Diversity and inclusion: buzzword or real value?

Letizia Clementi, Riccardo Cavadini, Fabiola Casasopra, Marco Rabozzi, Sara Notargiacomo, Marco D. Santambrogio
Politecnico di Milano, Milan, Italy
{letizia.clementi, riccardo1.cavadini, fabiola.casasopra}@mail.polimi.it,
{marco.rabozzi, sara.notargiacomo, marco.santambrogio}@polimi.it

***Abstract*— The STEM field is characterized by a strong gender gap, both in Business and in Academia. Previous studies showed how the gender gap presents some peculiarities: women result to publish less than men across all disciplines, and this is the reason why this publication gap is often referred to as *"productivity puzzle"*. Strongly believing that gender should not influence the choice of the career to pursue, recent literature in organization has paid greater attention to gender related issues, analyzing the role played by team heterogeneity on performance. Such studies often obtained controversial outcomes, suggesting that the relationship between group heterogeneity and performance is a complex phenomenon. The dynamics taking place within working groups have been vastly studied in organizational psychology, showing that factors shaping group members' behavior are various. In this context, the working environment results to be a crucial factor. For these reasons, in this study we investigated the impact of heterogeneity on academic teams performance, taking into account gender representation in the overall working environment. More specifically, we evaluated the impact of diversity on the research conducted at the Dipartimento di Elettronica, Informazione e Biotecnologia of Politecnico di Milano, the first awarded technical university in Italy and at NECST Laboratory, a laboratory inside the Dipartimento di Elettronica, Informazione e Biotecnologia. Data are available for scientific paper published between 1965 and 2018. In this paper we studied the transformations occurred inside Dipartimento di Elettronica, Informazione e Biotecnologia in terms of gender representation between 1965 and 2018, taking into account teams characteristics, research outcomes and productivity puzzle. The results obtained showed how in both cases the impact of heterogeneity varied according to perceived value of diversity. Heterogeneity per se does not account for a boost in performance. Gender heterogeneity leads to an increase in performance only when also inclusion is achieved.**

***Keywords*—** Gender Gap, Productivity Puzzle, Performance, Diversity

I. Introduction

The presence of a strong gender gap in the Science, Technology, Engineering, Mathematics (STEM) field is a known fact: in 2017 the 82.8% of all Information and Communication Technology (ICT) specialists employed in Europe were men [1]. Statistics also reveal that, in 2015, the 73% of ICT students in Europe were males [2]. However, when considering tertiary education overall, it turns out that women are on average more educated than men: in fact, in 2017 only 38% of men aged 25-34 were tertiary-educated across the Organisation for Economic Co-operation and Development (OECD) countries [3] compared to the 50% of women. Such a gap has been widening over the past 10 year [1]. However, the labour market seems to favors men: on average across OECD countries, 80% of tertiary-educated young women are employed, compared to 89% of young men who pursued a same-level education. This gap is getting wider as years pass: in fact, while the number of ICT specialists in the European Union (EU) grew by 36.1% from 2007 to 2017, the share of female ICT specialists decreased by 5.3 percentage points [1]. In such a context, in the close future EU will face a lack of about $500,000$ employees in the ICT field [4]: increasing women representation in the STEM field would represent an important resource. Also when considering Academia, women are still underrepresented. However, while women publish less than men across all disciplines [5], quality of scientific publication does not seem to be influenced by author's gender. This phenomenon is known as the *productivity puzzle*. However, the impact of a more diverse environment in Academia is still unclear although various studies have been conducted on this topic [6] [7]. Previous studies in organizational psychology showed that heterogeneity team composition is strongly influenced by overall dynamics in terms of gender representation in work place [8] [9], and viceversa [10]. For these reasons in our analysis, while assessing the impact of heterogeneity on team's performance, we consider also gender representation rate in the working environment. We started this analysis from the assessment of the environment evolution along the years considered in the present study, which span from 1965 up to 2018. We analyzed the trend of the productivity puzzle as well, along the same time-frame. Finally, we evaluated the evolution of the impact of diversity on the academic groups' performance. Our results showed that the heterogeneity effects actually vary when the surroundings are modified, resulting in positive outcomes when the gender gap started decreasing.

This paper is structured as follows: in Section II we will illustrate the theoretical framework from which this analysis arises. In Section III we will describe the process of data collection. In Section IV we will introduce the employed Key Performance Indicators (KPIs) and motivate our choices. Section V and Section VI are respectively dedicated to the data analysis and to its discussion. Finally, conclusions are drawn in Section VII.

II. State of the art

In this Section we will provide a brief overview of previous literature on diversity and heterogeneity team performance. In Academia, which is the focus of this work, the gender gap assumes peculiar characteristics and implications, that will be described in Subsection II-A. When evaluating the impact of heterogeneity on team performance, considerations about gender representation in the external environment cannot be left aside, since they can deeply influence the behaviour of group members, as described in Subsection II-B. In Subsection II-C we analyze heterogeneity with respect to the different meanings it can assume and we describe various typologies of differences.

A. *The productivity puzzle*

As described in Section I a wide gender gap still holds in Academia: most positions are held by men, especially in the STEM field [11]. Nonetheless, in this context the gender gap is evident also in differences in the publication rate and in the impact of the published works, measured by the number of citations and the impact factor. In literature this fact is usually referred to as the *productivity puzzle* [5]. More specifically, women appear to publish less than men across all disciplines [12] [13]. However, the impact of their research works (and therefore, we could say, their value) does not reflect this gap: comparing papers to one another, their impact does not seem to change depending on the gender of the authors [5]. This kind of gap is peculiar, and results to be especially relevant when considering its effects in Academia. Indeed, in this context, productivity accounts for promotions and grants.

