AT THE ORIGINS OF LEARNING: ABSORBING KNOWLEDGE FLOWS FROM WITHIN OR OUTSIDE THE TEAM?

Documents de travail GREDEG
GREDEG Working Papers Series

Charles Ayoubi
Michele Pezzoni
Fabiana Visentin

GREDEG WP No. 2016-08
http://www.gredeg.cnrs.fr/working-papers.html

AT THE ORIGINS OF LEARNING: ABSORBING KNOWLEDGE FLOWS FROM WITHIN OR OUTSIDE 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

GREDEG Working Paper No. 2016-08

Abstract: Empirical studies document a positive effect of collaboration on team productivity. The most common explanation to the teamwork productivity gain is that teamwork stimulates knowledge sharing among team members. 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 team collaboration. We find that team characteristics play a key role in favoring knowledge flows among team members. Specifically, we find a significant effect of the social distance and an inverted U-shape effect of the cognitive distance of the team members on the probability of learning from within the team.

Keywords: team, learning process, knowledge flows, cognitive distance, social distance, and geographical distance

Acknowledgments

We are indebted to Francesco Lissoni and Jacques Mairesse for their invaluable advice. We also thank Dominique Foray, Cesar A. Hidalgo, Brian Uzzi, Reinhilde Veugelers, Bruce Weinberg, Patrick Llerena, Stefano Bianchini and the seminar participants at the University of Bordeaux, University of Strasbourg and International Symposium on Science of Science 2016. We thank also the SNSF for the technical support.

1. Introduction

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 the emergence of the "big science" paradigm and the decline of the "small science" paradigm, as defined by Price (1963). 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 favors 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 is greater than the sum of its parts" (Katz and Martin, 1997). Even if this view is not exempt from critiques (Bikard et al., 2015; Mowatt et al., 2002), several studies have found that collaboration has a positive impact on publication productivity. For instance, Wuchty et al. (2007) found that teams produce exceptionally high-impact research, even in fields that once were the traditional domain of solo-author works. 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 knowledge that prompts scientific breakthroughs (Uzzi et al., 2013). However, little has been done to measure precisely how knowledge flows among team members and even less to assess the effect of teamwork on the individual learning of the scientists.

In our study, we define as learning the increment in the individual stock of knowledge recorded after the entry in a team. We contribute to the existing literature on collaboration by identifying the factors that promote the learning of an individual from the other team members. We claim that such portion of learning can be affected by individual and team characteristics. As individual characteristics, we consider biographical characteristics, such as age, gender, and scientific reputation. As team characteristics, we compare the focal individual with her team members focusing on three distance measures, namely, geographical distance (Jaffe et al., 1993), social distance (Agrawal et al., 2008, 2003) and cognitive distance (Nooteboom et al., 2007).

In our analysis, we use a novel dataset of 231 grant applications to a Swiss funding program promoting team collaboration. We define a team as made by the researchers who express their willingness to collaborate by submitting a joint grant application. For all applicants we collected their publication records as well as information about their biographical characteristics. We use publication records to identify the knowledge components that are the building blocks for our measures of individual knowledge stock, individual learning, and knowledge flow among team members. Following Uzzi et al. (2013), we consider each journal cited in the bibliography of the articles published by a focal individual as a distinct knowledge component. The sum of all the knowledge components represents the individual knowledge stock, the acquisition of new components represents the individual learning, and, finally, the component acquired from another team member represents the learning from other team members. In our setting, the focal applicant enters the team when she submits with other researchers an application to the Swiss funding program. Her knowledge stock is represented by the stock of distinct journals cited in her publications at the application time. We consider as learning the new journals cited after the application submission. Finally, we matched the new journals cited by the focal individual with the journals cited by the other team members before the application submission and we flag correspondences as knowledge flows from within the team. The new journals cited that do not match are considered knowledge flows from outside the team or the result of a self-learning process.

We find that learning from team members is more likely when the team is already established and when there is at least one woman in the team. Young team members and team members with a limited stock of knowledge are more likely to learn from senior and more experienced peers. We find an inverted U-shape impact of the cognitive distance between the focal 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.

Our results speak to the existing literature on collaboration (Defazio et al., 2009; Katz and Martin, 1997; Lee and Bozeman, 2005; Wuchty et al., 2007). These studies have analyzed the aggregate team productivity and the individual productivity of scientists working in a team. In this paper we investigate the mechanisms of knowledge transmission and learning that often have been taken for granted in the collaboration literature. Precisely, we asses to what extent team characteristics affect the individual probability to learn from other team members.

The rest of this paper is organized as follows: Section 2 illustrates the main concepts and definitions, Section 3 describes the empirical context and data, Section 4 presents the empirical model and discusses the main variables, Section 5 provides the results, and Section 6 concludes.

2. Main concepts and definitions

In section 2.1 we illustrate our definitions of research team, knowledge flows and individual learning. Then, in section 2.2, we introduce the team characteristics that are relevant for the analysis. In comparing the focal individual with her team members, we consider three distance dimensions, namely social, geographical and cognitive distance.

