lower travel and communication costs increase scientists' mobility and favor the creation of multi-institution teams. Third, certain fields such as physics, chemistry, engineering, and biology are characterized by an increasing level of complexity that requires the joint effort of specialized scientists. It becomes implausible for a single individual to master all the technical skills and knowledge needed to set up a laboratory, run an experiment, analyze the data and manage the publication process.

There is a general consensus among scholars that "collaboration [outcome] is greater than the sum of its parts" (Katz and Martin, 1997). Even if some authors also expose some drawbacks to collaboration, such as higher coordination costs (Bikard et al., 2015; Mowatt et al., 2002), or ghost and honorary authorship (Mowatt et al., 2002), most empirical studies agree to say that collaboration has a positive impact on publication productivity. Not only teamwork has a greater value than solo-author work but teamwork positively affects the productivity of each team member (Defazio et al., 2009; Lee and Bozeman, 2005). The most common explanation of the greater value and higher productivity of teamwork is that it allows scientists to combine their knowledge prompting scientific discoveries (Uzzi et al., 2013).

These studies investigating the dynamics of scientific teams are the building blocks of the SciTS literature (Börner et al., 2010; Whitfield, 2008; Stokols et al., 2008). As described by Börner et al. (2010), SciTS is "an emerging area of research centered on examination of the processes by which scientific teams organize, communicate and conduct research". Our study aims at shedding light on a process often neglected in this literature, the exchange of knowledge among the members of a scientific research team and their ability to learn from one another.

While the process of learning has been investigated within the organizational literature starting from the 80's with the work of Levitt and March (1988), this process remains largely unexplored in the SciTS literature. The organizational literature claims that learning new skills and using them within the firm is critical to the innovation and productivity of the firm (Argote and Miron-Spektor, 2011). In a similar fashion, we believe that, for scientists, acquiring new knowledge and exploiting it in their research work is key to their productivity and for the novelty of their contribution. Precisely, it has been shown that broadening the researcher horizon and exchanging knowledge in the frame of interdisciplinary teams is key to the production of innovation and high impact scientific research (Börner et al. 2010). More recently, a study by Misra et al. (2015) suggests that scientists who are more open to other disciplines are more prone to produce higher-quality research. The process of acquiring new diverse knowledge is therefore central for the effectiveness of a scientific team.

We identify the factors that promote the learning of an individual from her teammates. We claim that this portion of learning can be affected by team characteristics. As team characteristics we consider the quality of the research project the team members are working on, the size of the team, and the discipline. We also distinguish co-ethnic teams from multi-ethnic teams, and teams with at least one female scientist from only men teams. Moreover, we consider as determinant of learning the distance between the individual and her team members along three dimensions, namely, geographical distance, social distance and cognitive distance (Agrawal et al., 2008, 2003; Bercovitz and Feldman 2011; Jaffe et al., 1993; Nooteboom et al., 2007).

In our analysis, we use a novel dataset of 231 grant applications to a Swiss funding program promoting team collaboration. This rich dataset allows us to clearly define a team, its boundaries and its creation date. We then use the bibliographical references of each scientist in the dataset to precisely define her knowledge stock and specify her learning from the other team members.

We find that learning from team members is more likely when there is at least one female scientist in the team and for larger teams. Young team members and team members with a limited stock of knowledge are more likely to learn from senior and more experienced team members. Also, having an already established collaboration is correlated with a higher probability of learning from the rest of the team. We find an inverted U-shape impact of the cognitive distance between the scientist and her team on the probability that learning originates from within the team. An individual with a knowledge stock that differs from the one of the others guarantees a buffer for learning something new. At the same time, the difference of the knowledge stocks should not be too large to avoid obstacles in effective communication between team members, that is the situation when team members ‘speak different languages’ and do not understand each other.

The rest of this paper is organized as follows: Section 2 illustrates the individual learning determinants, Section 3 describes the data and the variables, Section 4 describes the estimation strategy, Section 5 provides the results, and Section 6 concludes.

2. Individual learning determinants

As defined by Huber (1991), "learning consists in knowledge acquired by any unit of an organization and available for acting upon". Applying this definition of learning in our empirical setting, we proceed in two steps. We first assess the knowledge accumulated by a scientist before entering a team based on the literature she relied on in her work. Then we consider as learning any increment to this initial stock of knowledge. We finally focus on the part of this learning that originates from within the scientific team, i.e. the knowledge that is transmitted by her teammates.

To identify the determinants of this learning originating from within the team, we rely on the more extensive literature of SciTS on team productivity. In the current section, we discuss the factors influencing the probability of learning from other team members for an individual scientist. Specifically, we consider how the probability of an individual to learn from other team members is affected, on the one hand, by the characteristics of the team and, on the other hand, by the individual characteristics of the scientist in comparison to the rest of her team.

2.1 Team characteristics

As team characteristics, we consider: being endowed with research funds, working on a high quality project, having various team sizes, having a different ethnic and gender composition.

Creation of collaborative relationships might be facilitated in teams whose projects have been awarded due to the availability of funds for travelling, team building activities, workshops and meetings. Moreover, funds might be used also to buy research equipment and materials shared among team members. Collaboration activities favored by the availability of funds are expected to foster knowledge flows among team members and consequently individual learning.

We distinguish teams with high quality research projects from teams with lower quality research projects. High quality research projects and the promise of making breakthrough scientific discoveries might stimulate scientists' commitment in working actively together. This might foster knowledge flows and learning among team members.

We expect team size to be positively correlated with the probability of learning from other team members: a greater number of individual to interact with should increase the probability of learning.

Several works have investigated the effect of researchers' co-ethnicity on the probability of knowledge flows (Agrawal et al., 2008, 2003; Freeman and Huang, 2015). The prevalent result in literature is a positive effect of the co-ethnicity of researchers on the probability of observing a knowledge flow. Following the same line of reasoning, we expect co-ethnic teams to favor knowledge flows among team members and consequently individual learning.

Gender literature shows mixed results concerning the effect of the gender composition of the team on its productivity (Apesteguia et al., 2012; Woolley et al., 2010). Moreover, extant studies are rarely able to identify a causal relationship and provide tentative explanations to the empirical evidence. The mixed empirical results and the lack of testable explanations do not allow us to make any predictions about the impact of gender composition of the team on the team members' learning probabilities. However, relying on the concept of "homophily", we might expect that team members of the same gender might be more likely to benefit from reciprocal knowledge flows and learning (McPherson et al., 2001; Cummings and Kiesler, 2008).

2.2 Scientist vs. Rest of the team

In order to compare the scientist to the rest of her team, we consider how different she is from her teammates in terms of: geographical localization, social attributes and cognitive diversity (knowledge stock composition). We categorize these three dimensions into three types of distance separating the scientist from the other members of team: geographical distance, social distance and cognitive distance.

Over the last thirty years or so, in parallel to the increase in the average team size (Wutchy et al. 2007), we witness an even higher increase in the geographical dispersion of the team members (Jones et al. 2008). Adams et al. (2005) show that the average geographical distance of the collaborations more than doubled in the last twenty years, due to improvements in transports and telecommunications. In a sample of French scientists, Mairesse and Turner (2005) show that, except for immediate proximity (i.e. being affiliated to the same unity), geographical distance has no significant impact on collaboration. According to this evidence, we expect a limited effect of geographical distance among team members on the probability of learning from each other.