A first attempt to explain such a situation was made in 1984 by Cole and Zuckerman [5]: their analysis can be also considered the first formalization ever of this problem. Their study is based on the comparison between the work of 263 matched pairs of women and men to relate their performance one by one. From their analysis, women's aggregate productivity results to be the 57% of men's. Nevertheless, papers having women as authors, when considered one by one, are not actually less cited than those written by men. Various hypotheses were advanced to explain the productivity puzzle: whether the marital status may account for this gap, or if there is a bias in the review process, or if there may be a discrimination in the access to collaborations. The latter has been rejected: women appear to be as prone as men both to collaborations both to solo and primary authorship. In a later study Xie and Shauman (1998) [13] argued that such gap may be arise also from variables other than sex. They proposed further explanations for the productivity puzzle: the reasons behind this gap may be differences in personal characteristics and in structural positions. Moreover, they found out that, when looking at a more extended period of time (ranging from 1969 to 1993), the gap appears to be narrowing down. However, the reasons standing behind these social differences remain unexplained. Furthermore, while the explanations accounting for this gap and its magnitude slightly change between the previously cited studies, the fact that men and women often follow different career patterns holds.

B. The role of the "environment"

In organizational literature it is well known how the interpersonal dynamics taking place in a group may influence its outcomes [14]. For this reason, in the field of social psychology various theories have been developed to better understand the behaviour of individuals inside groups and, in this way, optimize teams' performance. Although when analyzing human behaviour a distinction is often drawn between interpersonal and intergroup dynamics one of the main findings in social psychology is the constant interweaving of these two typologies of interactions [10]. In fact, it has been observed that the presence of intergroup competition may actually lead to a higher cohesiveness and cooperation of the team itself [15]. On the other hand, according to social categorization theories, the mere perception of the existence of two distinct groups, each characterized by its peculiarities, has been found to foster intergroup hostility [16]. In such a context, the Realistic group Conflict Theory (RCT) highlights how the social *status* may play a role in these dynamics too: when prestige is perceived as a scarce resource, groups tend to polarize in dominating and subordinate ones. In other context than work groups this theory has been employed to explain ethnocentrism phenomena [10].

Other reasons why it is impossible to abstract interpersonal and intergroups dynamics are made clear by the Attraction - Selection - Attrition (ASA) framework. Such theory has been developed to explain the mechanisms through which groups are formed, and how the characteristics of a work place are actually shaped by the people living it. More specifically, the ASA frameworks states that people tend to form a group when they share personal characteristics, values and goals; on the other hand, people differing in those attributes tend to leave the organization [8], which in conclusion results to be shaped by its members peculiarities. Given the importance of similarities highlighted by the ASA framework, it is easy to understand how the impact of heterogeneity on the quality of work can be controversial. The results of analysis conducted in organizational literature on this topic in fact have sometimes been contradictory: while most studies have found a positive correlation between diversity and performance, sometimes the outcome of a more diverse team resulted to be worse [17]. As highlighted by Cox and Blake (1991) [14] diversity undoubtedly represents a challenge: it can be a competitive advantage by fostering creativity, innovation and flexibility, but in order to obtain such positive impact, it has to be managed.

C. Heterogeneity impact and typologies

Heterogeneity is a qualitative and complex concept, which is often employed in describing situations which may deeply differ from one another. More specifically, the characteristics among which team members may diversify are numerous, each playing a different role in team dynamics and accounting for various outcomes on the performance [18]. We can distinguish between: ***separation***, ***disparity***, and ***variety***. We are talking about *separation* when team members have diverging opinions, beliefs or values from each other. The impact of this property has been explained through the "Similarity - Attraction" [18] and the ASA [8] frameworks: according to such theories, people tend to bond with individuals whom they consider similar to themselves. Moreover, when the separating attribute is central to the task that has to be performed by the team, such differences may create a conflict within the group. *Disparity*, which is often referred to also as "inequality", implies a difference in socially evaluated assets such as prestige, income or decision-making authority inside the team. Relatively rare studies focused their attention toward the evaluation of disparity's outcomes: however, it seems evident that such a characteristic deters collaboration within team-members, while enhancing endogenous competition. Finally, talking about *variety*, it is defined as differences in categorical attributes which are not socially evaluated as more or less prestigious, such as functional background, ethnicity, or gender. Previous studies that dealt with the effects of variety considered each team as an information processing unit. In such a perspective the *Law of Requisite Variety* [19], coming from cybernetics, has been employed to explain its influence on team work. The presence of multiple categories inside a group broadens its cognitive and behavioural repertoire. Coherently with this interpretation, subsequent studies discovered that a more diverse group may be at the basis of more creative solutions [18].

In this analysis we focused our attention on the impact of variety on research teams performance. We analyzed scientific publications between 1985 and 2018 of scientists working at DEIB department of Politecnico di Milano, a first ranked technical university in Italy. We investigated the presence of a gender gap in scientific publication and the existence of a productivity puzzle. belonging to our University: given the importance of the background resulting from the cited studies we also evaluated the evolution of the environment they worked in, and the presence of a productivity puzzle to evaluate a possible correlation between such phenomena and the impact of diversity on team performance.

III. Data Collection

In this study, we are proposing an analysis of the impact of academic team heterogeneity on performance and evaluating the presence of a gender-gap in the context of the Dipartimento di Elettronica Informazione e Biotecnologia (DEIB) and NECST Laboratory at Politecnico di Milano. In this Section we will start by describing the scenario of our analysis, then we will illustrate how we collected the analyzed data, whereas in Section IV we will introduce the main indicators used for the analysis.

A. Scenarios description

In this analysis we evaluated two different environments to give wider support to our findings. Such environments are the DEIB Department and the NECST Laboratory at Politecnico di Milano. Both are focused on STEM research: more specifically DEIB comprises the research areas of Computer Science, Bioengineering and Electronics, while NECST Laboratory is mainly focused on the first two topics. More importantly, NECSTLab has always been characterized by a high rate of heterogeneity due to various factors, reason why we decided to evaluate it as a separate scenario. Firstly, one of its peculiarities is the mixture of teaching and research, which leads to differences in the educational level of people collaborating on the same project; secondly, there is a high background heterogeneity: most project are carried out by a collaboration between Biomedical and Computer Science Engineers; finally, there is a high gender heterogeneity, also due to the high number of Biomedical Engineering students, which are more often women. The data regarding both scenarios are collected following very similar procedures: we will start by describing the more complex process followed to gather data regarding the Department, and we will later describe how it differs from the one implemented to acquire information regarding the Laboratory.