2.1 Team boundaries, knowledge flows, and individual learning

In the team literature, a team is often defined relying on co-authorship relationships in scientific publications (Ding et al., 2010; Wuchty et al., 2007). In our study, we refrain to base our team definition on bibliometric data. We define 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 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 goal of the team is precisely stated in the grant application. Second, according to this definition we are able to identifying team boundaries that include members with no publication outcomes. Finally, we are able to determine the precise time when the team is formed.

According to this novel definition of team, individuals work together for a limited period in time pursuing a circumstantial goal. At the grant application time, i.e. the team formation time, each individual is endowed with a knowledge stock. We proxy the latter as the list of distinct scientific journals cited in the papers she published before entering the team. After the team formation time, each team member might add new journal citations to her knowledge stock by citing new references in her future published work. The learning outcome measured by the new references might be attributed to the interaction with other team members or not. We consider learning from within the team if the new 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 within the team, we classify its origin from outside the team or as self-learning.

2.2 Team members' characteristics

Social distance

In searching for the team characteristics that affect the probability of individual learning within a team, we borrow from the literature assessing the impact of team composition on team members' scientific productivity (Katz and Martin, 1997). In our study, we take into account five social dimensions that characterize the team.

The first social dimension we consider is the existence of previous professional collaborations among team members. We distinguish established and new teams according to the fact that team members have already experienced or not a co-authorship at the application time. Within established and newly formed teams the mechanisms that affect knowledge flows among team members are different. On one side, members of established teams 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). However, previous professional collaborations might increase the probability that team members shared the same knowledge stock and be detrimental for the probability that individuals learn from each other (Burt, 2004). On the other side, members of newly formed team face the cost of creating new collaborative routines and to create new relational ties. Weak relational ties prevent knowledge diffusion. However, there is a higher probability that the knowledge stock carried by an individual who collaborate for the first time is non-redundant knowledge. The contrasting effects of the mechanisms at work within established and newly formed teams do not allow as making a prediction on their effect on the probability of learning from other team members.

The second social dimension that we consider in our analysis is the ethnic composition of the team. 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 that co-ethnic teams favor knowledge flows among team members.

The third social dimension we consider is the gender composition of the team. In the literature results on the effect of the gender composition on team productivity are mixed (Apesteguia et al., 2012; Woolley et al., 2010). Moreover, extant studies are rarely able to identify a causal relationship and provide tentative non-testable explanations to the empirical evidence. The mixed empirical results in the team-gender productivity literature and the lack of convincing explanations do not allow us to make any predictions about the impact of gender composition of the team on the team members' learning opportunities. 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).

The fourth social dimension we consider in our analysis is the age distribution of the team members. We expect the age of an individual to be positively correlated with her stock of knowledge and to 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 senior team members, i.e. 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 hypotheses prevent us to formulate a prediction on the age difference effects.

Finally, 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. First, highly productive members might contribute to the team with larger knowledge stocks and enhance the probability of learning of less productive team members. Second, highly productive scientists might focus on knowledge exchanges with peers with similar publications stocks from which they benefit more and isolate low productive scientists from which they benefit less. As in the case of age composition, we have two competing hypothesis about the possible effect of scientific team composition reputation.

Geographical distance

Over the last thirty years or so, in parallel to the increase in the average team size, we witness an even higher increase in the geographic dispersion of the team members. 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 transport 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 other team members.

Cognitive distance

In the management literature another central 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). We proxy the cognitive distance between team members by 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 the team. 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 focal scientist to absorb knowledge similar to the one she already has from other team members. 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 focal scientist. The so called "*opening new horizons*" effect has a positive impact on knowledge flows within the team as the cognitive distance increases.

Hence, by 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. Empirical context and data

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 field, 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 and Medicine. Our final sample is represented by 231 grant applications that include 604 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.

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 48.8 years, with a minimum of 32 and a maximum of 71. Only the 18% of the entire population of applicants is below 40. Figure 1 shows the distribution of the count of

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

applicants' publications at the application time. The average number of applicants' publications is 42.62, 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 3.651 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 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 constrain, 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 and Medicine. Within the two broad disciplines, each application is classified in sub-disciplines. An application counts, on average, 3.38 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 observed 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 1,719,053 CHF, with a minimum of 349,901 CHF and a maximum of 6,854,573 CHF.

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, 52% 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>

4. Empirical strategy

4.1 Models and dependent variables

Following Agrawal et al. (2008), we adopt as unit of analysis the application-applicant-journal citation pair. This unit of analysis allows us to study the micro-dynamics of the team members' learning processes by isolating and tracing each knowledge flow component and its origin (Börner et al., 2010). Each observation is a single knowledge component attributed to the focal applicant involved in the team represented by the applicants listed in the grant application. For instance, an application with four different applicants, each of whom cites ten journals, generates forty observations.

For each applicant-application, we consider two time periods, before and after the application time. 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 application time.

