Women's Representation and Experiences in the High Performance Computing Community

Athina Frantzana

Doctor of Philosophy
The University of Edinburgh
August 2019

Abstract

Gender imbalance in STEM disciplines (Science, Technology, Engineering and Mathematics) has been a research subject for years. Studies have shown potential reasons leading to the underrepresentation of women in such disciplines and have suggested why and how to improve the gender balance and women's experiences in these areas. The High Performance Computing (HPC) community, which spans various STEM subjects and relies on advanced scientific research, might present a similar picture. The aim of this thesis is to understand the gender demographics of the HPC community, to identify the underlying reasons of a potential gender imbalance, and to suggest effective ways of improvement.

Since HPC is such a broad community, to obtain a first picture of the proportion of women in the HPC community, we decided to examine historical demographics of two different settings which are potential indicators of the participation and contribution of women in the community, namely HPC-related conferences and HPC training courses. From the analysis of these quantitative data, we found that women were fewer than men in all the categories of conference participation that we examined, and that women were outnumbered by men at all levels and years of the courses examined. Our study reveals an underrepresentation of women in the HPC community, along the lines of what already observed in STEM disciplines.

Additionally, we conducted a survey in order to further understand the reasons of the gender imbalance and to find out from the people within the HPC community what could be done to address the issue. Results from our survey indicate that the clear majority of both women and men forming the HPC community come from a STEM background, which is considered as the main reason of women's underrepresentation by the participants of this study. We also discovered that women are less likely to receive training and to develop software, both crucial factors for using HPC facilities for research purposes. Gender differences are also found in the impact of parenthood on career progression; the perception of

gender discrimination in workplace and conference environments; the importance of gender balance, mentoring, role models and Equality and Diversity awareness in the HPC community.

Similar findings and gender differences are also highlighted and confirmed by the results of further qualitative approaches of this study. We conducted interviews and focus group discussions with selected and recommended individuals of the community, to support and interpret previously obtained data, and to stimulate new ideas or hypotheses for future work. According to the interviewees and the participants to focus group discussions, one of the main challenges of the HPC community is its image of a closed, inaccessible, *"geeky"* area, which focuses on the size, speed and power of supercomputers, rather than on their use for solving problems in research and in life. This might be one reason that makes the community unattractive to women. Also, of significant importance for the current diversity status of the community is the fact that HPC is not well-promoted as a research tool, especially to more gender-balanced non-STEM subjects, in combination with the lack of formal (HPC and programming) training and of women in senior positions.

This thesis forms the first step to understand the womens representation and experiences within the HPC community. All the topics studied, and the evidence gathered in this thesis have provided significant insight to enable further research on the best practices for improvement in the HPC community and related STEM fields.

Lay Summary

This thesis forms the first step to understand the women's representation and experiences within the High Performance Computing (HPC) community, by quantifying the gender demographics of the community, identifying the underlying reasons of a potential underrepresentaiton of women, and suggesting effective ways of improving the community's gender balance.

Studies have shown potential reasons leading to the underrepresentation of women in STEM disciplines (Science, Technology, Engineering and Mathematics) and have suggested why and how to improve the gender balance and women's experiences in these areas. It is very likely that the HPC community presents a similar picture, since the community spans various STEM subjects and relies on advanced scientific research.

Firstly, we examined historical demographics of two different settings, HPC-related conferences and HPC training courses, which are potential indicators of the participation and contribution of women in the community. The results from the analysis of those data showed that women were fewer than men in all the categories of conference participation that we examined, and that women were outnumbered by men at all levels and years of the courses examined. Consequently, our study reveals an underrepresentation of women in the HPC community.

Through our HPC community survey, and the interviews and focus group discussions with people from within the community, we found that most women and men, who form the HPC community, study or work at a STEM-related environment. This fact was considered as the main reason for women's underrepresentation in the HPC community by the participants of our study. Thus, we examined if reasons for women's underrepresentation in STEM subjects, suggested by previous studies, apply in the HPC community. Indeed, we identified

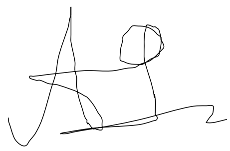
some reasons to be similar to those for STEM subjects, such as discirimination in the workplace and at conferences, impact of parenthood in one's career and others. However, there are some reasons suggested in our study, which are more HPC-specific. Such reasons are, the image of the HPC communtiy as a closed, inaccessible, *"geeky"* club; the lack of formal (HPC and programming) training; and the lack of women in senior positions.

Finally, from the analysis of our results, we were able to provide some suggestions on how to tackle the barriers and improve the gender balance in the HPC community. Additionally, we expressed our views on how to apply more effectively some of the current strategies aiming to the same goal in STEM subjects. We believe that all the topics studied, and the evidence gathered in this thesis have provided significant insight to enable further research on the best practices for improvement in the HPC community and related STEM fields.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

Part of this work, discussed in Chapter 3, has been submitted for publication in the Journal of Open Research Software: [1].



(*Athina Frantzana, August 2019*)

Acknowledgements

First and foremost I would like to thank my supervisors Professor Richard Blythe, Professor Marialuisa Aliotta, and Dr Will Hossack, for their invaluable guidance and advice. I am also grateful to all the women and men from the HPC community who willingly participated in my study, and without their contribution, this research would not be possible. I also thank Dr Toni Collis for introducing me to the topic of this study and the Women in High Performance Computing (WHPC) Network.

Special thanks to my life partner, Andrew Wright, who has listened patiently about my research and offered feedback and support throughout my studies; and to our dog, Chiquita, who sitting next to me while working, kept me sane and made everything easier. Also, a special thank you to my parents, Maria and Dimitris, and my brother Sotiris, who supported me in any way they could, and without their encouragement, I would not be here today.

Finally, thank you to Dr John Loveday for giving me the opportunity to study here, and the University of Edinburgh for funding my research.

Contents

List of Figures

List of Tables

Chapter 1

Introduction and Motivation

Women have been underrepresented in most STEM disciplines[1] and throughout the levels of education and academic or industrial positions. Much research has been done to identify the underlying reasons and the best ways to tackle the gender gap. Yet, women are still underrepresented and the gender imbalance in STEM is still an issue, which causes socio-ecomonic problems.

The focus of this study is the representation of women in the High Performance Computing community. High Performance Computing (HPC) or Supercomputers are mainly used for a variety of computational tasks for advanced research, such as molecular modelling and physical simulations, in a wide range of fields and facilities, from local clusters to national HPC centres. The HPC community is not clearly defined; however it has been traditionally limited to academia and government. Anecdotally, the majority of the users of HPC facilities and the technical experts of the field have a STEM background. If the HPC community consists of individuals from STEM disciplines, there is a high chance that women are underrepresented.

Nevertheless, it is worth mentioning that there is currently an increasing demand of using HPC in other - non-traditional HPC- fields, from Artificial Intelligence and Machine Learning to medical and financial modelling, as a result of the growing needs of data-intensive applications and processess. This interest and

[1] STEM stands for Science, Technology, Engineering and Mathematics, and it is an acronym commonly used to group together these and related disciplines. However, the disciplines included in this group are often a topic of debate. In this thesis, we use the acronym to include also Biology, Medicine and all the biomedical related subjects.

demand can expand the HPC community further to the commercial sector and business areas, which might have an impact to the community's gender balance.

One part of this study's purpose is to attempt to define the HPC community and to demonstrate evidence for women's underrepresentation. The other part is to identify the reasons for this gender imbalance and the strategies to tackle it. In the following sections, we present the current picture of women in STEM and a brief review of the research on the potential reasons of their underrepresentation and of the suggested practices for improvement, which motivated our research and formed the basis of this study.

1.1 Women in STEM

1.1.1 What do the numbers say?

According to UNESCO Institute of Statistics, as of June 2018, women form less than 29% of the people working in STEM research and development [2]. Globally the percentages of female and male individuals present a significant difference at PhD level of education and later at researcher level, with women falling to under 50% and under 35% respectively [3].

A similar situation occurs in the UK. According to recent statistics by the Women in Science and Engineering (WISE) Campaign[2], women form 24% of the core STEM (health related occupation are not included) workforce in the UK, with the lowest representation in Engineering (11%) and IT (17%-19%). The statistics also show a very low percentage for women in managerial positions in Science, Engineering and Technology (15%). The WISE statistics also emphasize the STEM Education *"Pipeline"*, which is the significant decrease of girls studying STEM subjects from A-levels onwards (Figure 1.1).

Statistics by the Higher Education Statistics Agency (HESA) present the numbers of female and male students in the various diciplines of Higher Education, which make obvious the underrepresentation of female students in STEM subjects, especially in Engineering & Technology and Computer Science. On the contrary, subjects such as Historical & Philosophical Studies and Business & Administrative Studies seem to have gender balance, and subjects such as

[2] `www.wisecampaign.org.uk/statistics`*Last accessed: October 2018*

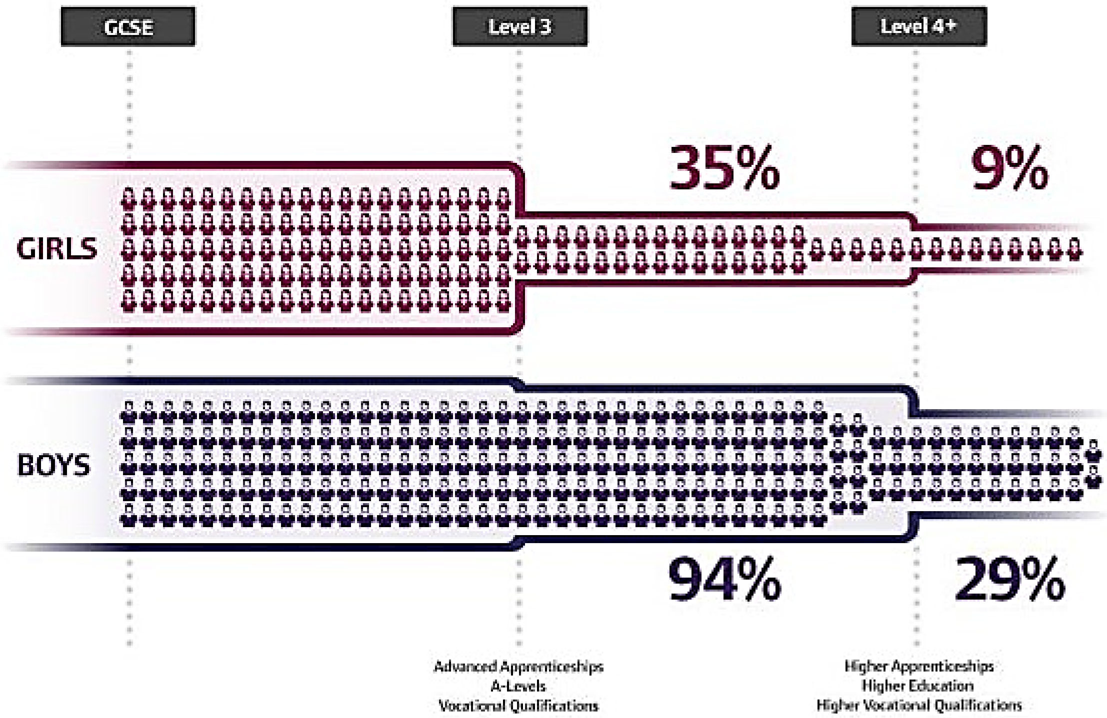


Figure 1.1 *Percentages of girls and boys at the different levels of STEM education, with a significant decrease of girls from A-Levels onwards. Source: WISE.*

Education and subjects related to medicine appear to have more female than male students. Figure 1.2 projects the percentages of female and male students of most of the disciplines for the academic year 2016/17, according to the table provided by HESA[3]. In addition to this, the table[4] which compares the numbers of female and male students' enrolments to the various disciplines for the last three years of data (2014/15, 2015/16 and 2016/17) does not show any change. This lack of improvement could be an indication that the current strategies and practices being used to increase the number of female students in STEM subjects are not effective.