B. Data retrieval procedure

The vast majority of data analyzed in this study are collected from Scopus [20]: the only exceptions are represented by the impact factor - for which we made reference to the "Scimago Journal & Country Rank "website [21] - and our Laboratory's researchers background, which has been manually collected. We decided to employ Scopus to collect the authors namefor two main reasons: firstly, it is the most complete database available; secondly, it provides some useful API to acquire information about an author or a paper given their identifier. The whole process was implemented through a PHP script and can be broken down in three sub-process:

- Name collection: we gathered all the names of the researchers working at DEIB Department;
- Publication list and co-authors retrieval;
- Inference of genders for all authors and co-authors.

Name collection. In this phase we collected the researchers' first and last names, which will be later used to acquire the publication list through the Scopus' Application Programming Interfaces (APIs). The list of the researchers that are, or have been, affiliated to DEIB can be found on DEIB website, sorted by role. Considered the goal of our analysis, we decided to include in the names' list only the categories of our interest, which are: full professors, associate professors, assistant professors, PhD students, research assistants, contract professors, research collaborators and *emeriti* professors.

Publication list and co-authors retrieval. Given the first and last names of the researchers affiliated to DEIB, we then acquired their list of publications through Scoupus. Firstly, we retrieved the Scopus identifier of each researcher. Nonetheless, some researchers do not have a unique identifier. This may be due to homonimy and to the fact that some scientists actually have more than one Scopus profile. Since it is impossible for us to solve this ambiguity, to avoid losing data we collected all the identifiers for each single first/last name pair. Given the author identifiers, we acquired their publications list and, for each published work we retrieved:

- the co-authors list, with first and last name;
- the number of citations;
- date and venue of publications.

Gender Inference The third and last phase regards the inference of each author's gender based on their name. Since the authors' gender is not disclosed on peer-reviewed publications inferring it from the name is the usual approach for large-scale studies [22] [23]. In order to do so, we employed the Genderize.io API [24]. The results of a prediction can be *male*, *female* or *unknown*, always followed by the associated accuracy.

When collecting data regarding NECST Laboratory instead, we used the publication list of NECST Laboratory director, which comprehends all the papers published at the laboratory. For this reason, in this case, we only needed to retrieve from Scopus the number of citations alongside the publication's date and venue. Then, we inferred the gender through Genderize.io as previously described.

C. Issues

In this brief Subsection we will discuss the most relevant issues we encountered while performing the previously explained tasks.

Information on Scopus is not always complete. This problem concerns each piece of information, which might have been missing or incomplete. For example, the authors' affiliation and first name is often absent affecting the possibility to determine the gender, and thus negatively impacting on the overall dataset accuracy.

Not all first names are gender specific. It is well known that some given names apply for both women and men, such as Ashley, Kim, Riley, Lee or Claude. Moreover, as highlighted in a Science-Metrix report about bibliometric indicators (specific for the measure of women contribution to science) [25], one serious problem arises when inferring the gender from Asian names. In such scenarios, the discriminating power of given names drops significantly: various Chinese first names are as common for women as they are for men, this ambiguity is less significant when names are written in Chinese ideogram, but this valuable information is lost when Chinese names are romanized [22].

D. Data characteristics and pre-processing

At the end of the process of data collection we obtained two datasets: one relative to DEIB, one to NECST Laboratory. The first initially featured $36,786$ entries regarding papers published between 1965 and May 2018, while the second is much smaller (190 entries) and refers to a narrower time horizon: it dates back to the foundation of our NECST Laboratory, in 2005.

Due to the above illustrated issues we decided to pre-process the Department dataset before proceeding with the analysis. From these initial records we removed:

- Entries with members whose gender was unknown, to improve accuracy. Such entries represented the 13.76% of all authors, affecting the 21.37% of records;
- Double entries: various papers appeared twice, or more, in the dataset, representing the 42.40% of the records. This is due to the already cited issues of duplicate profiles on Scopus;
- Papers published by teams composed by more than 30 members: in this case it would be unrealistic to assume group of such dimension to undergo the same dynamics as smaller ones;
- Entries missing fundamental information, such as the number of citations.

At the end of this process although we obtained a much smaller dataset - including $16,096$ entries - we observed the mean accuracy of the gender inference to increase from 0.929 to 0.978

IV. Key Performance Indicators

The aim of this analysis is to quantitatively evaluate the relationship between heterogeneity and performance in academic teams and to determine if a gender-based difference in performance is present at the level of DEIB and NECST Laboratory. Both diversity and performance are qualitative notions, thus, their mathematical formalization is of extreme importance for the outcome of our analysis. In the following Subsections we will illustrate the KPIs employed in this study and motivate our choices.

A. Heterogeneity indicators

Firstly, we must accurately define what we actually mean when we speak of heterogeneity and performance. As described in Section II the differences from which heterogeneity may arise can be various, each having a different impact on the dynamics taking place in a group [26]. Such typologies of heterogeneity, due to the different situations they reflect, need to be formalized differently. Moreover, the team composition characterized by the highest amount of diversity varies according to the heterogeneity type we are considering [18]. In this study our main focus is the heterogeneity in terms of gender differences, which are attributable to variety, as described in Section II. When we evaluate variety, we can ascribe team members to different categories, each represented by a gender, in our case. For example, when evaluating variety due to ethnicity instead, each category will represent an ethnic group. When assessing variety we assume to face the maximum heterogeneity when all categories are equally represented in the team [18]. Oppositely, a team is perfectly homogeneous when all team members belong to the same category.

The indicator most often employed to evaluate variety is the Blau's index (1), alongside its normalization, the IQV index [18] [27]. In fact, the IQV index ranges between 0 and 1, which represent, respectively, the most homogeneous situation and the most heterogeneous one. It is defined as:

$$B = 1 - \sum_{i=1}^{N} {p_i}^2 \tag{1}$$

N = number of categories
p_i = share of members belonging to each category
In this study the N categories are represented by genders, thus N is equal to 2.

Previous studies also employed the Information Entropy as heterogeneity measure [18]. The entropy concept in the information field [28] may be interpreted as the amount of information given by the realization of a specific event, a quantity inversely proportional to its probability to occur. This indicator is defined as follows:

$$E = -\sum_{i=1}^{N} p_i \cdot \ln p_i \tag{2}$$

N = number of categories
p_i = share of members belonging to each category
Again, in our case N is equal to 2.