We estimate with a Probit model the probability that the journal cited by the focal scientist after the grant application is a new citation that originates from within the team, i.e. probability of *learning within the team* for the focal individual (equation 1). Thus, our dependent variable is a dummy that equals one if the new journal cited is listed in the stock of journals cited by the other team members before the application, zero otherwise.

However, in case of co-authorship among team members after the application time, it is not possible to disentangle the contribution of each author to the list of references cited in the bibliographies of the articles. For this reason, we adopt two approaches for calculating our dependent variable. In the first approach we suppose that new cited journals in co-authored articles with other team members are the result of a process of learning of the focal individual. In the second approach, more conservative, we exclude the co-authored articles when we calculate the new cited journals. The latter approach allows us to discard the possibility of a joint contribution as the result of the work of two or more team members where each brings her knowledge stocks in an independent way without reciprocal individual learning.

Then, we consider two distinct regression exercises characterized by two different definitions of the dependent variable, one including the articles co-authored by the focal scientist with the other team members after the team formation (*learning from within the team – co-authored*) and one excluding the articles co-authored by the focal scientist with the other team members after the team formation (*learning from within the team – no co-authored*). In the two regression exercises we include the same set of explanatory variables. We group the explanatory variables in three vectors of characteristics, namely *team characteristics, focal individual characteristics*, and *journal characteristics.* Particular interest in our analysis is devoted to the impact of the team characteristics, in particular on the distance between the focal individual and her team according to three dimensions, i.e. geographical, social and cognitive distance (equation 2).

$$\Pr(Learning\ within\ the\ team) =$$

$$\beta_0 + \beta_1 * team\ c. + \beta_2 * focal\ individual\ c. + \beta_3 * journal\ c.$$

(Equation 1)

$$team\ c. = (Geographical\ dist., Social\ dist., Cognitive\ dist.)$$

(Equation 2)

4.2 Team Characteristics, Social distance

As detailed in Section 2, we consider the following social characteristics of the team: (1) existence of previous professional collaborations among team members; (2) ethnic distance among team members; (3) gender composition of the team; (4) age distance among team members; (5) distance in terms of publication stock among team members.

The *Established team* dummy equals one if at least two team members have already worked together in previous joint research projects, zero otherwise. Previous joint research projects are identified by co-authored scientific publications before the year of the grant application. We distinguish co-ethnic teams from teams where members have different ethnic origins with a dummy *Co-ethnic team*. *Co-ethnic team* dummy equals one when all the team members are characterized by having the same country of origin, zero otherwise. We introduce two variables concerning the gender composition of the team. A first variable, *Same gender focal ind.- team,* is a dummy that equals one if the team includes at least another team member of the same gender as the focal individual, zero otherwise. This dummy equals one for 91.6% of the 604 focal individuals considered. A second variable, *At least one female scientist in the team*, is a dummy that equals one if at least one team member is a female scientist. This variable equals one in 40.2% of the teams. We include in the regression a variable that aims to measure the age difference between the focal individual and her team. We calculate the arithmetic difference between the age of the focal individual and the average age of her team. The average difference is -0.12 years with a standard deviation of 8.76 years. Then, we standardized the arithmetic difference by subtracting its average and dividing by its standard deviation (*Standardized age difference focal ind. – team*). In the same fashion, we measure the productivity difference between the focal individual and the rest of her team. The average productivity difference is -0.81 scientific publications with a standard deviation of 40.40 scientific publications. We standardize the arithmetic difference between her stock of scientific publications before the year of the grant application and the average stock of papers published by the other team members (*Standardized stock of pub. Difference focal ind.- team*).

4.3 Team Characteristics, Geographical distance

We measure the geographical distance as the average time needed to travel from the affiliation of the focal individual to the affiliations of the other team members. We also attempt to use the distance in kilometers but, as expected, the time needed to travel and the distance in kilometers are highly correlated (about 0.9), then we include in the regression only the connection time variable. The average time needed to travel to the rest of the team members for the 604 focal individuals is 3.06 hours, with a standard deviation of 3.58 hours.

4.4 Team Characteristics, Cognitive distance

Cognitive distance measures the distance between the knowledge stock of the focal individual and her team. We calculate this distance in two steps. First, we define a matrix of distances between scientific journals. Second, we use the journal distance matrix to calculate the average distance between the journals cited by the focal individual (A) and the journal cited by the rest of her team (T).

The journal distance matrix bases on the assumption that the distance between two journals is a decreasing function of the number of scientific article bibliographies where the two journals are co-cited. If two journals are co-cited frequently in the bibliographies of the scientific publications, we attribute to the pair a small distance value in term of contents. On the contrary, if two journals are rarely co-cited, we attribute to the pair a large distance value in term of contents. Moreover, the probability of being co-cited depends also on the number of publications which cite each of the two journals for which we are measuring the distance. Thus, we calculate the distance for a pair of journals *i* and *j* 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. The distance measure ranges from 1 to infinite. In the case of infinite distance, i.e. *i* and *j* are never co-cited in a publication, for computational reason, we consider as distance the maximum non-infinite distance of the journal *i* from all the other journals. Equation 3 shows the calculation of the journal distance.