The same source also gives information about the numbers of female and male members of academic staff at the various disciplines. Figure 1.3 is based on these numbers for the year 2016/17 provided by HESA[5]. The small percentage of female members of academic staff in Engineering & Technology stands out and it is consistent with the percentage of the female students of the same discipline

[3] `www.hesa.ac.uk/data-and-analysis/students/chart-7` *Last accessed: October 2018*
[4] `www.hesa.ac.uk/data-and-analysis/sfr247/figure-14` *Last accessed: October 2018*
[5] `www.hesa.ac.uk/data-and-analysis/staff/cost-centres` *Last accessed: October 2018*

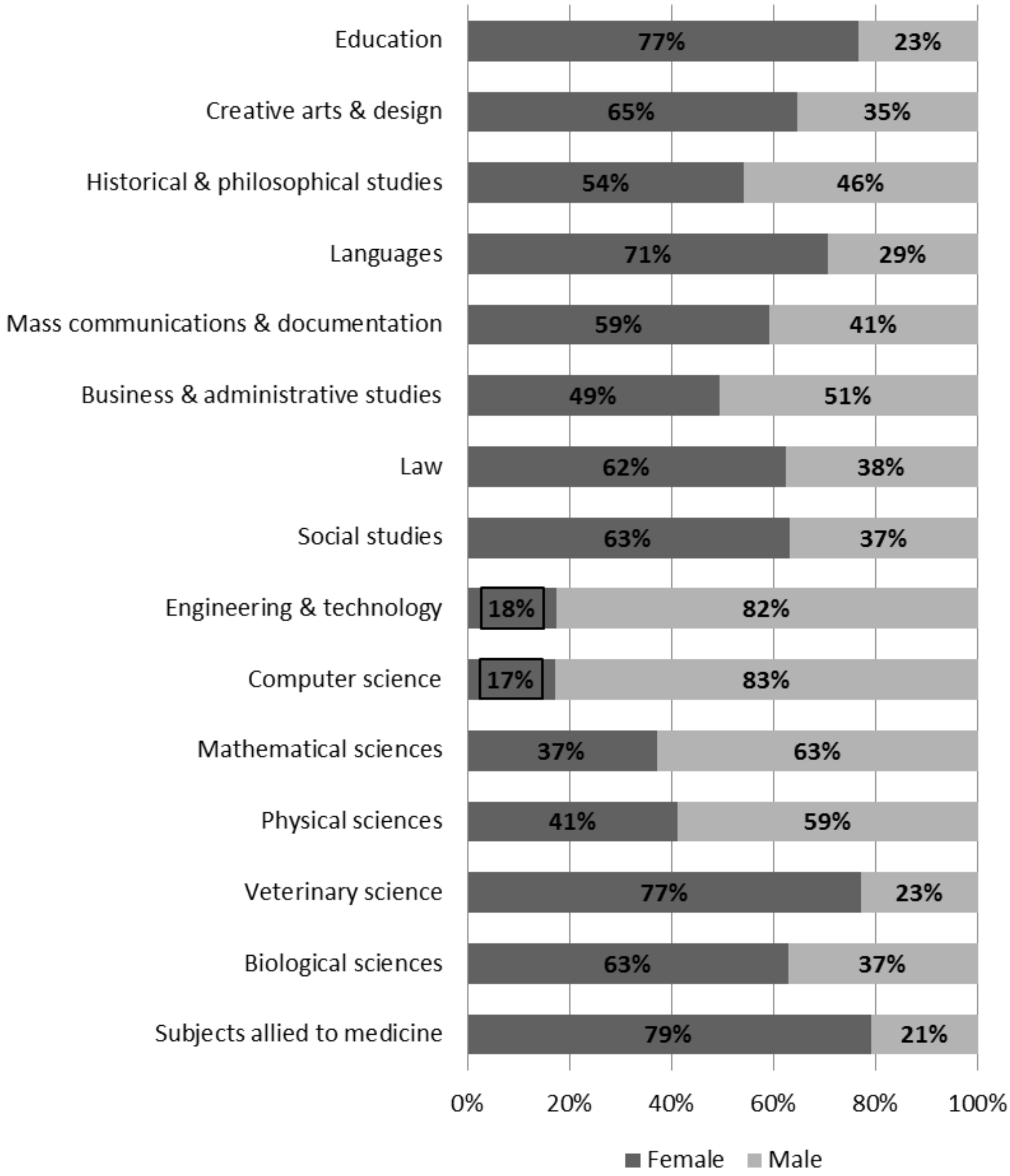


Figure 1.2 *Percentages of female and male students of various disciplines in the UK for the academic year 2016/17, based on data by HESA. The percentages in boxes highlight the remarkably low percentages of female students in Engineering & Technology and in Computer Science.*

that we discussed earlier.

Finally, it is important to mention that data related to the type of academic employment by HESA[6] have also made clear that: female professors are outnumbered 3:1 by male professors; women are less likely than men to be in

[6] `www.hesa.ac.uk/data-and-analysis/staff/employment` *Last accessed: October 2018*

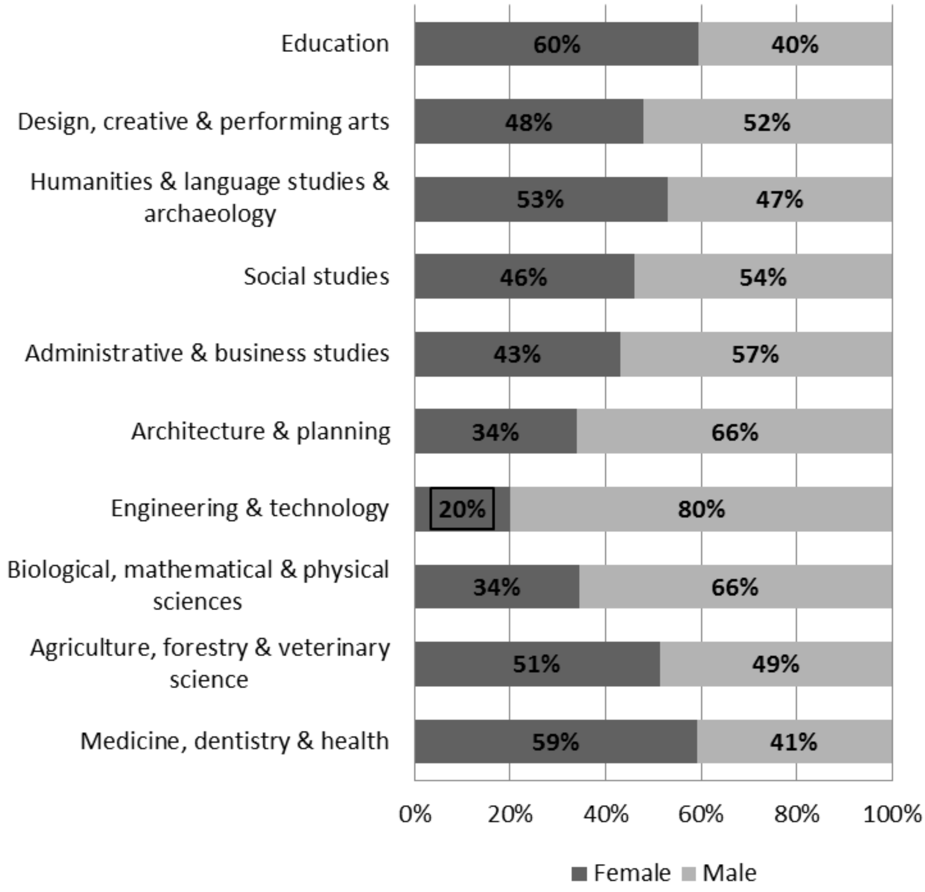


Figure 1.3 *Percentages of female and male members of academic staff across disciplines in the UK for 2016/17, based on data by HESA.*

other senior academic positions (36%-37%); and women are more likely to work part-time (56% of all part-time academic staff), especially as researchers (67% of all part-time researchers).

1.1.2 Why so few women?

Much research has been carried out to identify the reasons for the underrepresentation of women in STEM fields. We briefly present here some of the most important hypotheses and findings.

1.1.2.1 Biological differences and Life-style Choices

One of the initial hypotheses to explain the gender gap in STEM subjects was that there are biological differences in women's and men's abilities and interests. Various studies that have suggested that: 1) there are some differences in the performance of men and women at math-related tests [4]; 2) men have bigger brain on average [5], and they use different parts of it to complete the same tasks [6]; 3) prenatal hormones exposure might influence gender-related behaviours [7, 8]. However, the conclusions drawn from these studies are not clear, and there is research which suggests that sociocultural factors should be also taken into consideration when we try to explain the STEM gender gap [9–11].

An interesting finding of related research suggested that women with high competence in math-related subjects are more likely to also have high verbal competence, which broadens their career choices [12]. Usually, this results in math-proficient women preferring to pursue non-math related careers or dropping-out later if they choose a math-related career; possibly, this is due to existing stereotypes of male-dominated STEM environments or women's concerns for balancing motherhood with a career in STEM [13].

One study found that the ratios in which men outnumber women in STEM subjects are not consistent across different countries and cultures, which suggests that sociocultural factors may influence those numbers [14]. The study showed no correlation between women's math abilities and their entry to math-related field, however women are underrepresented in particularly math-intensive STEM filelds across all examined countries.

Women are more inclined to choose people-oriented careers, which might explain why they prefer to study and pursue careers in either non-STEM or in health-related subjects [15]. Particularly, research has indicated that women are more interested in "people", men are more interested in "things", and most people perceive careers in STEM fields as less people-oriented than in non-STEM fields [16–19]. Additionally, studies have shown that women believe that to succeed in STEM fields, one must be a "genius", and that women's success in STEM filelds is due to their hard work and not to their natural abilities [20–22]. These findings suggest that gender differences in interests and perceptions might play a role in the gender segregation of occupations in general and within STEM.

Women were also found to be more interested in home-centred lifestyles or prefer flexible work, whereas the majority of men prefer work-centered lifestyles [23]. Studies also show that having children has a more serious negative effect on women's career; that female academics are less likely to be married and have children than their male colleagues ; and that when female academics have children are more likely to work fewer hours or drop-off [24–26], which has an impact on their career since they have fewer opportunities for promotions or tenure-track positions [27–30]. Additionally, a study found that men with children are more appealing employees [31]. Interestingly, there are studies with similar findings on the impact of having children on women's career in non-STEM fields, where women are not underrepresented [24, 25]. It is also suggested that women most likely do not leave academia because of having children, but because of their work not being recognised and not having equal opportunities of progressing [12, 32, 33].

1.1.2.2 Stereotypes and Confidence

Women percieve differently interest and success in STEM fields, which is most likely based on existing stereotypes for STEM and women's abilities at those subjects; this might be a reason for the underrepresentation of women in STEM [13, 34]. Such stereotypes have been cultivated by the society for years and unfortunately have been adopted by parents and teachers, and consequently students' peers. Studies have shown that parents support their sons more with maths-related toys and activities, as they believe more in their sons' maths abilities [10]. A study found that fathers are more biased towards gender stereotypical toys, such as dolls, and they discouraged their sons to play with such toys [35]. A correlation between playing with toys such as Lego blocks and maths ability has also been found [36], and usually it is boys that are more encouraged to play with such toys. In the same pattern, boys are more likely to own and play with a computer at home, which might help their maths performance [37] and computing skills.

Similarly, teachers are more likely to encourange boys to ask questions [38], as well as to address and praise boys in science and maths classes [39]. Teachers' favouritism over boys in science and maths classes seems to have a positive impact on boys' performance, and encourages them to enrol in advanced science and maths courses, whereas teachers' biases have a negative impact on girls [40]. Research has also shown how peers' behaviour, and especially that of male peers, influences female students' performance and motivation in science and maths subjects [41, 42]. A study found that in a single-sex (girls only) classroom, girls had better physics-related perception and performance than in a mixed-sex classroom, where male students are dominant [43, 44]. Peer influence does not affect only children; studies have demonstrated how peer pressure and relationships within STEM fields can lead women to lose their ambitions or to leave their studies [45, 46].

Girls' and women's progress and performance in STEM influenced by parents', teachers' and peers' behaviour is hugely connected with Stereotype Threat. Stereotype Threat (ST) was defined as *"being at risk of confirming, as self-characteristic, a negative stereotype about one's group"* [47], and it often happens within fileds where women are underrepresented [48].Various studies have examined how ST affects women's performance, mainly on maths-related tests, and the majority of them have found a negative effect leading to decrease of preformance and motivation, and increase of negative thinking [34, 49, 50]. However, even though the effect of ST is well documented, it is not considered to be the main reason for the underrepresentation of women in STEM [10].

A similar impact on one's performance, but mainly focused on self-confidence and self-awareness, is that of the Impostor Syndrome. Impostor Syndrome (IS) was defined as *"an individual experience of self-perceived intellectual phoniness"* [51]. This initial study found that IS is particularly intense among high-achieving women; and even though other studies indicated that IS is common among women and men [52], women were found to express greater impostor fears related to performance and goal achievement [53]. Many people believe that IS might be an additional reason that women do not to pursue STEM careers and are greatly outnumbered by men in senior positions.