In our study we consider four variables measuring the social distance of the individual from her team that might affect her probability of learning: age, reputation, gender and previous collaborations. First, we expect the age difference between an individual and her teammates to be positively correlated with her attitude to engage in knowledge transmission activities. Thus, in a mentor-protégé relationship, the young team member is expected to learn, i.e. receive knowledge, from the senior team member who is expected to transmit knowledge. However, Zenger and Lawrence (1989) find that, in a firm environment, individuals with similar age tend to exchange information easier. These two competing effects prevent us from formulating a prediction on the age difference effects.

Second, we consider the scientific reputation of team members as proxied by their publication productivity before the team formation. We identify two possible mechanisms at play within the team. On one side, highly productive members might contribute to the team with larger knowledge stocks and enhance the probability of learning of less productive team members. On the other side, highly productive scientists might focus on knowledge exchanges with teammates with similar publication stocks from which they can benefit more, and thus decide to isolate low productive scientists from which they benefit less. As in the case of age differences, we have two competing hypotheses about the possible effect of scientific team reputation diversity on learning from within the team.

Third, based on the findings on "homiphily" (McPherson et al., 2001) we expect that a scientist working in a team where at least one member is of the same gender as her will have a higher probability of learning from their teammates.

Finally, the mechanisms that affect the learning of the scientist from her team might differ whether the individual has long lasting collaborations with her teammates or not. On the one hand, scientists with previous collaboration with their team, have a greater level of familiarity (Cummings and Kiesler, 2008; Bercovitz and Feldman, 2011), and thus benefit from the presence of routinized collaboration activities that facilitate the creation of strong relational ties (Porac et al. 2004). Strong relational ties, foster knowledge flows among team members and enhance their probability to learn from each other (Granovetter, 1973). On the other hand, having previous professional collaborations might increase the probability that team members share the same knowledge stock. A redundant knowledge decreases the probability that individuals learn from each other (Burt, 2004). The contrasting effects of the mechanisms at work when the scientist has an established collaboration prevent us from making predictions of the impact on her probability of learning.

In the management literature a major determinant of the knowledge flows and innovative performance of the team is the cognitive distance separating its members (Knoben and Oerlemans, 2006; Nooteboom et al., 2007). In the SciTS literature, the question of the disciplinary diversity of the scientific team is a central issue (Fiore, 2008). The tenants of SciTS discuss the impact of the various types of cross-disciplinary teams on the effectiveness of the collaboration (Stokols et al. 2008, Börner et al. 2010). In order to estimate the level of disciplinary diversity in the team, we use a proxy of cognitive distance between team members measuring the distance separating their knowledge stocks before the team formation. According to Nooteboom et al. (2007), the cognitive distance between the team members has two competing effects on the knowledge production capacity of a team in an organization. On the one hand, the capacity of absorbing new knowledge is higher when the cognitive distance between the members is low since it is easier for the scientist to absorb knowledge similar to the one she already has. Hence, the "*speaking the same language*" effect enhances knowledge flows within the team when the cognitive distance is low. On the other hand, having a low cognitive distance between individuals implies that their knowledge stocks are very similar and thus the probabilities of observing a knowledge flow from other team members are low because they have too little novelty to offer to the scientist. Thus, the so called “*opening new horizons*” effect has a positive impact on knowledge flows within the team as the cognitive distance increases. Therefore, combining the two effects, we expect the global impact of cognitive distance to have an inverted U-shape on knowledge flows within the team. For low cognitive distances, even if the absorptive capacity is very high, the low diversity of knowledge in the team implies that the knowledge flows within the team remain very limited. The probability of having knowledge flows within the team increases when cognitive distance increases until some optimal point. Then, too high cognitive distance blocks the understanding between the individuals and negatively affects the knowledge flows between team members.

3. Data

3.1 Team

In the SciTS literature, the dominant definition of a team relies on co-authorship relationships (Ding et al., 2010; Wuchty et al., 2007). In our study, we refrain to base our team definition on publication data. Following Cummings and Keisler’s (2008) definition of a team based on grant applications, we consider a team as made by all the applicants who express their willingness to collaborate by submitting a joint grant application. This definition has three main advantages with respect to the one based on co-authorship relationships. First, it fits the definition of team as a group of individuals working together to achieve a common goal (Katz and Martin, 1997). The members of the team are the scientists having their names on the grant application and the goal of the team is explicitly stated in the grant application. Second, conversely to the common definition of a team based on co-authorship, this definition with clear boundaries allows us to capture even the teams not producing any publication and the members of the team not mentioned in an eventual publication outcome. Finally, we are able to determine the precise time when the team is formed independently from the time when the first co-authored article is published. According to this definition of team, individuals work together for a limited period in time pursuing a circumstantial goal.

3.2 Learning

At the team formation time, each individual is endowed with a knowledge stock represented by the literature she relied on in her research work. We proxy the latter as the list of distinct scientific journals cited in the papers she published before entering the team. We measure individual learning as the citations to new journals added to her knowledge stock after the team formation. The learning might be attributed to the interaction with other team members or not. We consider learning from within the team if the new journal citation observed was present in the knowledge stock of another team member before the team formation. If the new citation cannot be attributed to a knowledge flow from another team member, we classify it as not originating from within the team. It could originate from an outside collaboration or be the result of a self-learning process.

3.3 Empirical setting and sources

Our study is conducted in the context of the SINERGIA Swiss funding program. The program is sponsored by the Swiss National Science Foundation (SNSF) that is the leading Swiss institution supporting the national scientific research. It plays in Switzerland the same role as the National Science Foundation (NSF) in the United States. SINERGIA was launched in 2008 and represents a flagship in the SNSF's funding schemes portfolio. It is designed to promote team collaboration. As mentioned in the application guidelines, researchers are required to collaborate as a condition for securing research funding, i.e. researchers need to submit a proposal for a "research work carried out collaboratively" (SNSF, 2011).

In most cases, a SINERGIA project involves three or four researchers who appear as co-applicants in the grant application. All disciplines are eligible for funding through the program. Applicants propose interdisciplinary projects or projects where co-applicants belong to the same discipline, but are specialized in different sub-fields. The criteria considered in evaluating the application are the value added of the joint research approach, the research complementarities of the applying groups, and the coherence of the projected collaboration. Applications are screened in a two-step evaluation process. In the first step, external reviewers assign a provisional score to each application. In the second step, an internal committee of SNSF, the Specialized Committee for Interdisciplinary Research based in Bern, assigns to each application the final score using an alphabetical scale, where A is the highest score and D the lowest one. Applications are ranked and funds are assigned until the annual budget quota is reached. Typically, applications receiving a score below B are not founded. Since its introduction, the SNSF received about 500 applications and financed 40% of them investing 35% of its total budget.

From all grant applications submitted to the SNSF in the period 2008-2012, we selected applications in Engineering, Science & Medicine. Our final sample is represented by 231 grant applications that include 604 unique applicants. The SNSF provided us with grant application data including final scores assigned and final funding decision and basic demographic information on applicants (gender, nationality and birth year)[1]. We matched this information with applicants' publication records using the Scopus database[2].