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

(Equation 3)

We collect all the values of the distances between each pair of journals in the journal distance matrix (D).

As second step, we use D to calculate the average distance between the focal individual (A) and her team members (T). We consider the journals cited by the focal individual and the journals cited by her team before the application time, then we calculate the average distance as in equation 4

$$D_{A,T} = \sum_{i=1}^{\#A} \sum_{j=1}^{\#T} D(i,j)/(\#A * \#T)$$

(Equation 4)

Where #A is the count of the journals cited by A and #T is the count of journal cited by the other team members.

4.5 Other team characteristics

In the regression we control for a set of other characteristics of the application. Specifically, we control for the quality of the application distinguishing top-quality and low-quality applications. We add a dummy *Research project quality (grade A)* that is equal to 1 if the application obtains the maximum score, 0 otherwise. We include a dummy *Research project quality (grade D)* that is equal to 1 if the application obtains the minimum score, 0 otherwise. The dummy *Awarded* concerns the final funding decision and it equals 1 if the SNSF awards the team of the focal scientist, 0 otherwise. The *Amount requested* and the *Number of team members* are proxies for the size of the project. Our sample includes applications in two macro-fields: Engineering and Science and Medicine. The dummy *Science and Medicine* is equal to 1 for applications in the latter field, 0 otherwise. Each application could be specialized in one or more sub-fields. We take in account the level of inter-disciplinarity of each application by including in our models the number of sub-disciplines listed in the grant application, *Number of disciplines*.

4.6 Focal individual's characteristics

We consider demographic characteristics such as age at the application time, *Age*, and gender of the focal individual. For the gender we include in the regression exercise a dummy *Female* that equals to 1 if the focal individual is a female, 0 otherwise. Since the probability of observing a new citation is correlated with the stock of existing citations, we control for the *Stock of publications pre-grant period* and the *Stock of journals cited pre-grant period*. As additional controls we consider the applicant's experience in SINERGIA projects. We include a dummy *Multiple current applications* that is equal to 1 if the focal applicant is participating to more than one project at the same time. Finally, we take into account the number of previous applications, *Previous applications* and the number of successful ones, *Previous awarded applications.*

4.7 Journal characteristics

Our measure of learning is based on the journals cited. Journal characteristics might impact on the probability of learning from other team members. In our regression exercise we control for the following journal characteristics. First, we include in the regression the number of articles where the focal journal is cited, *Journal frequency*. Second, we control for the fact that the focal journal is a generalist journal, *Generalists (NATURE, SCIENCE, PNAS, and PLOS)*. Finally, we control for the length of the history of the focal 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 7% 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*.

Table A1, in the appendix, reports the summary statistics for all the variables included in the regressions.

5. Results

Table 3 reports the results of the estimation of the probability of learning within the team. The dependent variable in the regressions in columns 1 and 3, includes the new cited journals appearing in team members' co-authored articles (*learning from within the team– co-authored*), while the dependent variable in column 2 and 4 exclude the co-authored articles (*learning from within the team – no co-authored*).

We find a limited impact of the individual characteristics. In particular, age and gender of the focal scientist have no impact on the probability of learning from the other team members. The stock of publications (*stock of publications pre-application period*) has a positive impact on the probability of learning from other team members. On the contrary, the stock of journals cited has a negative impact (*stock of journals pre-application period*).

Referring to the team characteristics, we find no impact of the geographical distance and we find an impact of the social and cognitive distance. Being part of an established team (*Established team*) enhances the probability that learning comes from other team members. We find the same effect if there is at least one female member in the team (*At least one female scientist in the team*). When the social distance between the focal individual and her team in terms of scientific reputation (*Standardized stock pub. difference focal ind. - team*) and age (*Standardized age difference focal ind. - team*) increases, the probability of learning within the team decreases. This means that young team members with limited scientific reputation benefit more than more senior and experienced scientists of knowledge flows from within the team.

Being member of a co-ethnic team has a significant positive effect only in the regression reported in column 3, which is in the case where we consider publications co-authored among team members.

<INSERT TABLE 3 ABOUT HERE>

Cognitive distance between the focal scientist and the team has a statistically significant impact on the probability of learning from other team members (Figure 3). We estimate the curve based on a simplified linear probability model adopting the same specification of the Probit model. We find an inverted U-shaped curve, which shows that when cognitive distance of the focal scientist from the rest of her team is high or 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[2].

<INSERT FIGURE 3 ABOUT HERE>

We control for other team characteristics. We find that awarded and not awarded applicants have the same probability of learning from other team members (*Awarded*). The quality of the application does not impact on the probability of learning from other team members (*Research*

[2] An extensive discussion on the statistical significance of this result in relation to the sample size can be found in BOX 1 in the appendix.

project quality). Finally, in larger teams the focal scientist has greater chances to learn from others (*Amount requested* and *number of team members*). Scientists in the fields of *science and medicine* have more chances to learn from other team members than team in engineering.