1.1.2.3 Biases and Discrimination

Significant research has also been done on institutional biases, which might lead to various forms of discrimination over women in STEM, and which are stongly connected with the stereotypes mentioned earlier. A popular study [54] showed that senior faculty members were biased in favour of men during a hiring procedure. Similarly, another study found that employers were hiring twice as many men as women, when they were aware of the candidates' appearance [55]. Other examined forms of discrimination over women in academia and workplace include sexism, tokenism and harassment [56]; lower rate of successful fund and promotion applications [57, 58]; difference in encouragement to undertake various roles [59]; fewer opportunities to participate in collaborative research projects [60]; lower rate of promotions, which also leads to a gender pay gap [32].

Another examined form of gender discrimination is related to peer-reviewed publications and conference participation, on which we based our quantitative research and analysis in Chpater 2. Men dominate research papers and conference visibility, as they are more likely to request a talk than a poster [61], to be invited as speakers [62, 63], and to publish papers [64–66]. This gender difference in visibility has a negative impact on women's recognition and progress in their field [67], and restrains women from specialising and becoming experts of their discipline [68]. Even though there are studies which claim that nowadays institutional discrimination is less significant or absent than in the past, and it should not be a reason for the underrepresentation of women in STEM [69], women still face institutional barriers, compared to men, during their career progress.

It is believed that the institutional discrimination towards women is due to unconscious bias (also known as subconscious or implicit bias), formed from the exposure to sociocultural stereotypes, which do not relate women to STEM subjects. Unconscious Bias (UB) is *"the unconscious attribution of particular qualities to a member of a certain social group"* [70]. In addition to the studies mentioned earlier, there are several other studies that have proved the existence of gender-related UB, for example in hiring procedures [71], employee evaluation [72], and studies that proved that certain changes in those procedures can reduce gender bias [73].

1.1.3 Why is gender balance important?

Better gender balance in STEM fields is important for many reasons. Apart from the obvious reasons of equal opportunities and broader options for everyone without socio-cultural and institutional barriers, so that everyone can equally take part in the progress of the world and life, more women in STEM are also important for other reasons.

Research has shown that gender-balanced teams demonstrated better performance, higher productivity, and better collaboration. In particular, a study found that the ability of a group to collaborate and perform a variety of tasks highly improved by the presence of a greater proportion of women in the group [74]. The same study also showed that teams with more women are more likely to achieve equality in individuals' participation. Another study found that women tend to recognize expertise of team members more easily and more often than men, and they tend to focus on educational qualifications of the individuals rather than on superficial and irrelevant features, such as gender, like men do [75]. Additionally, a field experiment showed that gender-balanced teams perform better than male-dominated and female-dominated teams, owing to more equal learning in gender diverse teams [76]. Gender diversity may also trigger different research by broadening the views, the questions and the purpose of the research, which might lead to new discoveries [77].

These findings suggest that the lack of gender diversity of teams in male-dominated fields, such as STEM, could have serious negative effects to team's productivity, as well as to STEM innovation in general. However, studies have shown that initial efforts to create gender diversity on male-dominated environments are likely to have negatives effects, due to existing negative gender stereotypes and the lack of equal influence and participation of men and women in a team [78, 79]. These results imply that just examining and increasing the number of women in STEM in order to close the gender gap might have serious social consequences and might not solve the women's underrepresentation problem. Nevertheless, a recent research to the heavily male-dominated environment of Open Source Software (OSS) found that overall increased gender (and tenure) diversity has a positive effect on team's productivity and results [80].

As mentioned earlier, innovation is greatly dependent on gender diversity. Reports [81, 82] showed that the majority of the technology patents and products are invented by male-only teams, which means that the technology used by everyone in the world is created by a homogeneous group of people, missing out on female talent, influence and needs. Consequently, gender diversity could improve technology products that cover all needs and are more competitive.

Related to this is also the *"Gendered Innovations"* movement, which is defined as: *"the process that integrates sex and gender analysis into all phases of basic and applied research to assure excellence and quality in outcomes"* [83]. This project addresses the gender bias in research, which results in limited scientific knowledge and innovation, and leads to problems which could even be life-threatening;

for example, full understanding of diseases and developing the most effective treatments might be restrained by medical research focused on one gender. Gender biased scientific research might be a result of the underrepresentation of women in the research fields and the lack of diverse viewpoints.

Finally, a good reason for increasing the number of women in STEM is to broaden the talent pool, to fill in the positions of the fast growing STEM (especially Technology) industries, and help the economy. Reports for the UK have indicated that STEM employers experience difficulties in recruiting individuals with appropriate skills and qualifications, and have predicted a high demand in STEM graduates to fill the vacancies of the current and new jobs that will be created in STEM sectors [84, 85].

1.1.4 How can we change the gender imbalance?

Based on the previously mentioned research which identified the barriers that girls and women meet during studying and pursuing a career in STEM, institutions and researchers have developed various policies and strategies to tackle those barriers and have measured their impact and effectiveness. We present here a brief summary of existing research.

1.1.4.1 Data and Outreach

Providing evidence and data of the discrepancies in the numbers of women and men studying and working in STEM-related environments is key to motivate action for change and imrpovement [86, 87]. For this reason, many networks of women in various STEM disciplines have been formed around the world to regularly publish statistical evidence of the underrepresentation of women and expose cases of discrimination. Those networks have also established a platform for women to network, express their concerns, and promote their work (for example, WISE Campaign[7]).

Also, it is significant to make sure that girls are aware of their potential and the various paths of a career in STEM by being exposed to female STEM professionals and breaking the gender stereotypes linked to STEM, and getting support by teachers, parents and peers to motivate and encourage their interest in STEM

[7] `https://www.wisecampaign.org.uk` *Last accessed: October 2018*

subjects. There are several studies which have proved the positive effect of outreach programmes on girls' interest and engagement in STEM subjects [88–90], as well as the significant impact of the support by teachers, parents and peers encouraging girls' interest in STEM subjects [91].

1.1.4.2 Role Models, Mentoring, and Change of Stereotypes

It is believed that changing stereotypes can increase girls' interest in STEM subjects [92]. A suggested way is by increasing the representation of role models for girls, since there is a lack of role models in STEM [93]; women appear to relate more to STEM role models who do not fit the current stereotypical image [94]; and women tend to persist with their studies more into fields with more women [95]. Indeed, there are studies that demonstrated successful results of using female role models to attract girls' interest and change their views on women's career in science [96, 97]. Role models and networking opportunities are also important for women to progress and succeed in their careers[33, 98]. In disciplines with better gender balance, it is easier for women to form networks, be more productive, and potentially more visible [99]. Additionally, at a conference with more female members in the organising committees, there might be a bigger number of female speakers [100].

These findings are in line with the findings of studies which examined the effect of in-group bias. In-group bias (or In-group - Out-group bias) is the action of favouritism among members of the same group over members of another group. In the findings of those studies, men appear to show stronger in-group bias than women at a younger age [101] and when they compete to another group [102]. Nevertheless, research shows that women have a stronger in-group bias for other women, when it comes to gender [103]; which probably emphasises women's need of other women in their environment to feel the sense of belonging.

Mentoring is often suggested as an effective way of supporting women to gain confidence and overcome barriers throughout their career route. Research has shown the importance of mentoring frameworks in the workplace and the positive impact of mentoring on women's career [104, 105], and on the progression and retention of doctoral students [106]. Similarly, mentoring has a significantly positive impact on women's participation specifically in STEM fields [107], as well as on female students' confidence in pursuing STEM careers [108] and in staying and continuing their studies in STEM subjects [109].

Nevertheless, the definition of mentoring is still debatable and experts have not decided whether formal mentoring is more effective than informal [110, 111]. For example, a study concluded that informal mentoring by teachers and peers was more effective in encouranging teenagers into computing [112]. On the other hand, institutionally organised formal mentoring has definitely a positive impact on women [113]; but also both formal and informal mentoring are important for the career development of men and women [111], emphasising that for mentoring to have an impact, opportunities for mentoring must be clearly available.

1.1.4.3 Equality & Diversity and Unconscious Bias Training

Another suggested, and currently popular, way to eliminate stereotypes and biases is Equality&Diversity and Unconscious Bias trainining, which raises awareness by presenting real-life examples of discrimination, and offers effective recommendations to identify and control biases. Such training can have positive results in changing behaviours and reducing biases in the workplace [114], and particularly in creating a climate that supports women's career in STEM disciplines [115].

However, studies on the effect of diversity training on managerial biases were disappointing, suggesting that such training had no effect or even had negative effect [116, 117]. Additionally, a recent report indicated that UB training can be effective to raise awareness and reduce implicit bias, but it is unlikely to eliminate implicit bias and change behaviour, and it might in fact have opposite effects [118]. The report emphasises the need to focus on the content of the UB training to be more educational and empowering; the need to constantly measure the impact of the UB training; and to consider the training as part of a wider programme for organisational change. A recent study revealed the same results for diversity training, highlighting the fact that just diversity training and initiatives cannot change the gender gap in STEM [119].

1.1.4.4 Parental Leave and Engagement of Men

We discussed earlier (Section 1.1.2.1) the significant impact of motherhood on women's career. Studies tried to identify the best ways to support mothers, and have found that with the right insititutional support for mothers and part-time employees (the majority of whom are women/mothers), they can be as productive as childless women, who work full-time [111, 120].

Shared parental leave was suggested as a solution to support mothers, and organisations in many countries adopted this policy. It was introduced in the UK in 2015 [121], and since then reports have shown that only a very small percentage of the entitled parents have used it [122]. The main barriers that have prevented the policy from properly being used seem to be organisational and communicational (lack of knowledge and promotion of the policy by the organisations), and sociocultural, as a result of the gender stereotypes around the roles of each parent. Additionally, parental policies vary from country to country and among organisations, institutions and levels of employement, and their effects in gender equality have not yet been established [123]. Researchers certainly recommend that shared parental leave should be better promoted and fathers more encouraged to make use of it, so that it can have an impact on gender equality in the workplace [124, 125].

1.2 Women in HPC

No published study, that we are aware of, has examined the representation of women explicitly in the HPC community. Nevertheless, since anecdotally the majority of the people who form the HPC community belong to the STEM group, we speculate that women are underrepresented. Additionally, if one considers HPC as a niche or advanced field within the broader computing area, women must be a minority, since women are particularly underrepresented in computing, as we discussed earlier (1.1.1). A recent study which investigated the gender gap in various computing domains by analysing data of paper authors adopted from various conferences, included two HPC-related conferences [126]. The study found that the proportion of female authors ranges from 4.5% to 10.4% at the various examined conferences. This study is probably the only on its kind to have mentioned HPC; however, since it includes many other computing fields, it cannot be conclusive just for the HPC community.

Traditionally, one of the basic skills that a person needs to use HPC facilities is the ability to program and develop software. A few studies have explored the gender representation and biases in programming and software engineering. These studies found that female contributors in online programming communities form 5.8% [127], but also that women's contributions (for programming improvement) are accepted more often than men's [128]; that there are significant gender differences in using various programming environments, exploring software features, and solving technical problems [129]; and that 11% of the UK Research Software Engineering (RSE) community are women [130]. Another crucial finding for the HPC community is that women are underrepresented in computational research even in areas with a large number of women, such as biology [131].

Even though there is no previous study or report that shows the underrepresentation of women in the HPC community, there are many initiatives and organisations which aim to address equality and diversity in various peripheral to HPC fields and might have an impact on HPC community's gender balance. They offer a variety of programmes that raise awareness, support underrepresented groups and early career reearchers, and share best practices on inclusivity. Such initiatives include: the Extreme Science and Engineering Discovery Environment (XSEDE)[8] and the Campus Champions programme[9], and the Engineering and Physical Sciences Reseasch Council (EPSRC) RSE Fellowship programme[10].

More specifically for the HPC community, the Women in HPC Network[11] was founded in 2014, aiming to provide evidence of the community's gender gap, suggest solutions, promote and support the women of the community. Also, diversity committees were established by international HPC conferences, such as the Supercomputing Conference (SC)[12], whose main tasks are to support the underrepresented groups of the community and educate on equality and diversity matters. For example, in 2016 the SC diversity committee offered child care and support for the first time attendees, gathered and published demographics and evaluation feedback of the conference. Similarly, the International Supercomputing Conference (ISC)[13] started publishing data related to gender and other characteristics. Additionally, in 2016, the Partnership for Advanced Computing in Europe (PRACE) launched the PRACE Ada Lovelace Award[14] for the promotion of female scientists of the HPC community; and the University College London named their new computing facility after Grace

[8] `https://www.xsede.org/community-engagement/diversity` *Last accessed: May 2019*

[9] `https://www.xsede.org/community-engagement/campus-champions` *Last accessed: May 2019*

[10] `https://epsrc.ukri.org/funding/calls/rsefellowships/` *Last accessed: May 2019*

[11] `https://womeninhpc.org` *Last accessed: October 2018*

[12] `http://sc16.supercomputing.org/diversity/index.html` *Last accessed: October 2018*

[13] `https://www.isc-hpc.com/diversity.html` *Last accessed: October 2018*

[14] `http://www.prace-ri.eu/about-prace-ri/prace-awards/adalovelaceaward/` *Last accessed: May 2019*

Hopper in order to promote female role models in Computing[15].