SINERGIA funding program is aimed at established researchers. In the majority of the cases applicants are associate or full professors with good publication records. They have to demonstrate their ability to conduct excellent quality independent research. The average age of an applicant is 48.4 years, with a minimum of 32 and a maximum of 69. Only the 18% of the entire population of applicants is below 40. Figure 1 shows the distribution of the count of applicants' publications at the application time. The average number of applicants' publications is 39.42 and 85% of the applicants have more than 10 publications.

<INSERT FIGURE 1 ABOUT HERE>

The representative team in our sample is a small one. Ninety percent of the teams have less than five members. A team is composed, on average, by 3.65 members, with a minimum of 2 and maximum of 11. Considering the team composition, 21 nationalities are represented. About 15% of teams have only Swiss members, while the others are multi-nationality teams. The average

---

[1] All concerned applicants were contacted by the SNSF and had the possibility to oppose the transmission of their data.

[2] We match the applicant's surname and the first letter of the name with the author's surname and the first letter of the name. Then, we filter the correct matches by hand-checking the scientist identity according to the applicant's characteristics such as her affiliation, discipline, college names, and age.

number of nationalities in a team is 2.46, with a maximum of 6 nationalities. SINERGIA funding program is favoring inter-institution collaborations. On average, each group has members from 2.42 different affiliations, with a maximum of 6. According to the SNSF's application requirements researchers with a foreign affiliation are admitted to apply for the grant only if her competencies and skills are not available in Switzerland. Due to this constraint, when we look at the country affiliations, we note that the 80% of the teams include only Swiss affiliations. When classified by disciplines, 36% of applications are in Engineering, whereas the rest are in Science & Medicine[3]. Within the two broad disciplines, each application is classified in sub-disciplines. An application counts, on average, 3.39 sub-disciplines; only 20% of the applications involve only one discipline, while the most diversified application involves 11 disciplines. When we look at the previous collaborations among applicants at the application time, we observe that in the 60% of the cases there was at least one co-authorship relationship among team members. When looking at the applicants' gender distribution, in our sample women constitute 15% of the total. A SINERGIA grant covers personnel costs, research costs, coordination costs and, to a limited extent, investment costs. The average amount requested per application is 1,712,492 CHF, with a minimum of 349,901 CHF and a maximum of 6,854,573 CHF.

[3] In this study, we excluded from the original sample applications in the Humanities and Social Sciences because book contributions represent a large part of the field publication outcomes and are not collected with accuracy in the Scopus database. Applications in the Humanities and Social Sciences represent 19% of the total sample.

Figure 2 represents the distribution of the number of grant applications by score assigned and final funding decision. 10% of the applications obtained the maximum score, A, 45% of the applications were awarded.

<INSERT FIGURE 2 ABOUT HERE>

Table 1 reports the applicants' characteristics and Table 2 reports the team characteristics.

<INSERT TABLE 1 AND 2 ABOUT HERE>

3.4 Variables

In this section we describe the dependent variables, the independent variables and the controls used in the regression exercise. Our independent variables are grouped into two categories: team characteristics, and individual characteristics of the scientist in comparison to the rest of her team. For the latter we consider the three dimensions of distance, namely, geographical distance, social distance, and cognitive distance. Finally, our set of control variables includes the individual characteristics and the characteristics of the newly cited journals, i.e. the basic component of our measure of individual learning.

3.4.1 Dependent variables

Following Agrawal et al. (2008), we adopt as unit of analysis the scientist-journal cited pair. For instance, a team composed of four scientists, each of whom cites ten distinct journals, generates forty observations. The scientist-journal cited level of analysis allows us to study the micro-dynamics of the team members' learning processes by isolating each knowledge component and tracing its origin (Börner et al., 2010). For each scientist, we consider two time periods, before and after the team formation. We compare the knowledge stock of each individual in the two periods in order to measure the individual learning, namely the new journals cited that appear after the individual enters the team[4].

The dependent variable *Learning from within the team* is a dummy that equals one if the new journal cited by the scientist after the team formation is included in the stock of knowledge of at least other team member before team formation, it equals zero otherwise. In our study sample, when looking at the origin of learning, 39% of the new journals cited by a team member after the team formation originate from within the team. A possible concern in the definition of learning from within the team is that in the specific case of publications co-authored by two members of the same team, it is not possible to disentangle the contribution of each author to the list of references. In other words, we cannot exclude with certainty that the new journal cited by the scientists has been included in the list of references by her coauthor teammate and that she actually has not learned from her teammate. In order to avoid this overestimation of the individual learning, we define a second variable *Learning from within the team– no co-authored pubs* for which we exclude the newly cited journals in co-authored articles.

[4] In the main analysis, we do not impose any constraints on the time span before and after the application time. In a robustness check, available upon request, we fix a time window of three years before and after the application time and the results remain stable.

3.4.2 Team characteristics

We distinguish co-ethnic teams from multi-ethnic teams with a dummy *Co-ethnic team*. *Co-ethnic team* equals one when all the team members are from the same country of origin, zero otherwise. We use the dummy *At least one female scientist in the team* which equals one if at least one team member is a female scientist, zero otherwise, to differentiate mixed teams from only-male teams. We take into consideration the quality level of the research project of the team based on the score attributed by the SNSF evaluation committee. We include in the regression a dummy *High quality application (grade A)* that equals to one if the application obtains the maximum score, zero otherwise and a dummy *Low quality project (grade D)* that equals to one if the application obtains the minimum score, zero otherwise. The dummy *Awarded* concerns the final funding decision and it equals one if the SNSF awards the team of the scientist, zero otherwise. We use the variables *Amount requested* and *Number of team members* as proxies for the size of the team's project. Our sample includes teams working in two macro-fields: Engineering and Science & Medicine. The dummy *Science & Medicine* is equal to one for Science & Medicine, zero otherwise. Each team can submit an application that involves one or more sub-fields. Finally, we take into account the number of sub-disciplines listed in the grant application with the variable *Number of disciplines*.

3.4.3 Scientist vs. Rest of the team

We measure the geographical distance as the average time needed to travel from the affiliation of the scientist to the affiliations of the other team members (*Distance hours*)[5]. The average time needed to reach the other team members for the 604 individuals is 3.06 hours, with a standard deviation of 3.58 hours.

We measure the social distance over four dimensions. First, the age difference between the scientist and the rest of her team is calculated as the arithmetic difference between the age of the scientist and the average age of her teammates. The average difference is -0.12 years with a standard deviation of 8.76 years. Then, we standardize the arithmetic difference by subtracting its average and dividing by its standard deviation (*Standardized age difference scientist vs. team*). Second, in a similar fashion, we measure the productivity difference between the individual and the rest of her team based on the number of publications (*Standardized stock of pubs. Difference scientist vs. team*). The average productivity difference is -0.81 scientific publications with a standard deviation of 40.40 scientific publications. Third, we introduce the variable *Same gender scientist vs. team* that equals one if the team includes at least another team member of the same gender as the scientist, zero otherwise. This dummy equals one for 91.6% of the 604 individuals considered. Finally, we consider the *Established collaboration* dummy that equals one if the

[5] Another possible variable for measuring the geographical distance could have been the distance in kilometers but, as expected, the time needed to travel and the distance in kilometers are highly correlated (about 0.9), therefore we include in the regression only the connection time variable.

scientist has already worked with at least another member of the team in previous joint research projects, zero otherwise. Previous joint research projects are identified by co-authored scientific articles before the year of the team formation.