The journal characteristics affect significantly 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.

The definition of team we adopted assumes that team formation is independent from the funding decision. A possible concern is that only awarded teams are incentivized to realize the prospective collaboration declared at the application time. To respond to this concern, we consider separately awarded and non-awarded teams. We replicate the analysis of table 3 for these two subsamples. 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.

6. Discussion and conclusions

This paper identifies the factors that promote the learning of the individuals from their team members when working in team. We look at the determinants of the origins of the learning. Unique to our study is the fact that we measure the basic component of the knowledge stock and we keep track, within the team, 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 individual characteristics and team characteristics for the social distance and the cognitive distance of team members affect the probability that the new knowledge acquired originates from within the team. We find that being part of an established team, being the junior member of a team, and working with scientists with higher scientific reputation, enhance the probability that the learning originates within the team. When we consider the cognitive distance between the focal individual and the other team members, we find an inverted U-shaped relationship between the cognitive distance and the probability of observing knowledge flows from within the team. This result shows that there is an optimal cognitive distance level 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 communication.

Our results have a direct implication for the literature on team collaboration. 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. Our findings on the social distance and on the optimal level of cognitive distance among team members lead us to suggest that, in promoting teamwork, particular attention should be devoted to team composition. Social aspects should be taken into account. Previous experience of joint research work favors the team members' learning as well as the presence of at least one female member. 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. Team members' geographical distance is not a determinant of learning. Finally, while it is common wisdom to promote multidisciplinary 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. Res. Policy 34, 259–285. doi:10.1016/j.respol.2005.01.014

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. J. Urban Econ. 64, 258–269. doi:10.1016/j.jue.2008.01.003

Apesteguia, J., Azmat, G., Iriberri, N., 2012. The Impact of Gender Composition on Team Performance and Decision Making: Evidence from the Field. Manag. Sci. 58, 78–93. doi:10.1287/mnsc.1110.1348

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

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. Sci. Transl. Med. 2, 49cm24–49cm24. doi:10.1126/scitranslmed.3001399

Burt, R.S., 2004. Structural Holes and Good Ideas. Am. J. Sociol. 110, 349–399. doi:10.1086/421787

Defazio, D., Lockett, A., Wright, M., 2009. Funding incentives, collaborative dynamics and scientific productivity: Evidence from the EU framework program. Res. Policy 38, 293–305. doi:10.1016/j.respol.2008.11.008

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. Manag. Sci. 56, 1439–1461.

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

Granovetter, M., 1973. The strength of weak ties. Am. J. Sociol. 78, l.

Jaffe, A.B., Trajtenberg, M., Henderson, R., 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. Q. J. Econ. 108, 577–598.

Jones, B.F., Wuchty, S., Uzzi, B., 2008. Multi-University Research Teams: Shifting Impact, Geography, and Stratification in. Science 322, 1259–1262. doi:10.1126/science.1158357

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

Knoben, J., Oerlemans, L.A.G., 2006. Proximity and inter-organizational collaboration: A literature review. Int. J. Manag. Rev. 8, 71–89. doi:10.1111/j.1468-2370.2006.00121.x

Lee, S., Bozeman, B., 2005. The Impact of Research Collaboration on Scientific Productivity. Soc. Stud. Sci. 35, 673–702. doi:10.1177/0306312705052359

Lin, M., Lucas, H.C., Shmueli, G., 2013. Too Big to Fail: Large Samples and the p-Value Problem. Inf. Syst. Res. 24, 906–917. doi:10.1287/isre.2013.0480

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. Annu. Rev. Sociol. 415–444.

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. Res. Policy 36, 1016–1034. doi:10.1016/j.respol.2007.04.003

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. doi:10.1016/j.respol.2004.01.007

Price, D. de S., 1963. Big science, little science. Columbia Univ. N. Y. 119–119.

Uzzi, B., Mukherjee, S., Stringer, M., Jones, B., 2013. Atypical Combinations and Scientific Impact. Science 342, 468–472. doi:10.1126/science.1240474

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. doi:10.1126/science.1193147

Wuchty, S., Jones, B.F., Uzzi, B., 2007. The Increasing Dominance of Teams in Production of Knowledge. Science 316, 1036–1039. doi:10.1126/science.1136099