The entropy index is null when the group is perfectly homogeneous, with all the members belonging to the same category. When, instead, all the categories are equally represented it reaches a value of $-1 \cdot \ln \frac{1}{N}$: the higher the entropy, the higher the information content and the diversity. As stated in Section II, the impact of variety on groups dynamics has been explained through the Law of Requisite Variety [19], which considers team members as information carrier. Thus, this index describes heterogeneity through the amount of information it carries with itself. This peculiar point of view is coherent to the explanation of heterogeneity impact given in the State of Art. For these reasons, the main diversity indicator employed in this analysis is the Information Entropy, nonetheless we also used the IQV index to intuitively describe the amount of heterogeneity present in a team.

B. Performance indicators

The measurement of performance is central in this work, since it represents the metric to evaluate both the presence of a gap and the quality of a research work. In this analysis we assessed the presence of a gender gap through differences in performance. As previous studies showed [5], the gender gap in Academia is often revealed by differences in publication rates and in the impact of published papers. Gender gap might vary when evaluating the number of citations and the impact factor [12]. Moreover, the evaluation of a productivity puzzle requires a distinction between the output and the impact of a scientist's work [13]. Here we provide a definition of output and impact as interpreted all along this paper.

- **output**: number of published papers - *quantitative* index;
- **impact**: appreciation received by a work, hence this index involves the *quality* of a publication.

Previous studies trying to evaluate the presence of a gender gap in Academia with respect to quality and quantity [27] of work and how gender may influence differently these two measures of performance, finding that women and men are characterized by different patterns in productivity [29]. In the present study, due to the specific characteristics of the dataset (as explained in Section III), we did not evaluate the productivity of the single author: we made instead reference to aggregate data and evaluate the trend along the 53 considered years, hence considering the impact of the work as a performance indicator of the whole team. The quality of work was esteemed based on two KPIs: the ***impact factor***, as indicated by the "Scimago Journal & Country Rank" website [21], and the ***number of citations per month***. This first metric is based on the h-index: this indicator, firstly developed to evaluate a scientist's work, when referred to a Journal or to a Conference indicates the number h of articles having at least h citations each published in that Journal or Conference. The data employed by Scimago to rank venues are retrieved by Scopus. However, this data only date back up to 1996. Such time frame is much narrower than the one covered by the dataset referring to DEIB: the impact factor reached by a Conference in the early 2000s would be inappropriate to describe the performance of a paper published in the same venue but 35 years before and would distort the evaluation. Hence we only employed such metric to evaluate NECST Laboratory teams performance, not DEIB one.

At this point we must draw a distinction between the KPIs used to measure the performance of DEIB and NECST Laboratory scientific works. For what concerns DEIB papers evaluation, after discarding the impact factor as a suitable performance indicator, we considered employing the number of citations as performance index. However, considering such a wide time frame, this metric would have favored older papers. Thus we divided the number of citations by the number of months passed since the paper publication. We considered months instead of years to include in our analysis also those works published during the course of 2018. Oppositely, when speaking of NECST Laboratory, this metric appears to be unsuitable: although conceptually appropriate, we observed its variability to be extremely low. In fact, due to the mixture of teaching and research present at NECST Laboratory, often the papers published are the first research work of students: for this reason the number of citations is usually low. In this context the rating of the Conference, or Journal, in which they accomplish to publish represents a more reliable indicator to describe the quality of a paper.

C. Indicators summary

To summarize, in our analysis we focused on assessing the existence of gender based differences and the impact of team heterogeneity on scientific publication performance. To do so we employed the following parameters:

- Heterogeneity indicators:
  - The Blau's Index (1) alongside its normalization (the IQV index);
  - The Information Entropy (2);
- Performance indicators:
  - The number of citations per month, for DEIB;
  - The impact factor, for NECST Laboratory.

TABLE I
COMPARISON BETWEEN MALE AND FEMALE FIRST AUTHORSHIP

Gender	First author	Percentage over total
Male	12499	78 %
Female	3597	22 %

TABLE II
COMPOSITION OF RESEARCH TEAM BASED ON GENDER

Team composition	Number of teams	Percentage over total
Males only	7784	48.35 %
Females only	261	1.62 %
Mixed	8052	50,02 %

V. ANALYSIS

The aim of this study is on one hand to assess the presence of a gender gap at DEIB, and on the other hand to evaluate the existence of a relation between team diversity and performance. In this Section we will describe and briefly comment the results obtained in our analysis, while a more comprehensive explanation will be given in Section VI. In Subsection V-A we provide a brief overview of our sample and main evidences found. In Subsection V-B we describe the main changes which took place at DEIB between 1965 and 2018, while in Subsection V-C we illustrate the results of the gender gap assessment conducted at DEIB. Subsection V-D is dedicated to illustrate main findings about the relation between diversity and performance, both at DEIB and NECST Laboratory.

A. Data overview

Our dataset is composed of 16,097 papers published between 1965 and 2018 by scientists affiliated to DEIB. As presented in Table I, the 78% of publications were published by a men as a first author, compared to only the 22% of women.

For what concerns team composition, as presented in Table II, the 50% of teams were heterogeneous team. When analyzing homogeneous teams, we can see that groups composed of only women represents a short portion of the sample, only the 1.62%.

Finally, Table III illustrates main DEIB research areas in terms of number of papers published and number of female and male authors. Women seem to be mostly represented in Bioengineering and Computer Science research areas. However, if we compare the numbers of female and male authors in both areas, we can see that in Bioengineering and Computer Science women represents respectively only the 48% and the 34% of total. Women are still underrepresented in fields such as Telecommunications, Systems and Controls, Electronics and Electrical Engineering.

B. "Environment" evolution

The temporal span we are taking into account in this analysis ranges from 1965 up to 2018: in this years our society underwent many changes, and so did the technologies employed in the academic work. Those changes have had a strong impact on how scientists approach scientific research. First, we investigated changes in team composition: the average team composition underwent significant changes. In particular, their dimension has been growing between 1965 and 2018. In the first examined decade the mean team size was 2.83, whereas in 2018 reached 5.40 members. The mean performance of published papers measured in terms of monthly citations has been growing as well. More specifically, the mean number of monthly citations in the '60s was 0.023, while in the last decade reached the 0.138 citations per month. A graphical representation of this trend can be found in Figure 1.