All these initiatives and actions have made clear that there is a gender issue noticed in the community, that needs to be addressed and solved.

1.3 Thesis Structure

The aim of this thesis is to provide evidence of the underrepresentation of women in the HPC community and identify the reasons and the best ways to improve the gender balance. Having laid out the background work and motivation in the current chapter, we now describe the structure of this study in more detail.

Chapter 2 includes the quantitative approach of this study by presenting the analysis of two sets of data used, namely, data from HPC-related conferences and from HPC-related training courses. The gender was assumed by the first name of the individuals who participated in various capacities at the examined conferences to quantify the representation of women at the community's conferences. The responses to training courses feedback forms, where the participants have stated their gender, were also analysed to quantify women's participation at the various levels of HPC-related training courses.

Chapter 3 presents a semi-quantitative analysis of extant data of a survey carried out by the Software Sustainability Software (SSI) and it was used as a guide to design the HPC community survey which was carried out explicitly for this study. This chapter also mentions other pilot surveys and questionnaires used to help further design of research methods. This semi-quantitative approach offers information in an effort to define who forms the HPC community, and investigates the views of the community and some of our initial hypotheses on the reasons of the underrepresentation of women.

Chapter 4 comprises a qualitative analysis of the interviews and focus group discussions conducted with people from within the HPC community for this study. The interview and discussion questions were generated based on the results from Chapters 2 and 3, as well as on the background work discussed in Chapter 1. This approach aimed to collect evidence of the underrepresentation of women in the HPC community, and identify the reasons and ways to tackle the gender gap

[15] `https://www.ucl.ac.uk/research-it-services/grace-launch` *Last accessed: May 2019*

through personal views and experiences of women and men of the community.

Finally, Chapter 5 offers some conclusions and discussion of the research presented, and an outlook for future work based on the findings of this study.

Chapter 2

Representation of Women in the HPC Community: Quantitative Approaches

In this Chapter we will present the data and discuss the results from the quantitative approach of the study. To obtain a first picture of the proportion of women in the HPC community, we needed to examine historical demographics of two different settings, which are potential indicators of the participation and contribution of women in the community. Therefore, we produced two datasets: one consisted of lists of the participants at HPC related conferences, where we assessed their gender based on their first name; and the other included the responses to the feedback forms of attendees to HPC related training courses, where the respondents self-reported their gender.

For the analysis of the data in Chapter 2 we mainly used “descriptive” statistics [132] and “inferential” statistics with tests of statistical significance [132], [133]. Finally, we marked the standard error of averages with “SE” and $\pm number$ (more details on the statistical analysis of the study can be found in the Appendix A.1).

2.1 Conferences

Conferences are traditionally the main meeting hubs of individuals and/or companies in a certain field. For this reason, quantifying the participation of women at HPC related conferences could help to produce a first picture of the representation of women in the HPC community. However, studies in other fields found that men dominate conference visibility, as they are more likely to request a talk than a poster [61], more likely to be invited as speakers [62], and more likely to publish papers [64]. If the HPC community follows the same patterns, then our results might not reflect the actual demographics of women in the community, but their representation at the community's conferences and other dimensions of gender inequality.

We examined the numbers of women who participated at conferences of the HPC community, as speakers, paper and poster authors, and attendees. Additionally, we recorded other forms of contributions, such as moderating a workshop, asking questions, and chairing a session.

In the first part (2.1.1) of this section, we will describe our preliminary findings by attending conferences and using the research method of "participant observation". In the next part (2.1.2), we will discuss the results of the gender analysis of the participants of HPC related conferences, for which we adopted data from publicly available programmes and proceedings of the examined conferences.

2.1.1 Participant Observation

Participant observation is the process which enables a researcher to learn and collect data of the community under study through participating in the community and observing, and it is widely used as a qualitative method of data collection [134]. It is recognised as one of the best methods to begin a study with ethnographic nature, as with observation the researcher can build a theory and generate or examine hypotheses [134]. Finally, participant observation is believed to increase the validity of the study, as it gives the opportunity to the researcher to collect several types of data, both quantitative and qualitative, and to develop meaningful research questions [135].

For this study, we used the "selective observation", which means that the researcher focuses on specific activities and outlines behaviours and potential differences of the study sample [136]. Our focus was on the gender of presenters, session chairpersons and moderators, and on the frequency of attendance of both men and women at various sessions. The number of questions asked by women and men during sessions was also recorded, with the hypothesis that women ask fewer questions than men at conferences, as other studies have concluded [137], [138].

The participant observation method was also complemented by selecting data from the printed programme and the delegates list which were provided at each conference. These printed sources were used to extrapolate the gender by the first name of the registered participants and speakers at the conference sessions which we were not able to attend.

2.1.1.1 Data Collection

In 2015 we had the opportunity to attend and participate at the following HPC related conferences:

- the **Exascale Applications and Software Conference (EASC)**, which focuses on issues of applications for exascale and the associated tools, software programming models and libraries. In April of 2015, EASC took place in Edinburgh (UK) and included a total of 20 sessions.We attended the 4 keynote sessions, the poster talks' session, and 8 of the parallel sessions;
- the **PRACEdays**, which is one of Europe's most important conferences on HPC and is organised annually by the Partnership for Advanced Computing in Europe (PRACE). In May of 2015, PRACEdays was organised in Dublin (Ireland) and consisted of 22 sessions (5 of which were keynotes). We attended all of the sessions;
- the **International Supercomputing Conference (ISC) High Performance**, which is based in Germany and is the world's oldest HPC conference with the first one dated in 1986. ISC 2015 took place in July in Frankfurt (Germany). Since ISC is one of the biggest HPC related conference, it was not possible for us to attend enough of the sessions for observation and counting, so the data gathered for this conference were mainly adopted by the printed material.

At conferences, keynote speakers are those who highlight the importance of the topic of the conference and are usually highly recognised within the community. According to a study [139], keynote speakers are often the most attractive element of a conference. Thus, we counted the keynote (and invited, where applied) speakers seperately from the paper presenters and other sessions' speakers. We also handled the attendees and the questions asked at keynote sessions seperately, to detect any potential differences in attendance and question frequency among keynote and other seesions.

2.1.1.2 Women's Participation

Table 2.1 summarises the percentages of women for the various categories we counted while attending the conferences, as well as information gathered from the printed conference material. The printed programmes of PRACEdays and ISC included only the names of speakers and presenters, and not of all authors of the presented papers. Additionally, a delegate list was not provided at ISC. Finally, tutorials or workshops were not organised at EASC and PRACEdays and there is no data for organisers and moderators.

Table 2.1 *Total numbers of participants (N) and percentages of women (F%) at the various categories from the analysis of the data gathered for EASC, PRACEdays (PRACE) and ISC 2015. Where "n/a" stands for no data available.*

	EASC N	**F%**	**PRACE N**	**F%**	**ISC N**	**F%**
Papers authors	142	13%	n/a	n/a	n/a	n/a
Delegates	155	16%	162	17%	n/a	n/a
Speakers	41	15%	17	0%	487	7%
Keynote speakers	5	20%	6	0%	3	33%
Poster presenters	17	12%	18	28%	n/a	n/a
Chairpersons	20	15%	6	0%	59	8%
Organising committee members	n/a	n/a	9	22%	n/a	n/a
Organisers/Moderators	n/a	n/a	n/a	n/a	78	9%

As shown in Table 2.1, the percentages to be noticed at a first glance are the keynote speakers at EASC and ISC, where in both cases there was one woman. Also, when comparing speakers with poster presenters, we could say that PRACEdays conference follows the findings of previous studies that women prefer to present a poster than give a talk [61]. Overall, the percentages of women at EASC, however lower than men's, show that women distributed equally across the various roles of the conference. The same conclusion cannot be made for PRACEdays, which appears to have had some important "problematic" areas, as the complete absence of female speakers, keynote speakers and chairwomen.

It is difficult to draw a general conclusion for ISC because of missing data.

However, the very small percentages of female speakers, organisers/moderators and chairwomen at a conference of the capacity of ISC should draw our attention and prompt further investigation.

Furthermore, we recognised that the numbers of female presenters and organisers/moderators at ISC would have been affected if we had added the numbers of women who represented the Women in HPC (WHPC) network (1.2), as the numbers of these women were “outliers”. In descriptive statistics, “outlier” is a number that differs from the other numbers of the data set, and it usually means that this entry is affected by other factors. It is better to exlcude this number from the data set for analysis, or use this “uncharacteristic performance”, but explain the circumstances [133]. When we recalculated the total number of female presenters by including the number of the WHPC network’s presenters, the percentage changed from 7% to 8% female presenters, which did not seem as a remarkable difference, but it is statistically significant ($p < 0.05$). When we did the same for the organisers and moderators, the percentage increased from 9% to 19% ($p < 0.05$). We could say that this difference in the percentages demonstrates the role of a network which supports the women of a community, such as WHPC, in engaging more women with the conferences of the community.

2.1.1.3 Further Analysis and Discussion

We also counted the number of female and male attendees and the number of questions they asked during the sessions which we attended at EASC. We found that even though the number of questions asked by men was always higher, the number of questions asked by women was in proportion to the number of female attendees. Female attendees at PRACEdays did not ask any questions except at two of the keynote sessions, which automatically means that there was a clear difference between the ratio of female attendees and the ratio of questions they asked at each session. This lack of questions by women at some of the examined sessions explains the large error bars in Figure 2.1.

Research shows that more women in the organising committee may increase the number of female speakers at a conference [100]. However, this was not the case for PRACEdays, where we had the data to compare these two features: with 22% of the organising committee being women, there was only one female keynote speaker in the programme and no other female speakers. An interesting fact to be mentioned is that the female speaker of PRACEdays’ printed programme did

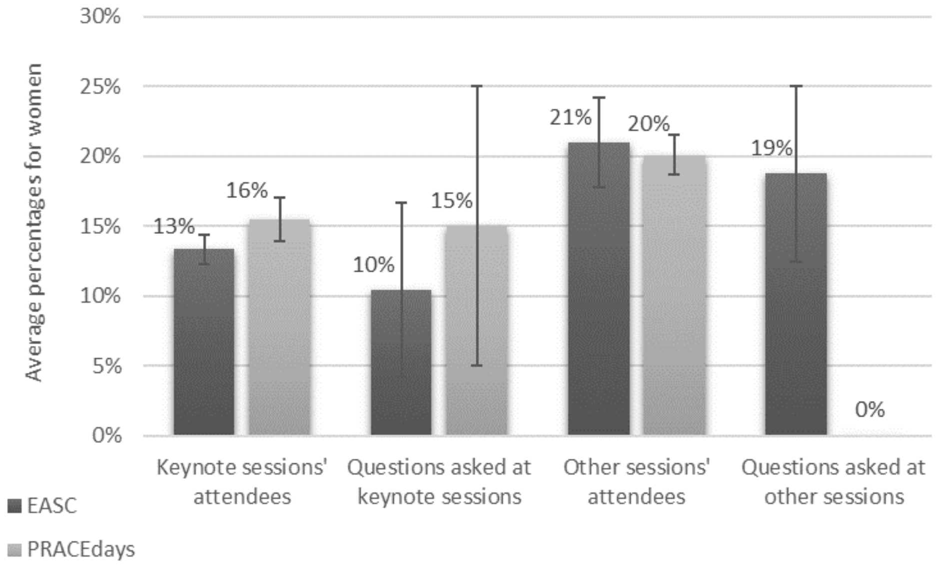


Figure 2.1 *Comparison of average percentages of female attendees and questions asked by women at keynote and other sessions of EASC and PRACEdays conferences.*

not manage to attend the conference for personal reasons, as we were informed during the day of the conference; this explains the 0% of keynote speakers in Table 2.1. However, this is an isolated case which lacks further details, and it cannot lead us to any conclusions about the reasons why women do not attend or present at conferences as often as men.