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

Figures and Tables

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

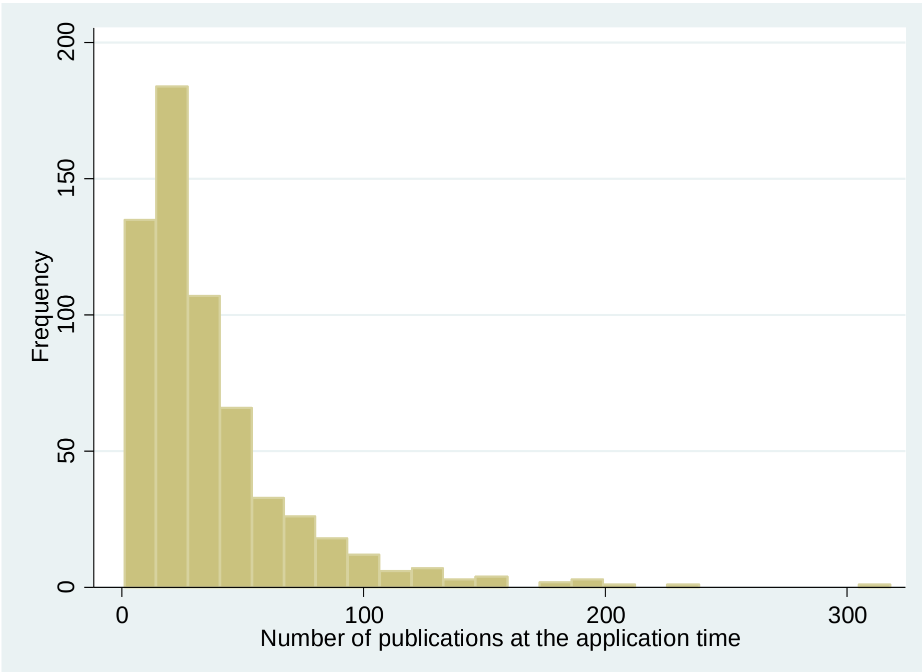


Figure 2: Distribution of grant applications by score assigned and final funding decision

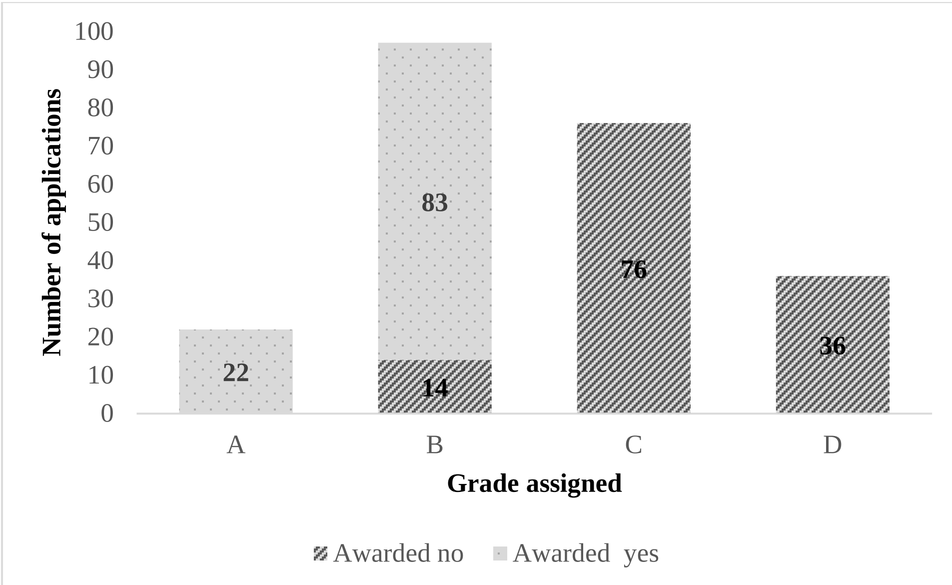


Table 1: Applicants' characteristics at the application time (number of applicants=604)

	Mean	Std. Dev.	Min	Max
Age	48.78	7.87	32	71
Female	0.15	0.36	0	1
Number of publication	42.62	39.32	1	318

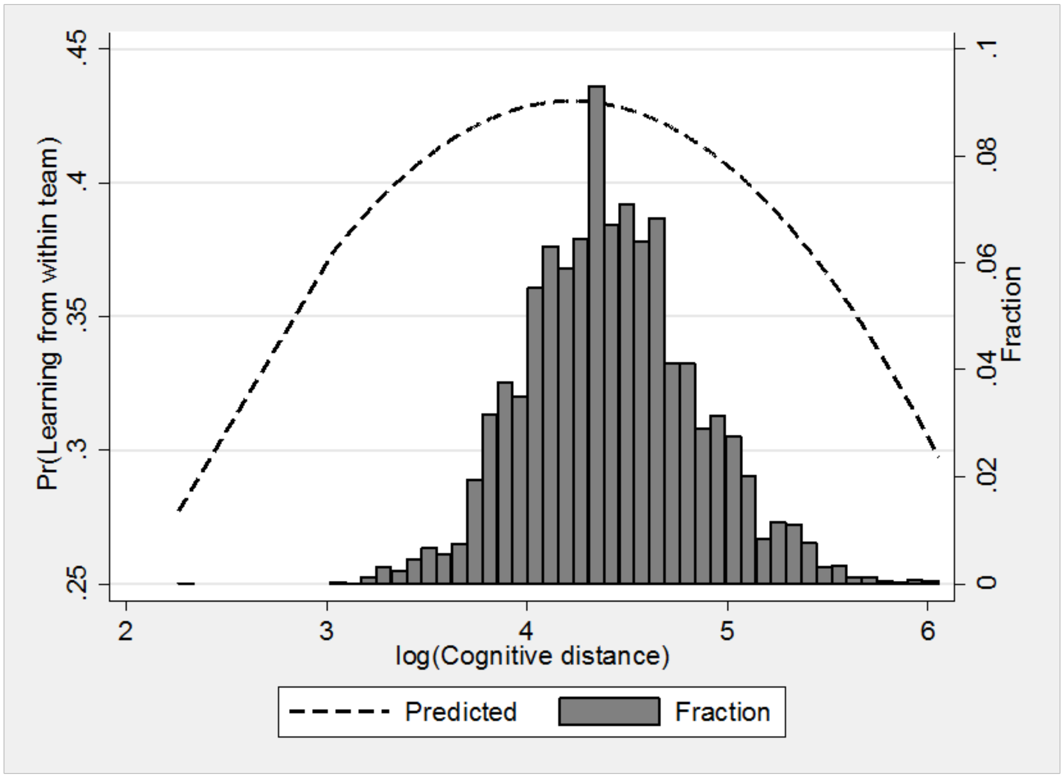
Table 2: Team characteristics at the application time (number of teams=231)