We calculate the cognitive distance between the scientist and her team in two steps. First, we define a matrix of distances between journals. Second, we use the journal distance matrix obtained to calculate the average cognitive distance between the scientist (S) and her team (T). The journal distance matrix is based on the assumption that the more two journals *i* and *j* are cited together in the references of the same scientific publication, the closer they are. If the two journals are frequently co-cited, we attribute to the pair a small distance value. On the contrary, if the two journals are rarely co-cited, we attribute to the pair a large distance value. We collect all the values of the distances between each pair of journals in the journal distance matrix (D). Specifically, we compute each cell D(*i*,*j*) of the matrix as the inverted ratio between the number of publications in which *i* and *j* are co-cited and the minimum number of publication where *i* or *j* are cited (Equation 1). The denominator of the ratio accounts for the fact that the probability of being co-cited depends also on the number of publications where only one of the two journals for which we are measuring the distance appears in the references. The distance measure ranges from one to infinity. In the case of an infinite distance, i.e. *i* and *j* are never co-cited in a publication, for computational reason, we attribute the maximum non-infinite distance of the journal *i* from all the other journals. The matrix distance relies on the references of all the articles included in our database, i.e. the ones published by the 604 scientists included in the analysis[6].

$$D(i,j) = \frac{1}{\frac{\#\text{pubs where } i \text{ and } j \text{ are co-cited}}{\min(\#\text{pubs where } i \text{ is cited}, \#\text{pubs where } j \text{ is cited})}}$$

(Equation 1)

As a second step, we use D to calculate the average cognitive distance between a scientist (S) and her team members (T). We consider the journals cited by the individual and the journals cited by her team before the team formation, then we calculate the average distance as in equation 2.

$$Cognitive\ distance_{S,T} = \sum_{i=1}^{\#S} \sum_{j=1}^{\#T} D(i,j)/(\#S * \#T)$$

(Equation 2)

Where #S is the count of the journals cited by S and #T is the count of journal cited by the other team members. We consider the average distance calculated in equation 2 as our measure of the cognitive distance between the scientist S and her team T ($Cognitive\ distance_{S,T}$).

[6] As robustness check we calculated the same matrix based on a database of all the articles published in the 100-top journals targeted by Swiss scientists. The journal distance matrix remains substantially unchanged as well as the results of our analysis.

3.4.4 Individual characteristics

We consider demographic characteristics as determinants of individual learning from within the team, such as age and gender of the scientist. We include in the regression exercise a dummy *Gender* that equals one if the individual is a female, zero otherwise. We include the age of the individual at the time of team formation (*Age*). Since the probability of observing a new citation is correlated with the scientist's productivity and her knowledge stock before entering the team, we control for the *Stock of publications pre-team entry* and the *Stock of journals cited pre-team entry*. As additional controls we consider the scientist's experience in SINERGIA project applications. We thus include a dummy *Multiple current applications* that is equal to one if the individual is participating to more than one project at the same time, zero otherwise. Finally, we take into account the number of previous applications, *Previous applications* and the number of successful ones, *Previous awarded applications.*

Table 3 reports the descriptive statistics concerning the total number of new citations after the team formation, and the proportion of these citations originating from within the team according to the characteristics of the applicants. Females tend to have a higher proportion of new citations than males. Young and less experienced scientists learn more than older and more experienced scientists. However, t-tests show that these differences are not significant.

<INSERT TABLE 3 ABOUT HERE>

3.4.5 Journal characteristics

We measure learning relying on the new journals cited by the scientist. Journal characteristics might affect the number of citations that a journal receives. Hence, in our regression exercise we control for the following journal characteristics. First, we include in the regression the number of articles where the journal is cited, *Journal frequency*. Second, we control for the fact that the journal is a generalist journal, *Generalists.* As generalists we consider the following journals: Nature, Science, PNAS, and PlosOne*)*. Finally, we control for the length of the history of the journal proxying its foundation year by the year when the first article published on the journal appears in our database, *History of journal*. For about 8% of the journals we are not able to identify the funding year, then we control with a dummy when this information is missing, *Unknown history*.

In table 4 we consider descriptive statistics at the scientist-journal cited level of analysis adopted in the regression exercise.

<INSERT TABLE 4 ABOUT HERE>

4. Estimation strategy

We estimate with two Probit models the probability that a new journal cited by the scientist, after the team formation, is the result of a process of learning from other team members. In the first model we consider *Learning from within the team* as dependent variable while in the second model we consider as dependent variable *Learning from within the team – no co-authored pubs.*

In the two Probit models we maintain the same set of explanatory variables. We group the explanatory variables in four vectors, namely *team characteristics*, *scientist vs. team*, *individual characteristics*, and *journal characteristics* (Equation 3).

$$\mathrm{P}(Learning\ from\ within\ the\ team = 1|\boldsymbol{x}) =$$

$$G(\beta_0 + \boldsymbol{team} * \boldsymbol{\beta_1} + \boldsymbol{scientist\ vs.team} * \boldsymbol{\beta_2}$$

$$+\boldsymbol{individual} * \boldsymbol{\beta_3} + \ \boldsymbol{journal} * \boldsymbol{\beta_4} + \boldsymbol{controls} * \boldsymbol{\beta_5})$$

(Equation 3)

Where G is the standard normal cumulative distribution function. In our estimations we clustered standard errors at the scientists' level.

5. Results

Table 4 reports the results of the estimation of the probability of learning from within the team. In the regressions of columns 1 and 3, we adopt as dependent variable *Learning from within the team* (61,068 scientist-journal cited observations) while in the regressions of column 2 and 4 we adopt as dependent variables *Learning from within the team – no co-authored pubs* (52,739 scientist-journal cited observations). Columns 1 and 2 consider only the controls, individual and journal characteristics, while columns 3 and 4 add the team characteristics and the scientist vs. team measures.

We find a weak correlation between the individual characteristics and the probability of learning from within the team. In particular, age and gender of the scientist have no impact on the probability of learning from the other team members. The stock of publications (*Stock of publications pre-application period*) is positively correlated with the probability of learning from other team members. On the contrary, the stock of journals cited is negatively correlated (*Stock of journals pre-application period*).

The journal characteristics are significantly correlated with the probability of learning from other team members. In particular, when journals are frequently cited in the bibliographies of the articles in our dataset (*Journal frequency*) and have a long history (*History of journal*), it is more likely that the citation to the corresponding journal originates from other team members. If the citation refers to a generalist journal (*Generalist*), it is less likely to originate from other team members.