Following the same thinking process on a likely correlation of the number of female committee members and the number of female speakers, we hypothesised a potential increase of the number of female attendees at the sessions with female speakers. Given that we did not have the numbers of attendees for ISC and that there were no female speakers at PRACEdays, we looked at this aspect only for EASC. Even though the numbers of female speakers and attendees were small, we calculated the average percentage of female attendees at the sessions with female speakers that we attended as $(21 \pm 3)\%$, without the keynote sessions being included. Similarly, we found that the average number of women who attended sessions with only male speakers was $(23 \pm 7)\%$. We also calculated the average percentage of attendees at the keynote sessions, where there was only one female speaker. The female keynote speaker attracted 10% female attendees, whereas the keynote sessions with male speakers attracted on average $(14 \pm 1)\%$ female attendees. These results show that in this case the number of female attendees was not significantly affected by the gender of the speaker. However, the hypothesis of a correlation of the speaker's gender with the number of female attendees seeks further investigation.

One of the main limitations using the participant observation method for our study was that it was not possible for us to attend all the sessions of some of the conferences, nor to attend more conferences. Therefore, the amount of data obtained with this method is not enough to draw general conclusions on the representation of women in the HPC community. However, these data and their analysis helped with the design of the next steps of our research. Additionally, the validity of the data and the results generated with this method is threatened by the fact that the collection of the data was conducted by humans, whose biased judgement of gender and likely miscounting could affect the analysis and the interpretation of the results.

2.1.2 Programmes and Proceedings

Following the results of section 2.1.1, we wished to investigate further the representation of women at HPC related conferences through the years and detect potential patterns and differences between genders, various roles, and conferences. To do so we gathered historical data of the names of conferences' participants. Based on a method that has been used before in other similar studies [65],[126], we assumed that we can identify the participants' gender by their first name.

2.1.2.1 Data Selection

Firstly, we identified the conferences which would be appropriate for our research. Major international conferences about supercomputers were the main source of data. Such conferences include:

- the **Super Computing (SC) Conference**: the U.S. based annual international conference for HPC networking, storage, and analysis since 1988;
- the **International Supercomputing Conference (ISC) High Performance**: the worlds oldest HPC conference (1986) based in Germany;
- the **International Conference in Supercomputing (ICS)**: the premier international forum for the presentation of research results in HPC systems

since 1987;

- the **Cray User Group (CUG) Conference**: a conference organised by the CUG featuring insightful presentations by HPC experts, as well as technical sessions, tutorials, and demonstrations;
- the **PRACEdays**: one of Europe's most important conferences on HPC in science and industry conference organised annually by the Partnership for Advanced Computing in Europe (PRACE).

We also included smaller conferences that concern HPC related topics and groups of people. The conferences that we chose to take into account for this study are:

- the **Partitioned Global Address Space (PGAS) conference**: the premier forum to present and discuss ideas and research developments in the area of PGAS models, languages, compilers, runtimes, applications and tools, PGAS architectures and hardware features;
- the **EuroMPI conference**: the annual meeting for users, developers and researchers to interact and discuss new developments and applications of message-passing parallel computing, particularly related to the Message Passing Interface (MPI);
- the **Parallel Computing (ParCo) conference**: a conference which occurs every two years and is focused on fundamental aspects of high speed computing, Big Data and parallel computing methods and technologies;
- the **Exascale Applications and Software Conference (EASC)**: a conference focused on issues of applications for exascale and the associated tools, software programming models and libraries.

Programmes and proceedings of these conferences that were available on the internet, were the sources of the required information for this study. We used both programmes and proceedings to cover any data gaps and obtain as much information as possible. Additionally, sometimes the online published conference programmes are preliminary and their information are subject to changes, whereas the proceedings include only the final published papers, which made it essential to receive information from both sources to improve the quality of the results. From the programmes, we could acquire additional data such as the names of the invited and keynote speakers, the workshop and tutorial organisers and presenters, the poster authors and presenters, and the organising committee members. Table 2.2 gathers the programmes and proceedings that were found publicly available online for each conference, at the time of our data selection.

Table 2.2 *Years of programmes and proceedings of each conference that were availble online at the data selection time. Where "n/a" stands for data no available.*

Conference Name	Years of Programmes	Years of Proceedings
SC	2001-2016	1988-2016
ISC	2012-2017	n/a
ICS	1995-2007,2012,2015-2016	n/a
CUG	2012-2017	2012-2017
PRACEdays	2014-2017	n/a
PGAS	2005-2006,2012-2015	2009-2011,2013-2014
EuroMPI	2003,2007,2010-2016	2006-2016
ParCo	1995,1997,1999,2005, 2007,2011,2013,2015, 2017	2007,2009,2011,2013, 2015
EASC	2013,2015	2016

Participants' entries whose last name was accompanied by the initial letter of their first name instead of their full first name could not be used for this study. This had as a result the exclusion of whole years of programmes and proceedings, which will not appear at all in the analysis section. More than 27,000 names formed the final data set (more details on the final data of each conference in the Appendix A.2)

2.1.2.2 Analysis and Results

We firstly identified those entries, where the gender could be assumed by the first name. Then, we tracked down the entries whose first name was not gender-specific or was considered ambiguous. We used a combination of tools and actions to determine the gender of those entries. Initially, we used "Gender API"[1], which is an online platform that determines gender based on first name. Difficulties were encountered mostly for names of Asian and Indian origin, either because they were not included in the database of the online tool, or because some of them can be used for both female and male. In cases where the gender was unknown or the samples of the "Gender API" platform were not enough (under

[1] `https://gender-api.com` *Last accessed: October 2018*

100), we inserted the full name of the individual in an online search engine, and we investigated the results for images and social media profiles, where we could extrapolate their gender from the content. When the results were absent or not conclusive to indicate the gender, we added the name of the institution or the company that the individuals were representing next to their full name to narrow down and specify the results of the search engine. In cases where we did not have the information of the affiliation, we added the word “computing” or “supercomputing” next to the full name of the individual, again to narrow down our search. The names whose gender we were not able to determine were marked as “unidentified” (0.3% of the final data set, Appendix A.2) and left out of the analysis. In the following sections, we present the results of the gender analysis of all the conferences based on the form of contribution.

2.1.2.2.1 Paper Authors and Presenters This group of data includes the authors of the papers presented at each conference as adopted from the proceedings, as well as the authors of the papers and the presenters as adopted from the online programmes. The authors from the proceedings were most of the times available in alphabetical order, and in the vast majority of the programmes the presenters of each paper were not explicitly indicated, but they appeared as authors. That missing information left us unable to compare and draw conclusions on gender differences of the paper presenters in particular. Hence, we gathered a rather satisfactory amount of data for the authors of the papers, as this information was provided by the majority of the available programmes and proceedings.

Figures 2.2 and 2.3 illustrate the average percentages of female and male paper authors as found from the programmes and proceedings of each year. Overall for all the conferences, the number of women is consistently lower than men’s for all the years considered, with no significant differences between programmes and proceedings. With EuroMPI displaying the lowest percentage and PGAS the highest percentage of female authors, we could say that on average women’s representation as paper authors at HPC related conference comes up to 9%.

For ICS 2006, only the names of the presenters, and not all the authors, were available. We found that only 3% of the presenters of ICS 2006 were women; a percentage which is even lower than the average percentage of female paper authors that we found from the examined years of ICS (9%) (Figure 2.2).

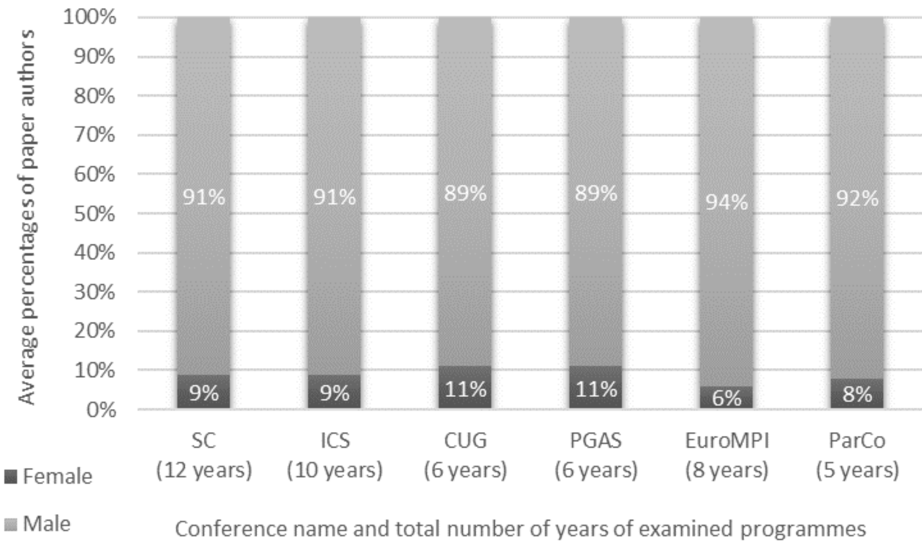


Figure 2.2 *Percentages of female and male paper authors averaged over years for each conference, according to available programmes ($SE \leq \pm 1\%$, except for PGAS where $SE = \pm 2\%$).*

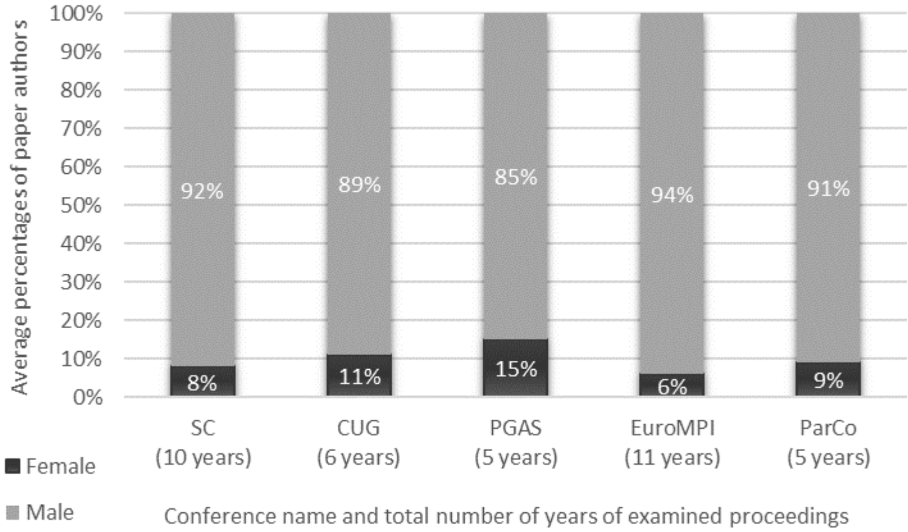


Figure 2.3 *Percentages of female and male paper authors averaged over years for each conference, according to available proceedings ($SE \leq \pm 1\%$, except for PGAS where $SE = \pm 3\%$).*

For EASC, we found data only for three years (two programmes and one proceedings), and we did not include them in the aggregate charts, where we included conferences with data available for 5 years and over (Figures 2.2 and 2.3). At EASC 2013, women made up 10% of the 164 paper authors, and in

2015 there were 12% women of a total of 130 paper authors. These percentages follow our findings from observation at EASC 2015 in section 2.1.1. From the proceedings of 2016 we found that women were making up 17% of the total of only 24 authors, which shows a small increase from the previous years; however, we did not know why the total number of authors of the proceedings 2016 is much smaller than the programmes of the previous years.

Additionally, for ISC we found data for “speakers”, as they were labelled in the source from where we adopted data for ISC (Appendix A.2). “Speakers” included all types of participation and contribution, from exhibition hall and conference speakers to poster presenters and tutorial instructors, hence not exclusively the speakers of scientific or plenary sessions. The number of female “speakers” appeared to be similar for every year, with an average 8% for the years 2012-2017, which seems to agree with our previous finding in section 2.1.1.

For PRACEdays, only the programme of 2014 included all the authors of the presented papers (total number of paper authors N=43), with women making up 7% of them. For the years 2015-2017, there were only the names of the presenters available. Table 2.3 shows the percentages of women who presented a paper at the conferences in 2016 and 2017, including the data from 2015 which we examined in section 2.1.1. There is a significant increase of the female paper presenters from 4% in 2015 to 22% in 2016. In 2017 there is a small drop to 17% from 2016, but this percentage is still higher than the 4% of 2015. Even though men still made up the majority of the presenters at PRACEdays, the considerable change of women’s representation as presenters from 2016 onwards seems promising. Since PRACEdays was first organised in 2014 and there are no available data of the presenters from earlier than 2015, future data comparison through the years is recommended, when more data are available, for drawing more accurate conclusions.