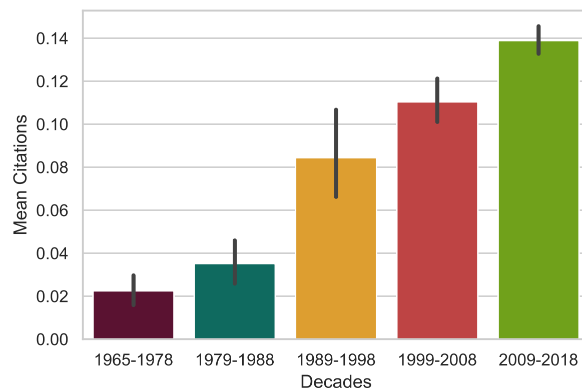

Fig. 1. Monthly citations trend over decades.

It is possible to explain this fact by thinking of the technological advancement taking place in such time frame. Past literature is nowadays widely accessible to more and more people thanks to the new technologies, and everybody has a higher possibility to share and publish their studies. These are two important factors in the observable increment of this KPI. We also evaluated the trend of the heterogeneity indices all along the decades we took into account. As shown in Figure 2, the amount of diversity has been growing during the time period considered in our analysis. Such a growth in the overall heterogeneity is related to an increasing presence of women in Academia during the last decades. Today the total number of authors is 464.6 times what it was back then. When distinguishing between males and females, we can notice that

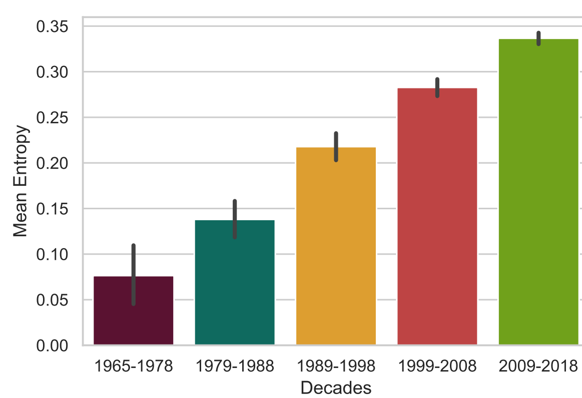

Fig. 2. Heterogeneity trend over decades

TABLE III
NUMBER OF MEN AND WOMEN AUTHORS PER RESEARCH AREA

Research area	Papers	Female authors	Male authors
Bioengineering	2273 (14.12%)	4099	8479
Telecomunications	2762 (17.16%)	1225	8605
Systems and Controls	2456 (15.26%)	1237	6592
Computer science and Eng.	5049 (31.37%)	4608	13475
Electronics	1852 (11.51%)	1739	7772
Electrical engineering	1426 (8.86%)	761	4044

while the male population grew by a factor of 3.7, the growth in the female population is much broader. In fact, nowadays, the female author population at DEIB is 24.67 times what it was in the '60s. This growth also reflects into the average percentage of women in teams that we can observe over years and that is represented in Figure 3. In fact, while in the first considered decade the mean

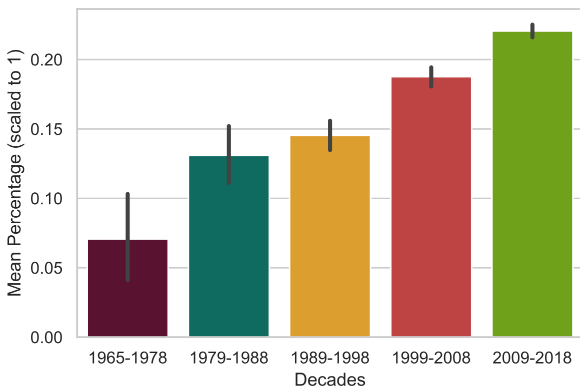

Fig. 3. Percentage of female authors trend over decades

percentage of female authors per team was of 7.07%, nowadays it reached the 22.15%. However, a strong gap is still present: almost the 80% of authors affiliated at DEIB are still men.

C. The productivity puzzle

To evaluate the presence of a productivity puzzle at DEIB, we started by comparing the proclivity of men and women to first and solo authorship. We had a dual reason for investigating that: first, such positions are the most prestigious in Academia; second, women's proclivity to publish as first or solo authors has been studied in literature to explain the presence of a productivity puzzle [5] [12].

We started our evaluation by considering the solo-authorship. The global number of solo-authored paper has been growing over the last few years due to the rise in the number of overall published works. Actually, accordingly to SubsectionV-B, the solo-authorship tendency has been declining in terms of percentage of work done: while in the '60s they represented the 13% of all published works, in the last decades they accounted only for the 3.7% of publications. In such a context, however, the number of publications solo-authored by women has been growing. Moreover, it is of great relevance to notice that the first solo-authored work done by a woman was published only after 1998. This tendency may be explained through a narrowing of the gender gap, especially when considering that the opposite trend is found when looking at the male population. We can therefore state that our data do not show a lower proclivity of women to solo-author their work. We observed an important change during the considered time frame for what concerns primary authorship as well. While in the '60s the totality of the published work featured a man as first-author, in the last decade this percentage drops to the 80.97%. This percentage matches the distribution of authors between genders: again, women's proclivity to primary authorship does not seem to account for differences in productivity.

Nevertheless, when investigating the relationship between performance and women primary and solo-authorship, the situation appears to be more complex. Papers featuring women as first authors seem to perform worse than those featuring men. Women's performance results to be 0.8 times the men's one, although showing an increasing trend. We can conclude that, although women are getting more and more represented in Academia and the gender gap is narrowing down, the parity in work acknowledgement is yet to be reached.

D. Diversity and performance

Dipartimento di Elettronica, Informazione e Biotecnologia We started this phase of our analysis by evaluating the aggregate data relative to DEIB. In this step we did not distinguish papers based on their publication date, but considered all the works together. What emerged in the beginning was puzzling: homogeneous groups resulted to have a higher average level of performance. Given the vast transformations observed during the period we are analyzing, as described in Subsection V-B, we started assessing each decade separately. Firstly, we evaluated the correlation between heterogeneity (measured through entropy 2) and performance (measured through number of monthly citations). Results are showed in Table IV.