When adding the team characteristics, we find that having at least one female member in the team (*At least one female scientist in the team*) is associated with a higher probability of learning from other team members. Being member of a co-ethnic team (*co-ethnic team*) has a significant positive effect only in the regression reported in column 3, which is the regression considering publications co-authored among team members. The coefficient of the dummy *Awarded* is not significant, i.e. we find that awarded and not awarded applicants have the same probability of learning from other team members. The quality of the application is not significantly correlated with the probability of learning from other team members, precisely, both the coefficients of the dummies *High quality application (grade A)* and *Low quality application (grade D)* are not significant. Finally, in larger teams the scientist has greater chances to learn from her teammates since the coefficients accounting for the size of the team (*Amount requested* and *number of team members*) are both significantly positive. Scientists in the fields of Science & Medicine (i.e dummy *Science & Medicine* equal to one) are associated with higher chances of learning from other team members than teams in Engineering (dummy equal to zero).

Looking at the individual characteristics of the scientist in comparison to the rest of her team, we find that the coefficient of the variable capturing the geographical distance (*Distance hours*) is not significant. This result is coherent with the findings of Mairesse and Turner (2005) stating that there is no impact of geographical distance on collaboration. As for the social distance between the scientist and her team, we observe that when scientific reputation (*Standardized stock pub. difference scientist vs. team*) and age (*Standardized age difference scientist vs. team*) increase, the probability of learning within the team decreases. This means that young team members and team members with limited scientific reputation benefit more of the learning from other team members than senior and experienced scientists. We observe that matching with teammate of the same gender (i.e. dummy *Same gender scientist vs. team* equal to one) is not significantly correlated with the probability of learning from within the team. Finally, we find that if the scientist has an established collaboration with at least one member of the team (*Established collaboration*) it is significantly more likely that she learns from her teammates.

<INSERT TABLE 4 ABOUT HERE>

Concerning the cognitive distance between the scientist and her team, we find an inverted U-shaped curve. This relationship means that when the cognitive distance between the scientist and the rest of her team is too high or too low, the scientist is less likely to learn from within the team. A medium level of cognitive distance maximizes the probability of learning from within the team. More precisely, we find a statistically significant and positive coefficient for the linear term of the variable *Cognitive distance* and a statistically significant and negative coefficient for its quadratic term[7]. We estimate the curve plotted in the Figure 3 below based on a simplified linear probability model adopting the same specification of the Probit model. These results confirm the findings of Nooteboom et al. (2007) concerning the two opposing effects of the cognitive distance on the ability to learn something new.

<INSERT FIGURE 3 ABOUT HERE>

The definition of team we adopted assumes that team formation is independent of the funding decision. This assumption might rise the concern that only awarded teams realize the prospective collaboration declared at the grant application time. To respond to this concern, we replicated the analysis of Table 4 considering separately awarded and non-awarded teams. The analysis that are available upon request show that the results remain stable for both the subsample of awarded and non-awarded teams. In particular, our main results on the significance of the effect of social and cognitive distance remain.

[7] BOX 1 in the appendix reports an extensive discussion on the statistical significance of this result in relation to the sample size.

6. Discussion and conclusions

This paper contributes to the SciTS literature by identifying the factors that promote the learning of the individuals from their team members when working in a scientific research team. Unique to our study is the fact that we measure the basic component of the knowledge stock and we keep track, within the team, of the knowledge flows from an individual to another one.

When an individual enters the team she contributes to it with her knowledge stock and, at the same time, she has the occasion to learn from other team members. We find that team characteristics and the characteristics of the scientists in comparison to the rest of her team affect the probability of receiving knowledge from the members of the team. Precisely, in terms of social distance, we find that having an established relationship with at least a team member, being the junior member of a team, and working with scientists with higher scientific reputation, enhance the probability of learning from within the team. As for the cognitive distance between the individual and the other team members, we find that it generates an inverted U-shaped relationship with the probability of observing knowledge flows from within the team. This result suggests that there is an optimal level of cognitive distance that favors learning. An individual should have a knowledge stock that differs from the one of the others in order to guarantee a buffer for learning something new. At the same time, the knowledge stock difference should not be too large to avoid that 'speaking a different language' obstacles an effective flow of knowledge among team members.

Our results have a direct implication for the SciTS literature. While a large part of this literature shows that researchers working in team have a higher productivity, we focus on the individual and team characteristics that stimulate team members' learning from their colleagues.

Nowadays, funding agencies are increasingly promoting the constitution of interdisciplinary teams for conducting research. For instance, the latest SINERGIA call for grants states as a new requirement for applicants the need to prove the interdisciplinary composition of the team. In this context, understanding the micro-dynamics of research teams is crucial. Our findings on the social distance and on the optimal level of cognitive distance among team members suggest that, in promoting interdisciplinary teamwork, particular attention should be devoted to team composition. While geographical distance has little impact on the knowledge flows among team members, social aspects should be taken into account. Previous experience of joint research work favors the team members' learning. Moreover, age and scientific reputation differences of team members direct the knowledge flows from senior and productive members to junior scientists with less experience in research. Finally, while it is common wisdom to promote cross-disciplinary research in order to stimulate creativity, this could have unexpected consequences. It is important to maintain a common knowledge base among team members in order to guarantee the knowledge flows absorption.

7. Bibliography

Adams, J.D., Black, G.C., Clemmons, J.R., Stephan, P.E., 2005. Scientific teams and institutional collaborations: Evidence from U.S. universities, 1981–1999. Research Policy 34, 259–285.

Agrawal, A., Cockburn, I., McHale, J., 2003. Gone but not forgotten: Labor flows, knowledge spillovers, and enduring social capital. National Bureau of Economic Research.

Agrawal, A., Kapur, D., McHale, J., 2008. How do spatial and social proximity influence knowledge flows? Evidence from patent data. Journal of Urban Economics 64, 258–269.

Apesteguia, J., Azmat, G., Iriberri, N., 2012. The Impact of Gender Composition on Team Performance and Decision Making: Evidence from the Field. Management Science 58, 78–93.

Argote, L., Miron-Spektor, E., 2011. Organizational Learning: From Experience to Knowledge. Organ. Sci. 22, 1123–1137.

Bercovitz, J., Feldman, M., 2011. The mechanisms of collaboration in inventive teams: Composition, social networks, and geography. Research Policy 40, 81–93.

Bikard, M., Murray, F., Gans, J.S., 2015. Exploring Trade-offs in the Organization of Scientific Work: Collaboration and Scientific Reward. Management Science.

Börner, K., Contractor, N., Falk-Krzesinski, H.J., Fiore, S.M., Hall, K.L., Keyton, J., Spring, B., Stokols, D., Trochim, W., Uzzi, B., 2010. A Multi-Level Systems Perspective for the Science of Team Science. Science Translational Medicine 2, 49.

Burt, R.S., 2004. Structural Holes and Good Ideas. American Journal of Sociology 110, 349–399.

Cummings, J.N., Kiesler, S., 2008. Who collaborates successfully? prior experience reduces collaboration barriers in distributed interdisciplinary research, in: Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work. ACM, pp. 437–446.

Defazio, D., Lockett, A., Wright, M., 2009. Funding incentives, collaborative dynamics and scientific productivity: Evidence from the EU framework program. Research Policy 38, 293–305.

Ding, W.W., Levin, S.G., Stephan, P.E., Winkler, A.E., 2010. The impact of information technology on academic scientists’ productivity and collaboration patterns. Management Science 56, 1439–1461.