	Mean	Std. Dev	Min	Max
Established team	0.60	0.49	0	1
Team size	3.65	1.56	2	11
Number of nationalities in the team	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' publication count	51.43	32.17	8	233
Grant 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: Regression results for the probability of learning from within the team

	(1)	(2)	(3)	(4)
	co-authored	no co-auth.	co-authored	no co-auth.
VARIABLES	learning from within the team	learning from within the team	learning from within the team	learning from within the team
Individual characteristics				
Age	-0.0023	-0.00066	0.0019	0.0014
Gender (female)	-0.030	-0.054	-0.029	-0.032
log(stock of publications pre-application period)	-0.029	-0.031*	0.050***	0.052***
log(stock journals cited pre-application period)	-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, Geographical distance				
log(1+distance Hours)			0.0086	0.0089
Team characteristics, Social distance				
Established team			0.056***	0.034**
Same gender focal ind. – team			0.0043	0.017
At least one female scientist in the team			0.032**	0.029*
Standardized stock pub. difference focal ind. - team			-0.097***	-0.099***
Standardized age difference focal ind. - team			-0.031**	-0.024*
Co-ethnic team			0.037*	0.010
Team characteristics, Cognitive distance				
log(individual knowledge distance)			0.56**	0.40*
log(individual knowledge distance)^2			-0.066**	-0.048*
Other team characteristics				
Awarded			-0.0061	-0.013
Research project quality (grade A)			0.031	0.029
Research project quality (grade D)			-0.037	-0.035
log(amount requested)			0.080***	0.063***
log(n. of team members)			0.28***	0.28***
log(n. of disciplines)			0.018	0.0047
Science and Medicine			0.077***	0.072***
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**
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



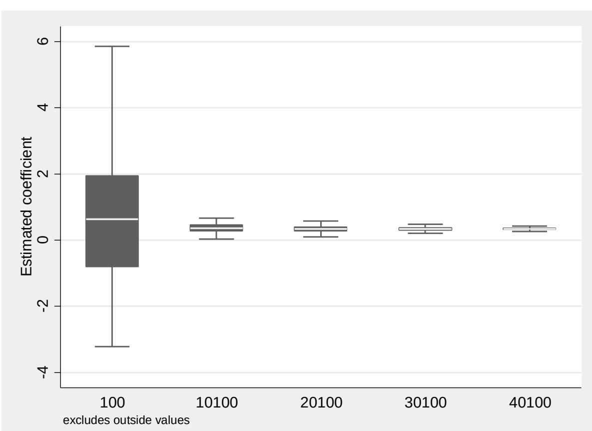
Appendix

Table A1: Regression descriptive statistics considering the study sample (52739 observations used in the regression presented in column 4 of Table 3)

VARIABLES	obs.	mean	std	min	max
dependent variable including co-publications	61068	0.39	0.48	0.00	1.00
dependent variable excluding co-publications	52739	0.36	0.48	0.00	1.00
Individual characteristics					
Age	52739	49.50	7.17	32.00	71.00
Gender (female)	52739	0.13	0.33	0.00	1.00
log(stock of publications pre-grant period)	52739	3.48	0.82	0.00	5.76
log(stock journals cited pre-grant period)	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, Geographical distance					
log(1+distance Hours)	52739	1.11	0.72	0.00	3.58
Team characteristics, Social distance					
Established team	52739	0.35	0.47	0.00	1.00
Same gender focal ind. – team	52739	0.91	0.29	0.00	1.00
At least one female scientist in the team	52739	0.35	0.48	0.00	1.00
Standardized stock pub. difference focal ind. - team	52739	-0.04	0.96	-5.62	6.35
Standardized age difference focal ind. - team	52739	-0.01	0.99	-2.83	3.13
Co-ethnic team	52739	0.17	0.37	0.00	1.00
Team characteristics, Cognitive distance					
log(individual knowledge distance)	52739	4.42	0.44	0.96	6.05
log(individual knowledge distance)^2	52739	19.73	3.92	0.93	36.62
Other team characteristics					
Awarded	52739	0.44	0.50	0.00	1.00
Research project quality (grade A)	52739	0.10	0.30	0.00	1.00
Research project quality (grade D)	52739	0.17	0.38	0.00	1.00
log(amount requested)	52739	14.34	0.42	12.77	15.74
log(n. of team members)	52739	1.34	0.37	0.69	2.40
log(n. of disciplines)	52739	1.02	0.67	0.00	2.40
Biology and Medicine	52739	0.74	0.44	0.00	1.00
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