TABLE IV
COEFFICIENT OF CORRELATION BETWEEN HETEROGENEITY AND PERFORMANCE OVER YEARS

Years	Correlation coefficient
1965 - 1978	-0.05786
1979 - 1988	0.10917
1989 - 1998	-0.02647
1999 - 2008	0.01233
2009 - 2018	0.00769

Although these coefficients do not show a strong correlation between the two considered variables we can observe an inversion of tendency along the considered time-frame. At this point, we decided to evaluate the distribution of data. This was done through the scatterplots shown in Figure 4 and 5, which refer respectively to the first and to the last decades we are considering. The outer barplots, instead, show the number of observation per value of heterogeneity. When looking at the data distribution we can notice

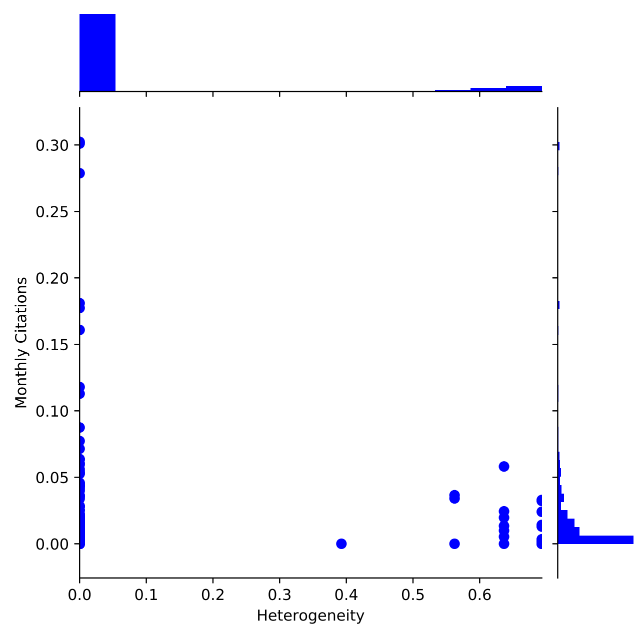

Fig. 4. Scatterplot relating heterogeneity and performance between 1965 and 1978 at DEIB. The outer barplots display the data distribution.

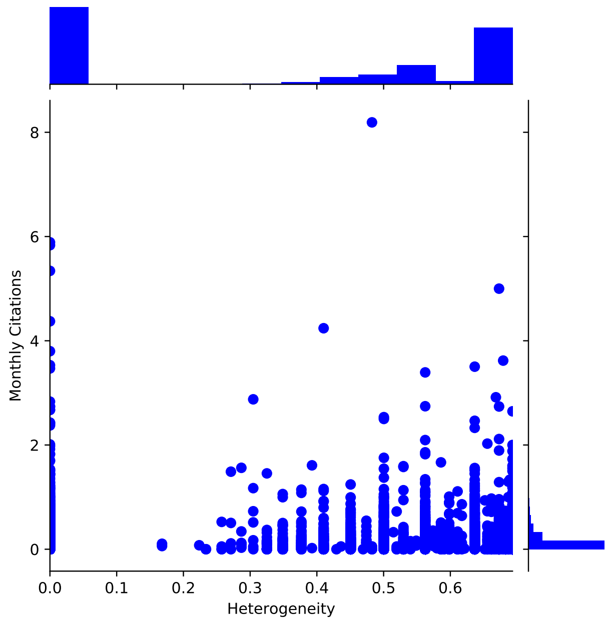

Fig. 5. Scatterplot relating heterogeneity and performance between 2009 and 2018 at DEIB. The outer barplots display the data distribution.

how the relation between heterogeneity and performance underwent a change. More specifically it becomes evident that in the earlier years the most performing teams were the most homogeneous ones, while this trend is not confirmed in later years. In fact, although homogeneous groups keep performing well, the more heterogeneous ones' performance has strongly improved. The lowest number of citations is instead reached when the diversity amount is low. Such a situation matches the explanation given in literature for the diversity impact on performance: although heterogeneity may have a positive impact thanks to the wider pool of information it makes available, when differences do not reach a balance its outcomes may be negative.

NECST Laboratory For what concerns NECST Laboratory, only data of the last 13 years are available. Nonetheless, it is possible to retrieve some information. The scatterplots relating to the periods 2005-2006 and 2017-2018 are shown respectively in Figure 6 and 7. The obtained results are coherent to those shown above, observable from the other dataset. Therefore, we can state that there is a trend in the STEM field at DEIB department and that heterogeneous teams have been bringing to value research. Nonetheless, this data also show that more homogeneous teams have been increasing their performance over the last few years, compared to the more hetero-geneous ones. Although their results remain worse, this evidence must be taken into account and discussed. Actually, such evidence is not in contrast to the previous ones: a possible explanation to this fact could be that, over the last few years the Laboratory has been undergoing a change in its composition: an increasing number of Biomedical Engineering students has been joining the Laboratory, introducing an additional heterogeneity source in the environment. We also performed some background analysis on all the people that took part to published works within the Laboratory, but since the vast majority are still Computer Science Engineers, the validity of this analysis is limited.

From the results shown also emerges that gender diversity is not sufficient to evaluate the expected performance of a team: even if a clear trend appears from our data, we were unable to individuate a causal relationship linking heterogeneity and performance. It is

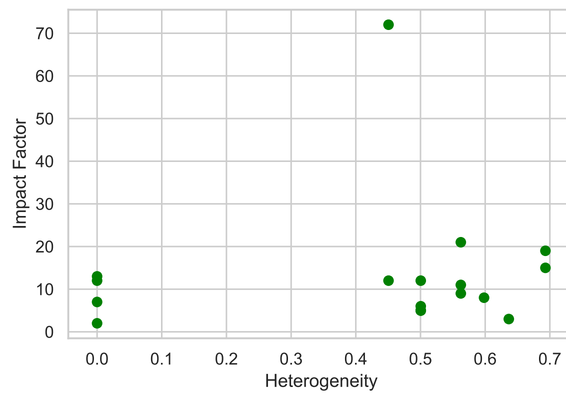

Fig. 6. Scatterplot relating heterogeneity and performance between 2005 and 2006 at NECST Laboratory.