Fiore, S.M., 2008. Interdisciplinarity as Teamwork: How the Science of Teams Can Inform Team Science. Small Group Research 39, 251–277.

Freeman, R.B., Huang, W., 2015. Collaborating with People Like Me: Ethnic Coauthorship within the United States. Journal of Labor Economics 33, S289–S318.

Granovetter, M., 1973. The strength of weak ties. American journal of sociology 78, l.

Huber, G.P., 1991. Organizational learning: The contributing processes and the literatures. Organization Science 2, 88–115.

Jaffe, A.B., Trajtenberg, M., Henderson, R., 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. the Quarterly journal of Economics 108, 577–598.

Jones, B.F., Wuchty, S., Uzzi, B., 2008. Multi-University Research Teams: Shifting Impact, Geography, and Stratification in Science. Science 322, 1259–1262.

Katz, J.S., Martin, B.R., 1997. What is research collaboration? Research policy 26, 1–18.

Knoben, J., Oerlemans, L.A.G., 2006. Proximity and inter-organizational collaboration: A literature review. International Journal of Management Reviews 8, 71–89.

Lee, S., Bozeman, B., 2005. The Impact of Research Collaboration on Scientific Productivity. Social Studies of Science 35, 673–702.

Levitt, B., March, J.G., 1988. Organizational learning. Annu. Rev. Sociol. 319–340.

Lin, M., Lucas, H.C., Shmueli, G., 2013. Too Big to Fail: Large Samples and the p-Value Problem. Information Systems Research 24, 906–917.

Mairesse, J., Pezzoni, M., 2015. Does Gender Affect Scientific Productivity? Revue Économique Vol. 66, 65–113.

Mairesse, J., Turner, L., 2005. Measurement and explanation of the intensity of co-publication in scientific research: An analysis at the laboratory level. National Bureau of Economic Research.

McPherson, M., Smith-Lovin, L., Cook, J.M., 2001. Birds of a feather: Homophily in social networks. Annual review of sociology 415–444.

Misra, S., Daniel S., Lulu, C., 2015. The Transdisciplinary Orientation Scale: Factor Structure and Relation to the Integrative Quality and Scope of Scientific Publications. Journal of Translational Medicine & Epidemiology, 3(2): 1042.

Mowatt, G., Shirran, L., Grimshaw, J.M., Rennie, D., Flanagin, A., Yank, V., MacLennan, G., Gøtzsche, P.C., Bero, L.A., 2002. Prevalence of honorary and ghost authorship in Cochrane reviews. Jama 287, 2769–2771.

Nooteboom, B., Van Haverbeke, W., Duysters, G., Gilsing, V., van den Oord, A., 2007. Optimal cognitive distance and absorptive capacity. Research Policy 36, 1016–1034.

Porac, J.F., Wade, J.B., Fischer, H.M., Brown, J., Kanfer, A., Bowker, G., 2004. Human capital heterogeneity, collaborative relationships, and publication patterns in a multidisciplinary scientific alliance: a comparative case study of two scientific teams. Res. Policy 33, 661–678.

Stokols, D., Hall, K.L., Taylor, B.K., Moser, R.P., 2008. The Science of Team Science. Am. J. Prev. Med. 35, S77–S89.

SNSF, 2011. Regulation on Sinergia Grants. National Research Council. http://www.snf.ch/SiteCollectionDocuments/sinergia_reglement_e.pdf

Uzzi, B., Mukherjee, S., Stringer, M., Jones, B., 2013. Atypical Combinations and Scientific Impact. Science 342, 468–472.

Whitfield, J., 2008. Collaboration: Group theory. Nature News 455, 720–723.

Woolley, A.W., Chabris, C.F., Pentland, A., Hashmi, N., Malone, T.W., 2010. Evidence for a Collective Intelligence Factor in the Performance of Human Groups. Science 330, 686–688.

Wuchty, S., Jones, B.F., Uzzi, B., 2007. The Increasing Dominance of Teams in Production of Knowledge. Science 316, 1036–1039.

Zenger, T.R., Lawrence, B.S., 1989. Organizational demography: The differential effects of age and tenure distributions on technical communication. Academy of Management journal 32, 353–376.

Figures and Tables

Figure 1: Distribution of the number of scientists' publications at the grant application time

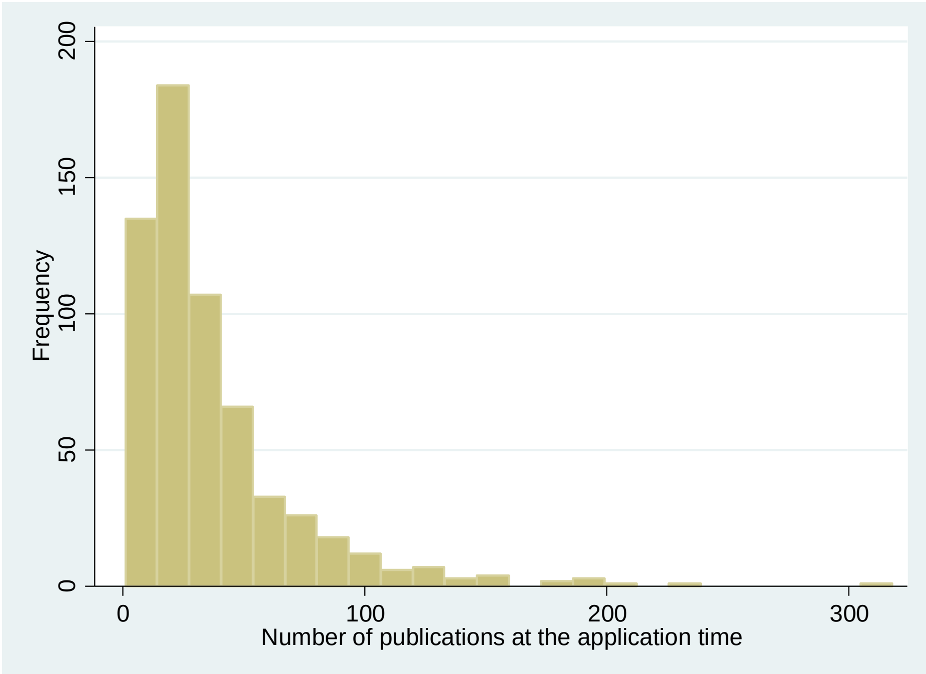


Figure 2: Distribution of grant applications by score assigned and final funding decision