Table 2.3 *Total number of paper presenters (N) and percentage of female paper presenters (F%) for years 2015, 2016 and 2017 of PRACEdays.*

Year	Paper Presenters (N)	F %
2015	23	4%
2016	27	22%
2017	29	17%

2.1.2.2.2 Invited and Keynote Speakers In section 2.1.1 we described the importance of keynote speakers. Invited and keynote speakers are usually chosen by the organising committee as individuals with important research background, position or influence in the respective field or topic of the conference. As we mentioned earlier (1.1.2.3), studies have shown that men are more likely to be invited to talk at a conference than women [62], [63]. Additionally, women are more likely to decline an invitation to speak, for various and complex reasons [62]; for example, beacuse they find it more difficult to travel due to childcare responsibilities [140] or because they lack self-promotion [141], confidence in their scientific ability [142], [143] and funding [57].Therefore, we examined the female and male invited and keynote speakers separately from the paper presenters and authors. Data for invited and keynote speakers were available in some years' programmes of SC, ICS, PGAS, EuroMPI and ParCo. From now on where we refer to invited speakers in this section, we mean invited and keynote speakers together, unless otherwise stated.

Figure 2.4 shows the percentages of female and male invited speakers of SC conferences. There is an increase of female invited speakers since 2012 with a remarkably high percentage approching a 50/50 balance in 2016. As we mentioned in Chapter 1 (1.2), since 2015, SC has focused on diversity and inclusivity by establishing a diversity committee, gathering and analysing diversity related data (such as gender) and setting future targets to promote diversity[2]. We believe that this initiative has affected and increased women's representation in some categories of contribution of SC, such as invited speakers.

At ICS there were no female invited speakers, except in 1995 where women made up 25% of a total of 4 invited speakers, and in 2015 where women made up 33% of a total of 3 invited speakers. PGAS and ParCo demonstrated a similar picture. There was only one woman who made up 25% of the 4 invited speakers at PGAS 2014. For ParCo, there were two women making up 40% of the 5 invited speakers in 2007, and only one woman making up 13% of a total of 8 in 2017. For the rest of the available years, there were no women invited speakers, so we cannot draw conclusions for a certain trend. We can only confirm that overall there were more male than female invited speakers.

At the majority of EuroMPI conferences of the examined years, there were no women invited or keynote speakers. In 2008 women invited speakers made up 14% of the 7 invited speakers, in 2014 they made up 50% of a total of 4, and in 2016 they made up 29% of the 7 invited speakers. Even though there seems to be

[2] `sc15.supercomputing.org/conference-program/technical-program/sc15-diversity-committee-focused-events.html` *Last accessed: October 2018*

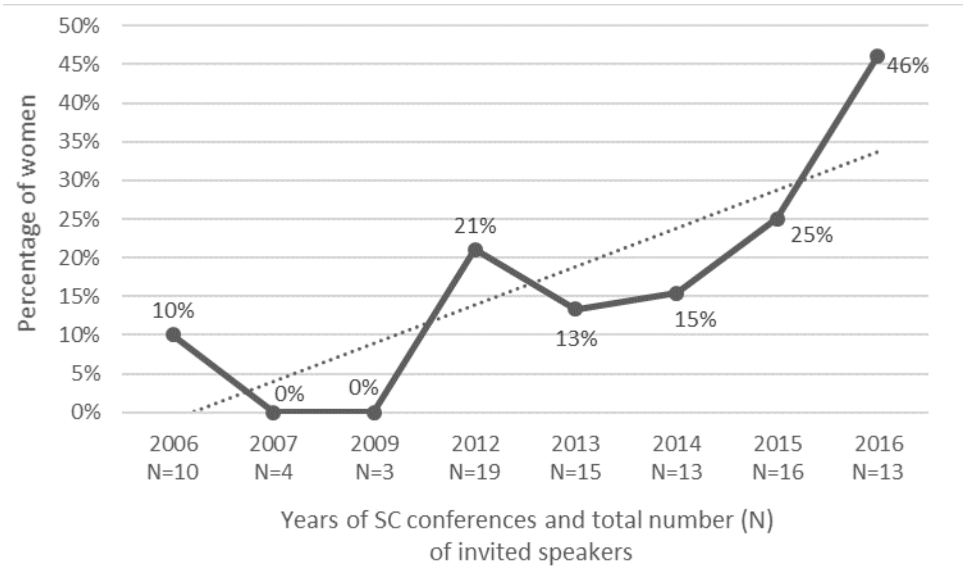


Figure 2.4 *Percentages of female invited speakers at SC conferences through the exmined years.The trendline shows the upward trend with a significant increase in 2016.*

an increase of female invited speakers recently, the years with 0% representation in between do not support an upward trend in female invited speakers.

2.1.2.2.3 Poster Authors and Presenters Research shows that men are more likely than women to request and receive a talk rather than a poster [61], and one could assume that there would be more female poster authors and poster presenters. Unfortunately, most of the programmes we used in our study did not highlight the names of the poster presenters, but they provided the names of all the authors of the posters.

PRACEdays' programmes were the only ones which provide the names of the poster presenters. For the years 2014-2017, we found an average of $(17 \pm 4)\%$ of female poster presenters. Table 2.4 shows the lowest percentage being in 2014 (8% of a total of 37 poster presenters) and a remarkable increase of 10-20% since 2015. This significant difference through the years is consistent with the increase of the PRACEdays paper presenters, that we discussed earlier (2.3), which however firstly appeared in 2016.

We also gathered data for poster authors from the programmes of SC, PGAS, EuroMPI and EASC. For EASC, we only have available data for 2013, where women make up 15% of the 20 poster authors. For PGAS we found data for

Table 2.4 *Total numbers of poster presenters (N) and percentages of female poster presenters (F%) for years 2014, 2015, 2016 and 2017 of PRACEdays.*

Year	Poster Presenters (N)	F %
2014	37	8%
2015	38	28%
2016	32	16%
2017	22	18%

2006 with 0% women out of 22, for 2013 with 21% women of a total of 14 poster authors, and for 2015 with 7% female poster authors of a total of 14. Figure 2.5 demonstrates the average percentages of poster authors for SC and for EuroMPI.

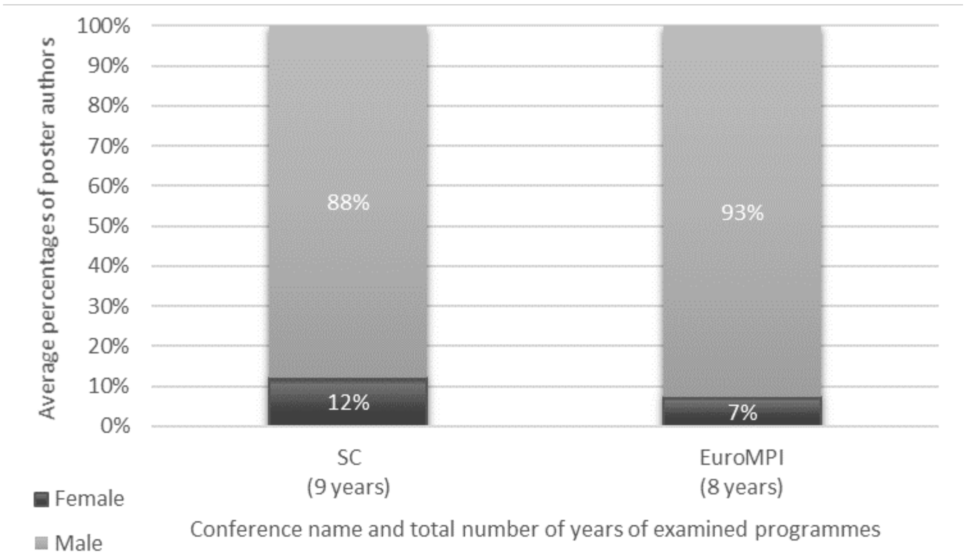


Figure 2.5 *Average percentages of female and male poster authors at SC and EuroMPI conferences for the examined years (For SC SE=±1% and for EuroMPI SE=±2%).*

Overall, men make up the majority of the poster authors and poster presenters. However, women appear to participate as poster authors slightly more than as paper authors at the conferences we examined. EuroMPI displays the lowest average of female poster authors, in line with the percentages of the female paper authors which we discussed earlier (2.1.2.2.1).

2.1.2.2.4 Workshop and Tutorial Presenters and Organisers We also examined the contribution of women in interactive training sessions at conferences, such as workshops and tutorials. Proposals for workshops and tutorials are usually submitted by instructors or organisers and are selected by the conference organising committee. There is no previous research on this aspect of a conference, but we hypothesised that it would probably demonstrate a similar picture as the one for paper presentations and authorship. Data for workshops and tutorials were only found for SC and EuroMPI.

Only one woman was a tutorial presenter at EuroMPI of a total of 33 presenters, that we found in the available data. Figure 2.6 shows the average percentages of female and male presenters and organisers of tutorials and workshops for the years of SC conferences.

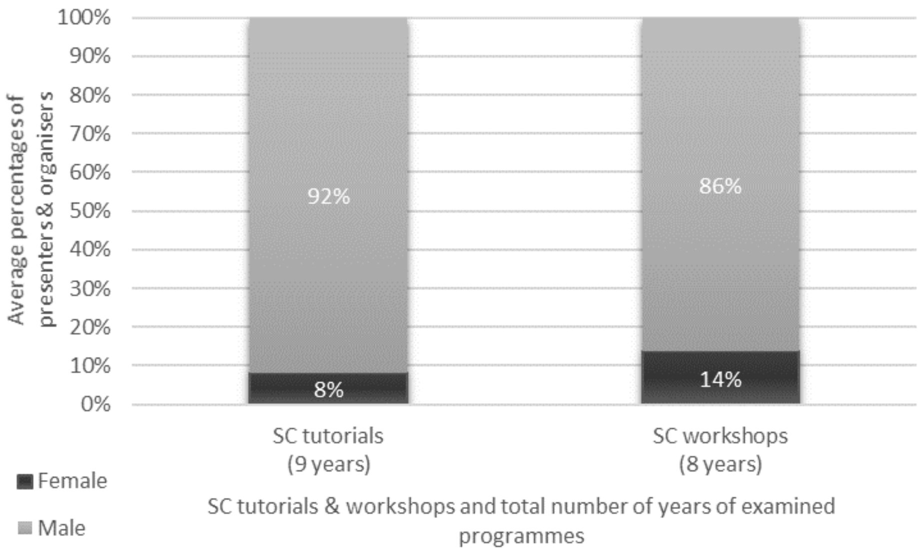


Figure 2.6 *Percentages of female and male tutorial and workshop presenters and organisers at SC conferences for the exmained years (For tutorials SE=±2% and for workshops SE=±1%).*

Once again, women were outnumbered 8-11 to 1 by men. However, women displayed a slightly higher percentage as presenters and organisers of workshops, with a remarkable increase of their participation since 2014 onwards (Figure 2.7). We could suggest that this rise was due to the estamblishment of the WHPC network in 2014 and its activity at the SC conferences since then, as well as to the SC diversity committee's work discussed in section 1.2.

2.1.2.2.5 Committee Members Research shows that at a conference with more female members in the organising committees, there might be a bigger number of female speakers [100]. For our study, we found the information of female and male members of the programme committee only for some years of SC,

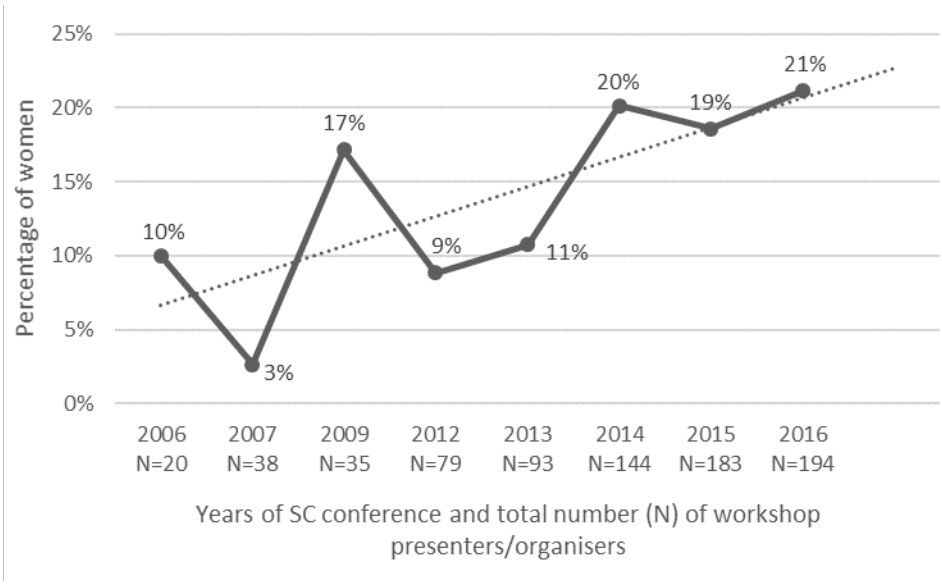


Figure 2.7 *Percentages of female workshop presenters and organisers at SC conferences through the examined years. The trendline shows the upward trend especially since 2014.*

CUG and EuroMPI conferences. We have also already presented the data for the committee of PRACEdays 2015 at section 2.1.1.3. Unfortunately, we could not examine if there was a correlation with the number of female speakers, because for most of the conferences the names of the presenters were not highlighted among the names of the authors of the papers. However, we inspected if the numbers of female committee members played any role in the number of women who contributed in ways other than presenting a paper or a keynote address.