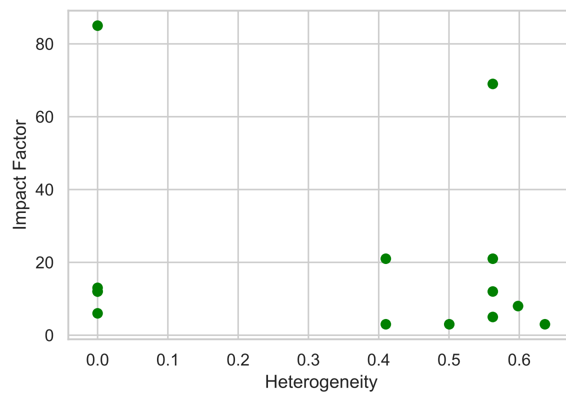

Fig. 7. Scatterplot relating heterogeneity and performance between 2017 and 2018 at NECST Laboratory.

our intention to conduct further analysis to consider other elements of diversity. More specifically we would like to expand the present study by including in the analysis other factors such as cultural background, level of education and ethnicity in order to better assess the impact of heterogeneity on intragroup dynamics.

VI. Discussion and Lessons Learned

As described in Section II, the impact of heterogeneity has been widely investigated in previous literature, sometimes with contrasting outcomes [30] [31] . The mechanisms through which group dynamics arise have been vastly studied as well [27]. The discussion is still lively, but from this context it emerges that the environment has a very important influence on the work quality. For this reason, we tried to understand if there is a link between the impact of diversity on the performance and the characteristics of the environment in which groups work. Also, in our research, the period in which a paper was published is of fundamental importance: this is due to the profound social changes that took place during the observed period, which inevitably impacted on the work environment that we are analyzing.

From our data, one of the first facts that can be noticed is that the KPIs we considered vastly changed along the considered time frame: the number of paper published each year grew enormously, passing from 63 in the first analyzed decades to $14,340$ in the last one; so did the number of people accomplishing to get their work published. Due to a higher accessibility to literature, the mean number of citations increased as well. The complexity of research work increased too: in fact, it is getting much rarer to find solo-authored papers, and the mean size of teams grew, going from 2.80 to 5.40 members. In such a context, DEIB Department experienced a significant growth in women representation. The percentage of female authors among the whole number of people who published papers in the considered period of time went from the 3.7% to the 20.8% and, as stated in Section V, their growth outperformed by far the men's one. Hence, from a merely numerical point of view, the gender gap is getting narrower as years go by. However, this is not enough to state that we are actually getting closer to a situation of equality: the productivity puzzle is a much more complex phenomenon, which involves many other factors, such as the opportunity to get recognition for one's work and to be included in teams, both as co- and first-author. To investigate the presence of a productivity puzzle, we decided to examine the trend of first and solo-authorship. The results showed that, while solo-authorship is a less and less adopted practice, the number of paper solo-authored by women grew: until 1998 there were not solo-authored paper by women, while there were 24 in the last analyzed decades. At the same time solo-authorship by men declined in terms of percentage of total published works. On the other hand, while in previous decades, women were underrepresented as first authors, the number of papers first-authored by women grew during the considered time period, until it reached, in the last decade, the number of female authors affiliated to DEIB Department.

The described data clearly shows how the gender gap, in merely numerical terms, is attenuating. Moreover more prestigious positions as authors are becoming more accessible to women. Nevertheless, we have to take into account the performance of their work. When looking at such indicators we notice how both as first authors and as solo, and also when part of an homogeneous team, women are outperformed by men for what concerns the number of monthly citations. More specifically, the mean women's performance results to be 0.8 times the male's one. Hence, we can state that despite the gender gap reduction, a productivity puzzle still holds. This fact must be taken into account when looking at the impact of heterogeneity on team performance, which in the last 53 years has undergone a clear shift. In fact, while in the first analyzed decades the best performing teams were the homogeneous ones, this trend has been inverting as years passed by. In fact, the best results observed were obtained by teams with a medium level of heterogeneity, characterized by an IQV value of 0.7. When considering NECST Laboratory, we must remember that we only had information about the last 13 years. Therefore, we can compare the results of the two environments only for the overlapping period of time. The average results of more heterogeneous teams in the NECST Laboratory were higher than those of fully homogeneous ones. However, this is not true when looking at the results from the whole Department, in which homogeneity still seems to prevail in terms of productivity. This is an interesting difference, which we tried to explain with the information to us available. The DEIB has slowly been getting more heterogeneous, while the NECST Laboratory has always been so, even if in the last few years its heterogeneity level has been increasing due to an increase in the background mixture. It is possible that, as happened at DEIB Department in the 60s, gender differences did not reach a balance yet.

Thus, from our analysis it arises that, while the gender gap - in all its forms - has been narrowing, the impact of heterogeneity has been getting more and more positive as years went by. A possible explanation to this fact can be found in the distinction between variety and disparity. As explained in Section II, when team members differ in socially evaluated assets, the group faces diversity in terms of disparity, which has been linked to negative impact on performance. Generally speaking, when one of the categories present in the team is perceived as socially inferior, the outcome of diversity on the quality of work will not be positive. Thus, heterogeneity *per se* does not account for a boost in performance: in order to experience such positive dynamics a group needs something more, which is inclusion.

VII. Conclusions

In this work we evaluated the impact of heterogeneity on team performance based on gender dynamics in the working environment. To do so, we assessed the social transformation our University underwent during the period taken into account in this study. We found out that some important changes took place. Firstly, during the last years the productivity puzzle significantly decreased. Secondly, the impact of heterogeneity on team performance changed, becoming more and more positive. This fact can be explained through the theoretical framework presented in Section II: how diversity is perceived by team members is crucial for its impact on intragroup dynamics. The negative impact of diversity that was overall faced in the last century may be related to the perception of gender diversity as disparity, which may lead to endogenous competition and to worse performance. Nonetheless, given the results obtained when analyzing later years, we can confirm that diversity may provide an additional value to a team by increasing available knowledge, problem solving capabilities and by stimulating creativity and we can thus conclude that inclusion is what makes the difference: only when all genders are perceived as equal, the real value of diversity becomes evident.