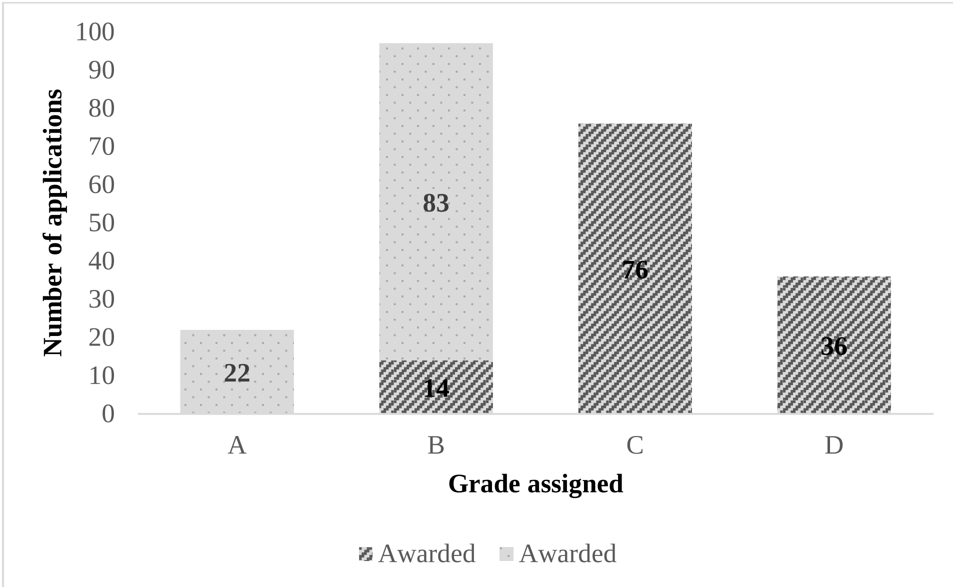


Table 1: Scientists' characteristics at the application time (number of applicants=604)

	Mean	Std. Dev.	Min	Max
Age	48.43	7.81	32	69
Gender (=1 for female, 0 otherwise)	0.15	0.36	0	1
Stock of publications pre-team entry	39.42	37.81	1	318
Stock of journals cited pre-team entry	102.70	84.97	1	496

Table 2: Team characteristics at the application time (number of teams=231)

	Mean	Std. Dev	Min	Max
Number of team members	3.65	1.56	2	11
Number of nationalities represented	2.46	1.03	1	6
Number of country affiliations	1.20	0.40	1	2
Number of affiliations	2.42	1.00	1	6
Number of disciplines	3.39	2.20	1	11
Average team members' age	51.66	5.57	35	69
Share of women	0.19	0.28	0	1
Average team members' stock of pubs	51.43	32.17	8	233
Awarded	0.45	0.50	0	1
High quality application (grade A)	0.10	0.29	0	1
Low quality application (grade D)	0.16	0.36	0	1
Amount requested	1712492	774832	349901	6854573

Table 3: Average number of new journals cited post application, and of learning from within the team by gender, age and stock of publications pre-team entry

	With co-authored pubs.			No co-authored pubs.		
	A. Number of new journals cited post application	B. Learning from within the team	B/A	C. Number of new journals cited post application	D. Learning from within the team	D/C
Female scientists	84.94	33.13	39.00%	68.73	22.61	32.90%
Male scientists	93.68	36.93	39.40%	81.71	29.85	36.50%
t-test	*0.44*	*0.4*		*0.23*	*0.08*	
Young scientists (Age<49)	89.86	37.25	41.50%	74.07	27.28	36.80%
Senior scientists (Age>49)	95.51	35.28	36.90%	86.83	30.63	35.30%
t-test	*0.48*	*0.54*		*0.10*	*0.25*	
Large stock of pubs. (Stock>43)	95.38	33.57	35.20%	86.90	29.09	33.50%
Limited stock of pubs. (Stock<43)	90.98	37.68	41.40%	76.45	28.63	37.40%
t-test	*0.61*	*0.24*		*0.21*	*0.88*	

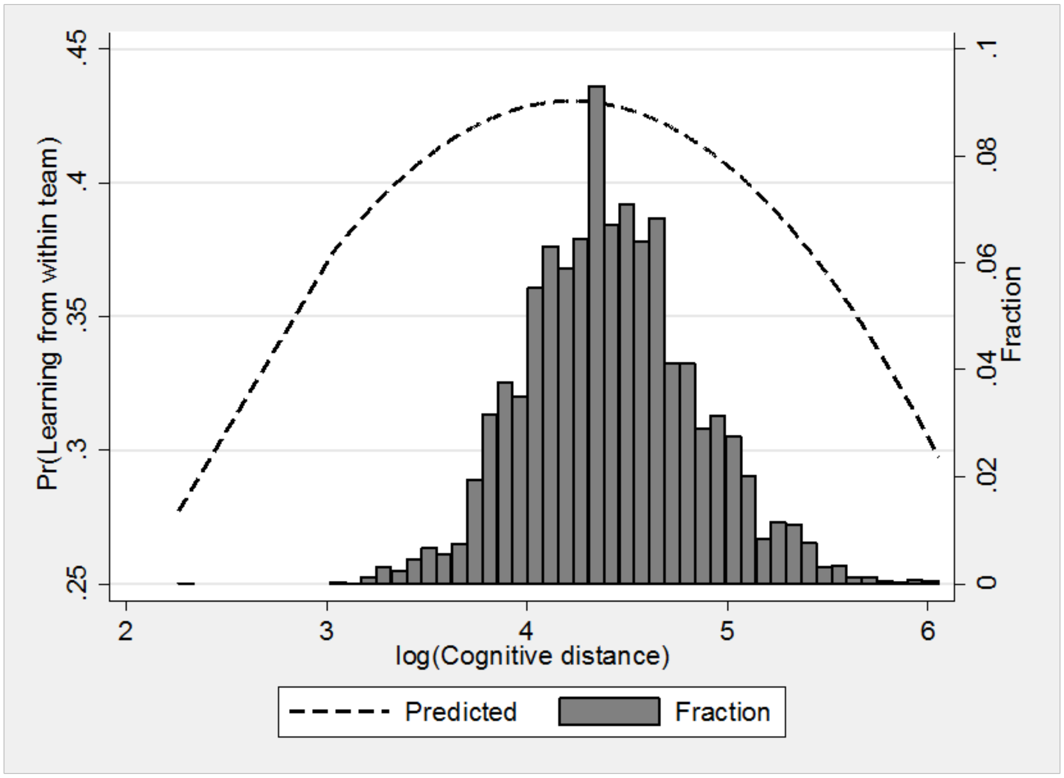
Table 4: Regression descriptive statistics considering the study sample used in Table 4, column 4 (52,739 scientist-journal cited observations)