We found that the average percentage of female committee members was (48±4)% for the six years of SC where this information was available, and that it was close to the average percentage of men in the committee, which was (52±4)%. However, this almost 50/50 gender balance does not justify the lower numbers of women at the other aspects of the SC conferences, escpecially when for some of the years there were more women than men in the committee (2013 and 2015, Figure 2.8).

For CUG, details of committee members were available only for 2016 with a total of 32 members, and for 2017 with a total of 28 members. Even though the percentage of women in the committee of CUG 2016 and 2017 came up to 28%, it did not seem to have affected the numbers of women at the other categories of conference participation which we investigated.

EuroMPI had only one female member in the programme committee in 2007 (14%

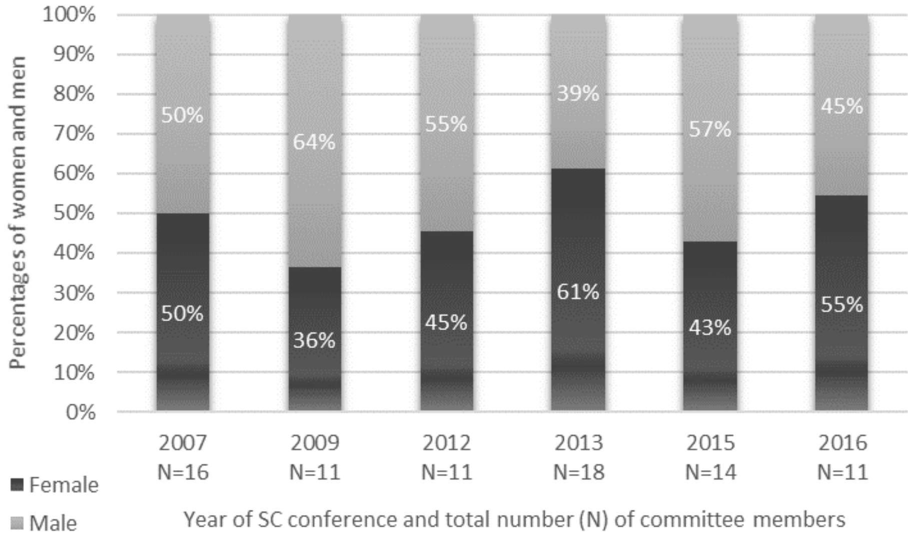


Figure 2.8 *Percentages of female and male committee members at SC conferences through the examined years.*

out of 7 members), out of a total of 34 members through the eight examined years; this might explain the very low numbers of female authors, invited speakers, poster and tutorial presenters that we found. EuroMPI presented the lowest numbers of women's participation in all of the aspects of the conference. As it is presented in Section (2.2), all courses related to MPI are categorised into intermediate and advanced levels; hence, MPI is considered as a more advanced field of the broader HPC area. We conjecture that this might be a reason for the low numbers of women at EuroMPI, and we encourage further research.

2.1.3 Summary and Discussion

We discussed the results from the analysis of data gathered by attending conferences where we counted and observed the attendees, as well as by collecting historical data with lists of names of conferences' participants which we found publicly available in conferences programmes and proceedings, where we assumed the gender of the participants by their first name.

The most significant results from the analysis of the aforementioned data were:

- Women were fewer than men in all the categories of conference participation that we examined. On average, we found: 9% paper authors, 0-1 female invited speakers for every 4 male invited speakers, 7%-12% female poster authors, 1 female workshop or tutorial presenter for every 8-11 male presenters.

- Women poster presenters and poster authors were slightly more than women paper presenters and authors for the conferences we had data for both categories. For example, on average at SC there were 9% female paper authors and 12% female poster authors.

- SC conference showed some recent cases of increase in numbers of women, and cases of gender balance (46% female invited speakers in 2016 in comparison to 13% in 2013, 21% female workshop presenters/organisers in 2016 in comparison to 11% in 2013). We believe that the appointment of a diversity committee by SC and the initiatives of the WHPC network were responsible for these changes at SC conferences.

- We noticed a significant difference in the number of women at ISC 2015, where female organisers/moderators increased from 9% to 19%, when we added the WHPC female representatives. This fact leads us to suggest that a network which support women can affect their representation and contribution in a community.

- EuroMPI conference presented the lowest numbers of female participation. For the examined years we found: 6% female paper authors, 7% female poster authors, 0%-1% female invited speakers, 1 female tutorial presenter out of a total of 33, 1 female committee member out of a total of 34.

- PRACEdays displayed a significant increase in female paper presenters (from 4% in 2015 to 22% in 2016) and poster presenters (from 8% in 2014 to 28% in 2015). Since PRACEdays is a new conference (since 2014), we encourage further observation of the numbers of women at the conferences of the next years, so that a likely pattern or conclusion can be drawn.

Taken together, these findings suggest that women are underrepresented in all the categories of conference participation, with a hope rising from the work of the WHPC Network and diversity committees.

However, due to the lack of previous studies and of reported demographics of the actual gender ratio of the population of the HPC community, the results of this section can only present the ratio of female and male conference participants without concluding that this accurately mirrors the demographics of the community, and that women of the community are underrepresented at the conferences of the field.

Even though we tried to include as many conferences and gather data from as many years of programmes and proceedings, information was missing in some cases (for example, whole years with entries which were excluded because of missing the full first name, or unavailable data). However, the number of missing or unidentified data for the years which formed our final dataset was too small to affect the results (participants' entries with initials instead of full first name, which were excluded, and those whose gender was not identified were $< 1\%$ of the total entries of the final dataset).

Finally, the validity of our results in this section is threatened because of the determination of the gender of the participants considering only one factor: their first name. We recognise that there might be cases where we might have wrongly assumed the gender of a participant or it was wrongly defined by the software[3], which however claims 94% accuracy of results.

2.2 Courses

The UK National Supercomputing service, ARCHER, offers free HPC related and programming training courses across the UK. We worked with ARCHER's team and obtained access to the participants' responses to the feedback forms of the various courses for 5 years (2013-2017). The respondents to those feedback forms self-reported their gender as a response to one of the questions. From this data set we were able to calculate the participation of women and men to the courses.

According to the courses section of ARCHER's website[4], the courses are classified into three levels: Introductory or Level 1, which covers introductory topics and require basic computer literacy; Intermediate or Level 2, which requires some programming knowledge or experience using some HPC techniques; and Advanced or Level 3, which covers the most advanced topics and consists only of courses suggested for "developers".

[3] `www.gender-api.com` *Last accessed: October 2018*
[4] `www.archer.ac.uk/training/courses/` *Last accessed: October 2018*

The titles of the courses of each level that are suitable for “users” are:

Level 1

- Hands-on introduction to HPC
- Scientific computing
- Online introduction to HPC
- Online ARCHER documentation
- Software carpentry
- Data carpentry

Level 2

- Data analytics with HPC
- Scientific Programming with Python

The titles of the courses of each level that are suitable for “developers” are:

Level 1

- Introduction to modern Fortran

Level 2

- Message-passing programming with MPI
- Shared memory programming with OpenMP
- Object-oriented programming with Fortran
- GPU programming with Cuda
- Practical software development

Level 3

- Writing scalable parallel applications with MPI

- Advanced MPI
- Programming the KNL
- Advanced OpenMP
- Single node performance optimisation
- Efficient use of ARCHER and KNL
- Efficient parallel IO on ARCHER
- Performance analysis workshop

Figure 2.9 presents a diagram which suggests possible routes through the recommended courses for “users” and “developers”, which can lead to “expert” level.

2.2.1 Data Selection

The feedback responses for a total of 25 courses for the year 2015, 25 courses for 2016 and 17 courses for 2017 formed the final data set for analysis. We did not include some courses whose level we could not identify and which do not exist anymore in the ARCHER’s list, as we were informed by ARCHER’s team. Additionally, the feedback forms for the courses of the years 2013 and 2014 did not include the question for the participants’ gender, so the data of these two years were not useful for this study and they were not included in the analysis’ data set.

The initial data set included data of two courses organised by the WHPC network in 2015, and one course for Life Scientists in 2017, which were all labelled as Level 1 courses (Hands-on introduction to HPC, Software Carpentry). Since these courses were specially targeted to women, and Life scientists (who present a better gender balance 1.1.1), the numbers of their participants considerd as “outliers” [133]. We did not include the data of these courses in the first part of the analysis’ section, as we believed that they could affect the overall results of this study. In the last part of section 2.2.3, we present the analysis including the data of these three “special” courses, and we explain their impact on our results.

To quantify the participation of women and men at the various courses, we located the responses of those who identified their gender as female and of those who

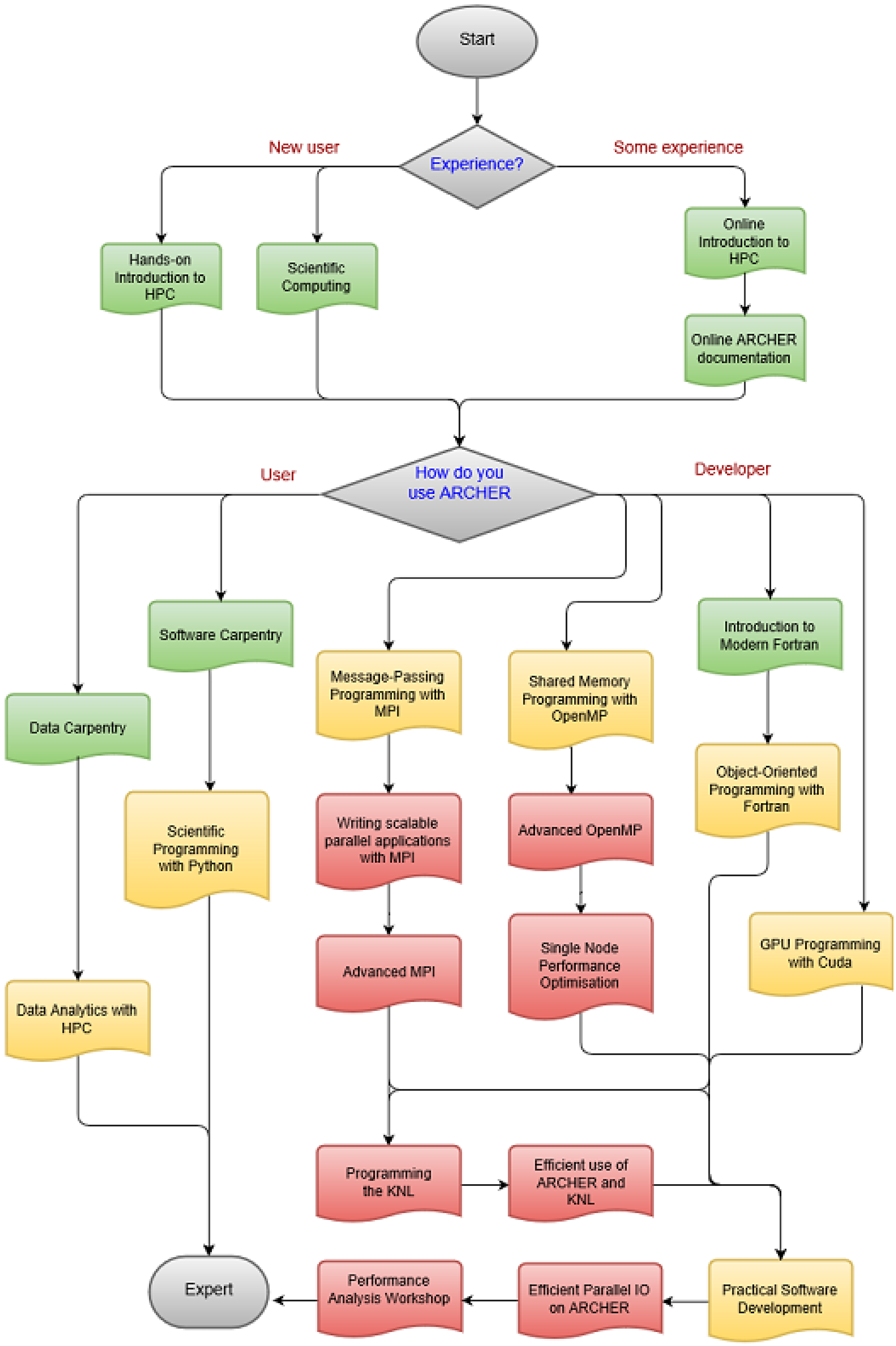


Figure 2.9 *Diagram showing possible routes through available training courses by ARCHER (Source: ARCHER).*

Table 2.5 *Total number of respondents (N) to the feedback forms, number of respondents who self-reported their gender as female (F) and number of respondents who self-reported their gender as male (M), for each year.*

Year	Respondents (N)	Female	Male
2015	238	56	169
2016	238	64	150
2017	162	34	113

identified it as male, for each course of every year. We did not take into account the responses of those who either did not reply to the gender question or chose the "prefer not to say" option. For a semi-qualitative analysis of these data, we also collected the participants' responses to other questions which we then correlated with their gender and we will discuss in section 2.2.3.