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 3, 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 component 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 40100 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 3 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) and of the P-value (right) of the linear term of the cognitive distance for each sample size.



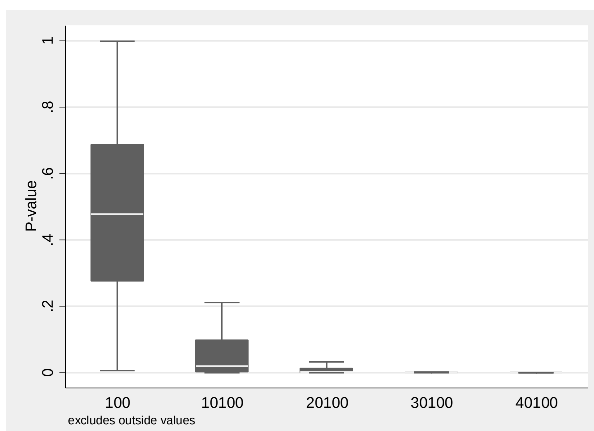
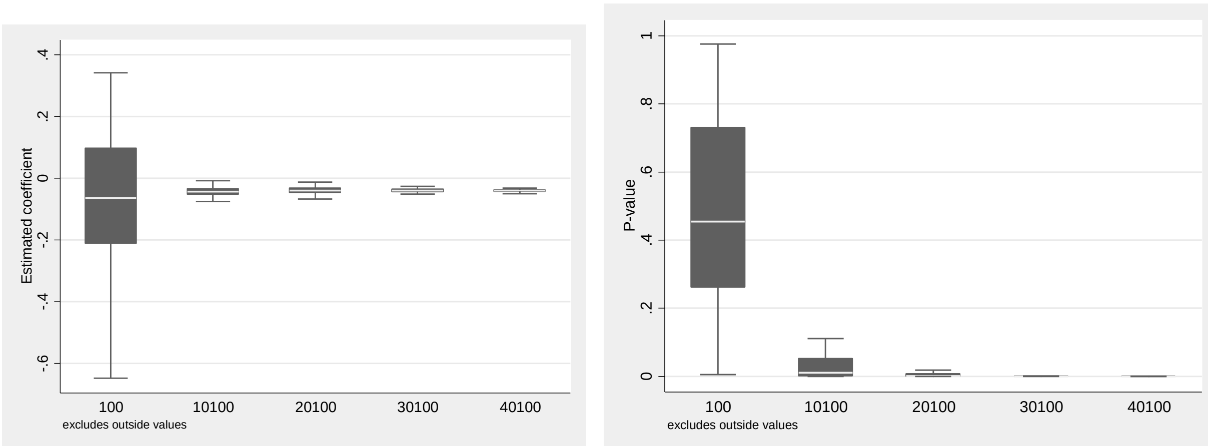


Figure Box1b: boxplots of the extent of the estimated coefficient (left) and of the P-value (right) of the quadratic term of the cognitive distance for each sample size.



DOCUMENTS DE TRAVAIL GREDEG PARUS EN 2016
GREDEG Working Papers Released in 2016

2016-01 Christian Longhi, Marcello M. Mariani & Sylvie Rochhia
Sharing and Tourism: The Rise of New Markets in Transport

2016-02 Nobuyuki Hanaki, Eizo Akiyama & Ryuichiro Ishikawa
A Methodological Note on Eliciting Price Forecasts in Asset Market Experiments

2016-03 Frédéric Marty
Les droits et libertés fondamentaux à l'épreuve de l'efficacité économique : une application à la politique de la concurrence

2016-04 Alessandra Colombelli, Jackie Krafft & Marco Vivarelli
Entrepreneurship and Innovation: New Entries, Survival, Growth

2016-05 Nobuyuki Hanaki, Angela Sutan & Marc Willinger
The Strategic Environment Effect in Beauty Contest Games

2016-06 Michael Assous, Muriel Dal Pont Legrand & Harald Hagemann
Business Cycles and Growth

2016-07 Nobuyuki Hanaki, Emily Tanimura & Nicolaas Vriend
The Principle of Minimum Differentiation Revisited: Return of the Median Voter

2016-08 Charles Ayoubi, Michele Pezzoni & Fabiana Visentin
At the Origins of Learning: Absorbing Knowledge Flows from Within or Outside the Team?