VARIABLES	Obs.	Mean	Std.	Min	Max
Learning from within the team	61068	0.39	0.48	0.00	1.00
Learning from within the team- no co-authored pubs.	52739	0.36	0.48	0.00	1.00
Individual characteristics					
Age	52739	49.50	7.17	32.00	71.00
Gender (=1 for female, 0 otherwise)	52739	0.13	0.33	0.00	1.00
Log(Stock of pubs. pre-team entry)	52739	3.48	0.82	0.00	5.76
Log(Stock journals cited pre-team entry)	52739	4.80	0.80	0.00	6.23
Multiple current applications	52739	0.17	0.38	0.00	1.00
Previous awarded applications	52739	0.12	0.33	0.00	1.00
Previous applications	52739	0.36	0.48	0.00	1.00
Team characteristics					
Co-ethnic team	52739	0.17	0.37	0.00	1.00
At least one female scientist in the team	52739	0.35	0.48	0.00	1.00
Awarded	52739	0.44	0.50	0.00	1.00
High quality application (grade A)	52739	0.10	0.30	0.00	1.00
Low quality application (grade D)	52739	0.17	0.38	0.00	1.00
Log(Amount requested)	52739	14.34	0.42	12.77	15.74
Log(Number of team members)	52739	1.34	0.37	0.69	2.40
Log(Number of sub-disciplines)	52739	1.02	0.67	0.00	2.40
Science & Medicine	52739	0.74	0.44	0.00	1.00
Geographical distance					
Log(1+Distance hours)	52739	1.11	0.72	0.00	3.58
Social distance					
Same gender scientist vs. team	52739	0.91	0.29	0.00	1.00
Standardized stock pubs. difference scientist vs. team	52739	-0.04	0.96	-5.62	6.35
Standardized age difference scientist vs. team	52739	-0.01	0.99	-2.83	3.13
Established collaboration	52739	0.35	0.47	0.00	1.00
Cognitive distance					
Log(Cognitive distance$_{S,T}$)	52739	4.42	0.44	0.96	6.05
Log(Cognitive distance$_{S,T}$)^2	52739	19.73	3.92	0.93	36.62
Journal characteristics					
Log(Journal frequency)	52739	5.37	1.02	3.93	9.34
Generalists (NATURE,SCIENCE,PNAS,PLOS)	52739	0.01	0.10	0.00	1.00
Log(History of journal)	52739	3.09	1.10	0.00	4.45
Unknown history	52739	0.08	0.28	0.00	1.00

Table 5: Regression results for the probability of learning from within the team

VARIABLES	(1) Learning from within the team co-authored pubs.	(2) Learning from within the team no co-authored pubs.	(3) Learning from within the team co-authored pubs.	(4) Learning from within the team no co-authored pubs.
Individual characteristics				
Age	-0.0023	-0.00066	0.0019	0.0014
Gender (=1 for female, 0 otherwise)	-0.030	-0.054	-0.029	-0.032
Log(Stock of pubs. pre-team entry)	-0.029	-0.031*	0.050***	0.052***
Log(Stock journals cited pre-team entry)	-0.037***	-0.018	-0.065***	-0.047***
Multiple current applications	-0.057*	-0.032	-0.027	-0.014
Previous awarded applications	0.025	0.016	-0.010	-0.030
Previous applications	0.014	0.013	0.0016	0.0094
Team characteristics				
Co-ethnic team			0.037*	0.010
At least one female scientist in the team			0.032**	0.029*
Awarded			-0.0061	-0.013
High quality application (grade A)			0.031	0.029
Low quality application (grade D)			-0.037	-0.035
Log(Amount requested)			0.080***	0.063***
Log(Number of team members)			0.28***	0.28***
Log(Number of sub-disciplines)			0.018	0.0047
Science & Medicine			0.077***	0.072***
Geographical distance				
Log(1+Distance hours)			0.0086	0.0089
Social distance				
Same gender scientist vs. team			0.0043	0.017
Standardized stock pub. difference scientist vs. team			-0.097***	-0.099***
Standardized age difference scientist vs. team			-0.031**	-0.024*
Established collaboration			0.056***	0.034**
Cognitive distance				
Log(cognitive distance$_{S,T}$)			0.56**	0.40*
Log(cognitive distance$_{S,T}$)^2			-0.066**	-0.048*
Journal characteristics				
Log(Journal frequency)	0.11***	0.11***	0.12***	0.12***
Generalists (NATURE, SCIENCE, PNAS, PLOS)	-0.064**	-0.037	-0.085***	-0.055**
Log(History of journal)	0.086***	0.082***	0.085***	0.080***
Unknown history	0.30***	0.30***	0.30***	0.28***
Pseudo R^2	0.06	0.06	0.12	0.11
Observations	61,068	52,739	61,068	52,739

Figure 3: Predicted probability of learning from within the team vs. cognitive distance



Appendix

Box 1. When econometric analyses are conducted on large samples, the standard econometric levels of significance of the estimated coefficients have to be treated carefully. With large samples even regression coefficients with a negligible economic impact (i.e. small size of the coefficient) might result to be statistically significant. In this paper, we are in the case of a large sample of 52739 observations (Table 4, column 4). In this appendix we comment the significance of the coefficients estimated for one of our main variable of interest. Figure 3 shows that the prediction of the probability of learning from within the team varies by an economically significant extent from a probability of about 27% to 37%. We go beyond the statistical significance of the coefficient by considering a Monte-Carlo CPS chart to show that the impact of the linear and quadratic components of the cognitive distance are significant already for smaller samples randomly drawn (Lin et al., 2013). The Monte-Carlo CPS approach draws random observations for different sample sizes, ranging from 100 to 9100 observations. For each sample size it extracts 100 samples. For each of the 100 samples our econometric model is estimated. In Figure Box1a we report the boxplots of the P-values of the linear term of the cognitive distance for each sample size. In Figure Box1b we report the boxplots of the P-values of the quadratic term of the cognitive distance for each sample size. We find that the P-value converge very quickly to the values observed with the complete sample, well before reaching the 52739 observations used in the regression in Table 4 column 4. Both the economically significant extent of the impact of cognitive distance (a variation of about 10% in the probability of learning from within the team) and the Monte-Carlo CPS chart, confirm that the impact of the cognitive distance is not a pure statistical artifact.

Figure Box1a: Boxplots of the extent of the estimated coefficient (left side) and of the P-value (right side) of the linear term of the cognitive distance for each sample size.

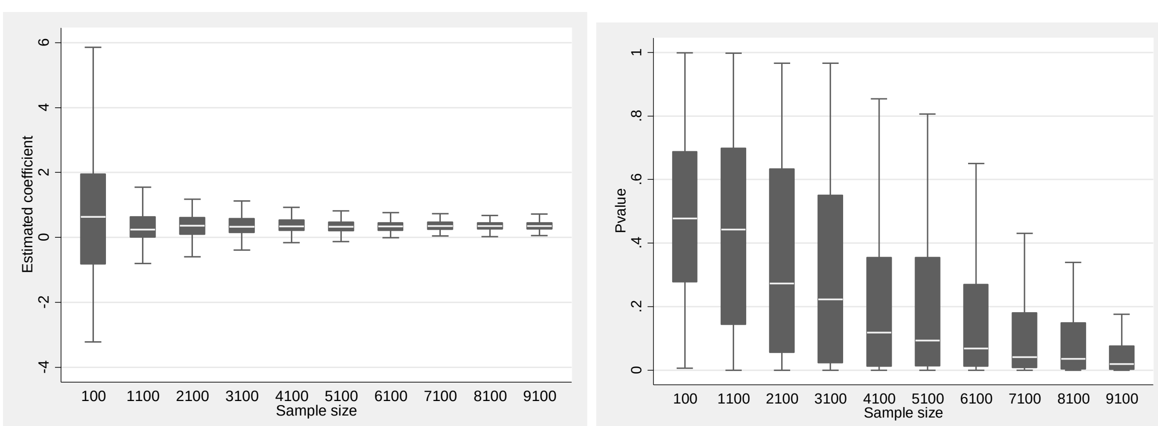


Figure Box1b: Boxplots of the extent of the estimated coefficient (left side) and of the P-value (right side) of the quadratic term of the cognitive distance for each sample size.