However, as our results showed, gender heterogeneity is not sufficient to fully assess the performance achieved by teams. As a future work we would like to evaluate the possibility to include other variables in our analysis. More specifically we believe that other typologies of diversity, such as differences in the background, in ethnicity and in level of education, should be considered and assessed to reach a more complete understanding of what is the heterogeneity impact on teams.

References

[1] "Eurostat - ICT specialists in employment," https://ec.europa.eu/eurostat/statistics-explained/index.php/ICT_specialists_in_employment, Accessed: 2018 - 10 - 30.

[2] "Eurostat - Girls and women under-represented in ICT," https://ec.europa.eu/eurostat/web/products-eurostat-news/-/EDN-20170426-1, Accessed: 2018 - 10 - 30.

[3] "OECD Countries List," http://www.oecd.org/about/membersandpartners/list-oecd-member-countries.htm, Accessed: 2018 - 10 - 30.

[4] E. European Commission Directorate General for Internal Market, Industry and SMEs, "High-tech leadership skills for europe - towards an agenda for 2020 and beyond," March, 2017.

[5] J. R. Cole and H. Zuckerman, "The productivity puzzle," *Advances in Motivation and Achievement. Women in Science. JAI Press, Greenwich, CT*, 1984.

[6] L. G. Campbell, S. Mehtani, M. E. Dozier, and J. Rinehart, "Gender-heterogeneous working groups produce higher quality science," *PloS one*, vol. 8, no. 10, p. e79147, 2013.

[7] J. B. Bear and A. W. Woolley, "The role of gender in team collaboration and performance," *Interdisciplinary science reviews*, vol. 36, no. 2, pp. 146–153, 2011.

[8] B. Schneider, "The people make the place," *Personnel psychology*, vol. 40, no. 3, pp. 437–453, 1987.

[9] A. Drach-Zahavy and A. Somech, "Team heterogeneity and its relationship with team support and team effectiveness," *Journal of Educational Administration*, vol. 40, no. 1, pp. 44–66, 2002.

[10] H. Tajfel and J. C. Turner, "An integrative theory of intergroup conflict," *The social psychology of intergroup relations*, vol. 33, no. 47, p. 74, 1979.

[11] "Eurostat - Tertiary education statistics," https://ec.europa.eu/eurostat/statistics-explained/index.php/Tertiary_education_statistics, Accessed: 2018 - 10 - 30.

[12] L. S. Rigg, S. McCarragher, and A. Krmenec, “Authorship, collaboration, and gender: Fifteen years of publication productivity in selected geography journals,” *The Professional Geographer*, vol. 64, no. 4, pp. 491–502, 2012.

[13] Y. Xie and K. A. Shauman, “Sex differences in research productivity: New evidence about an old puzzle,” *American Sociological Review*, pp. 847–870, 1998.

[14] T. H. Cox and S. Blake, “Managing cultural diversity: Implications for organizational competitiveness,” *Academy of Management Perspectives*, vol. 5, no. 3, pp. 45–56, 1991.

[15] F. E. Fiedler, “The effect of inter-group competition on group member adjustment,” ILLINOIS UNIV URBANA, Tech. Rep., 1967.

[16] J. C. Turner, I. Sachdev, and M. A. Hogg, “Social categorization, interpersonal attraction and group formation,” *British Journal of Social Psychology*, vol. 22, no. 3, pp. 227–239, 1983.

[17] B. Frick, A. Rose, A. Kolle *et al.*, “Gender diversity is detrimental to team performance: Evidence from a field experiment,” Paderborn University, Faculty of Business Administration and Economics, Tech. Rep., 2017.

[18] D. A. Harrison and K. J. Klein, “What’s the difference? diversity constructs as separation, variety, or disparity in organizations,” *Academy of management review*, vol. 32, no. 4, pp. 1199–1228, 2007.

[19] W. R. Ashby, “Requisite variety and its implications for the control of complex systems,” in *Facets of systems science*. Springer, 1991, pp. 405–417.

[20] “Scopus,” https://www.scopus.com/, Accessed: 2018 - 12 - 11.

[21] “Scimago Journal & Country Rank,” https://www.scimagojr.com, Accessed: 2018 - 10 - 30.

[22] S. Mattauch, K. Lohmann, F. Hannig, D. Lohmann, and J. Teich, “The Gender Gap in Computer Science A Bibliometric Analysis,” 2018.

[23] L. Santamaría and H. Mihaljević, “Comparison and benchmark of name-to-gender inference services,” *PeerJ Computer Science*, vol. 4, p. e156, 2018.

[24] “Genderize.io,” https://genderize.io/, Accessed: 2019 - 01 - 01.

[25] S.-M. Inc., “Analytical support for bibliometrics indicators - development of bibliometric indicators to measure womens contribution to scientific publications,” Tech. Rep., 2018.

[26] A. Solanas Pérez, R. M. Selvam, J. Navarro Cid, and D. Leiva Ureña, “Some common indexes of group diversity: upper boundaries,” *Psychological Reports, 2012, vol. 111, num. 3, p. 777-796*, 2012.

[27] S. K. Horwitz and I. B. Horwitz, “The effects of team diversity on team outcomes: A meta-analytic review of team demography,” *Journal of management*, vol. 33, no. 6, pp. 987–1015, 2007.

[28] C. E. Shannon, “A mathematical theory of communication,” *Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[29] M. R. Symonds, N. J. Gemmell, T. L. Braisher, K. L. Gorringe, and M. A. Elgar, “Gender differences in publication output: towards an unbiased metric of research performance,” *PloS one*, vol. 1, no. 1, p. e127, 2006.

[30] B. Frick, A. Rose, A. Kolle *et al.*, “Gender diversity is detrimental to team performance: Evidence from a field experiment,” Paderborn University, Faculty of Business Administration and Economics, Tech. Rep., 2017.

[31] U. Cantner, M. Goethner, and M. Stuetzer, “Disentangling the effects of new venture team functional heterogeneity on new venture performance,” Jena economic research papers, Tech. Rep., 2010.