We collected the responses to the gender question for each course of each year and we calculated the numbers of those who self-reported as women and those who self-reported as men. Table 2.5 displays the total number of the participants who respondend to the feedback forms, as well as the numbers of female and male respondents for each year.

2.2.2 Women's Participation

Figure 2.10 displays a graph with the percentages of female and male respondents from the total number of respondents when we excluded the number of those who did not report their gender, for each year. The percentages of the female participants range among values two to three times smaller than those of the male participants for all years.

2.2.3 Further Analysis and Results

For a qualitative approach and to identify any gender differences in behaviours and perceptions, in this section we present the results we obtained when we correlated the respondents' gender with: the levels of the courses they attended, the percentages for those who potentially are HPC "users" or "developers" according to ARCHER's diagram (Figure 2.9), their responses to the question asking how difficult they found each course, and their responses to discipline and

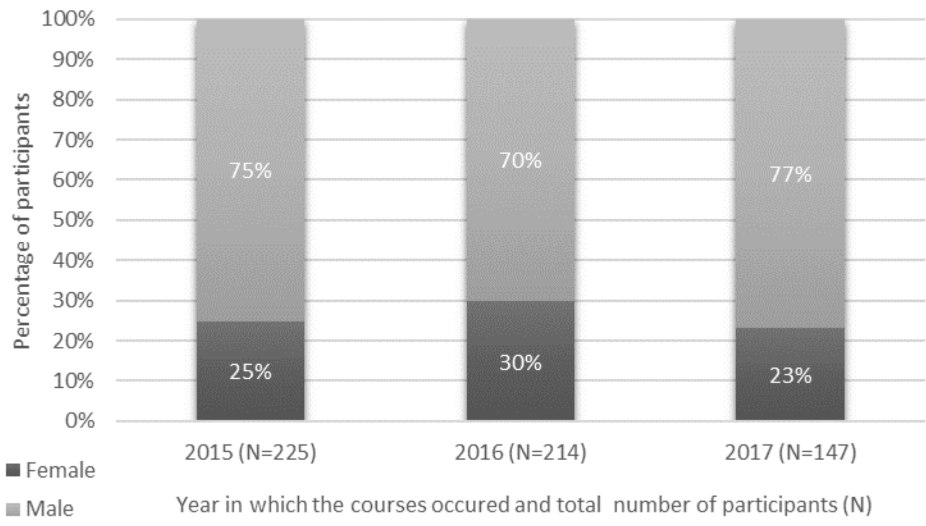


Figure 2.10 *Percentages of female and male participants at the ARCHER training courses of all levels for the years 2015, 2016 and 2017.*

sector of employment.

2.2.3.1 Levels of Courses

In Table 2.6 we see how many courses of each level for each year we used for the analysis in this section.

Table 2.6 *Numbers of courses of each level used for analysis by year.*

Year	Level 1	Level 2	Level 3
2015	7	11	5
2016	9	10	6
2017	6	10	3

We grouped the attendees from all the courses of each year by level, and we calculated and compared the average numbers of female and male participants to each level for every year, which are shown in Table 2.7.

From Table 2.7, it is obvious that men outnumbered women at the majority of the courses. We noticed that overall, for every year the average number of female participants dropped from Level 1 towards Level 3, whereas the average number of male participants was similar through the levels. In 2017, male participants dropped slightly from Level 1 (7) towards Level 3 (4), whereas female participants did not display any trend through levels. This small difference in numbers,

Table 2.7 *Average numbers of female and male participants to all courses of each level by year ($SE \leq \pm 1$).*

Year	Gender	Level 1	Level 2	Level 3
2015	Female	4	2	1
	Male	8	7	8
2016	Female	6	1	0
	Male	6	6	6
2017	Female	3	1	2
	Male	7	6	4

especially for men, encourages further investigation of participants' number of courses for future years.

We then calculated the average numbers of female and male respondents of all years for each of the three levels (Figure 2.11).

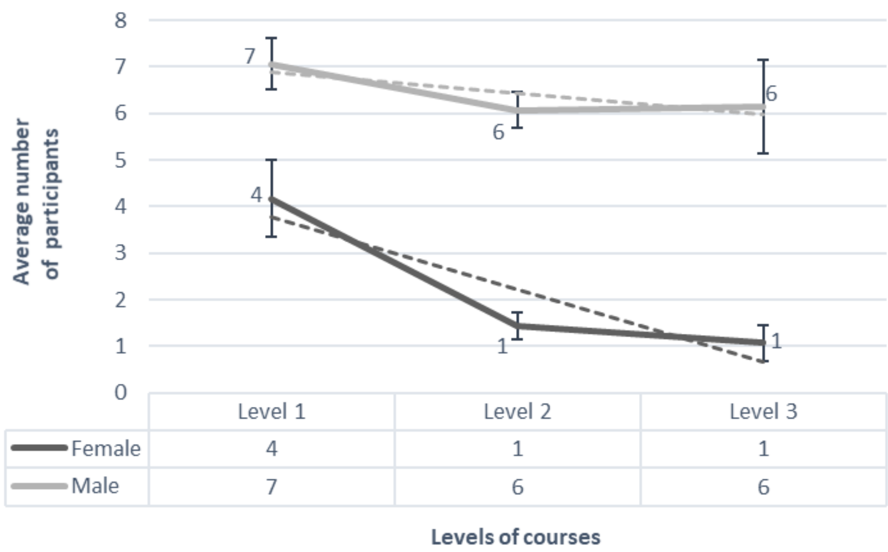


Figure 2.11 *Average numbers of female and male participants of all courses for the years 2015-2017 by level.*

The lines in the chart in Figure 2.11 show that the total average number of participants of all courses of all years dropped from Level 1. We see a statistically significant difference in the average numbers of female participants at all levels when comparing with the respective averages of male participants ($p < 0.05$). Additionally, the trendline of the female attendees shows a deeper drop from Level 1 towards Level 3 than the male attendees' trendline.

2.2.3.2 Difficulty of Courses

The last question of the feedback form of all the courses was asking the participant how difficult they found the course. The response options were: very easy, quite easy, moderate, quite hard, very hard. To facilitate the analysis of these data, we converted the words of these options into a 1 to 5 scale where 1 stands for very easy, 2 for quite easy, 3 for moderate, 4 for quite hard and 5 for very hard.

To determine how difficult the female and male respondents found each course, we correlated the reposonses to this question with gender, and we calculated the weighted average, where values are the numbers of the scale which stand for difficulty ratings, and weights are the number of responses for each number of the scale. Table 2.8 shows the average difficulty of each course level for each year according to female and male participants' rating.

Table 2.8 *Average difficulty of each course level according to the ratings of female and male respondents by year ($SE \leq \pm 0.3$). The difficulty scale ranges from 1= very easy to 5=very hard.*

Year	Gender	Level 1	Level 2	Level 3
2015	Female	2.2	2.9	3.3
	Male	2.8	2.8	2.7
2016	Female	2.4	2.8	4.0
	Male	2.5	2.8	2.5
2017	Female	2.9	3.1	2.3
	Male	2.6	2.8	3.0

In Table 2.8 we observed that women's difficulty rating increased wih the increase of the level of the courses for the years 2015 and 2016. Interestingly, in 2017 women found Level 1 courses harder than Level 3 courses, and Level 2 courses the hardest. On the other hand, men rated all levels for all years in the same way: close to "moderate" (2.5-3).

We also calculated the total average difficulty of each level for all years, which is shown in Figure 2.12. In this chart, we can see that female and male participants rated the levels of the courses similarly, with small, but not significant differences ($p > 0.05$), between genders for each level. As in Table 2.8, women's difficulty rating increased gradually with the levels, whereas men found the courses of the three levels equally hard.

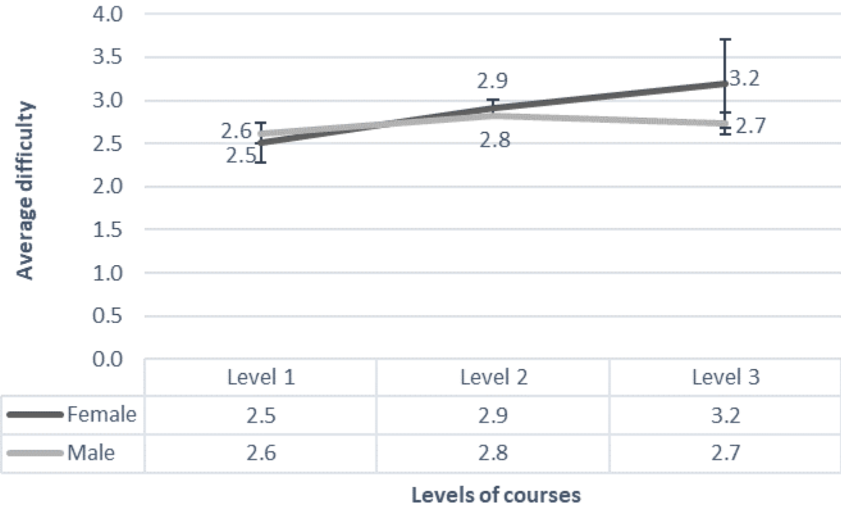


Figure 2.12 *Averages of the course difficulty rated by female and male participants by course level for all years. The difficulty scale ranges from 1= very easy to 5=very hard.*

2.2.3.3 Users and Developers

According to ARCHER's diagram (Figure 2.9), the choice of each course shows how the participant potentially uses HPC. According to the courses the respondents attended, we labeled them as "users" or "developers". To detect any potential gender differences in the use of HPC, we then calculated the percentages of all the female and male "users" and "developers" (Figure 2.13). In this graph, we see that male participants were more likely to be "developers" than "users"; whereas female participants did not appear to have a significant preference, but only a slight tendency towards "users".

2.2.3.4 Discipline and Job Sector

By correlating the respondents' gender with their responses to the open-ended question on the discipline of their work or studies, we wanted to identify if the background of the participants at HPC related courses is mostly STEM or non-STEM, and if there were any differences in the background of female and male participants.

Not all courses' feedback forms included the question on discipline. There were 12 courses in 2015, 10 in 2016, and 1 in 2017, which included ths question. The

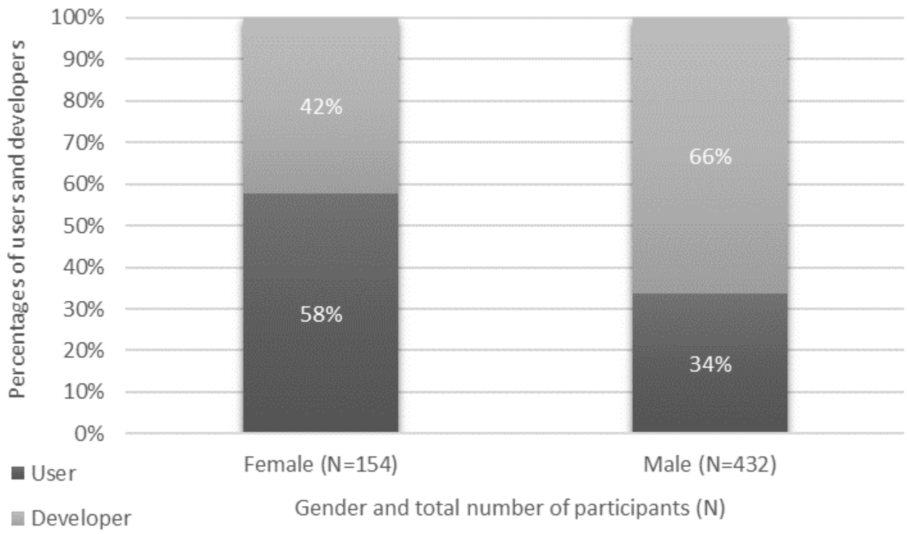


Figure 2.13 *Percentages of female and male participants who use ARCHER as "users" or "developers" according to the courses which they attended, and based on the ARCHER diagram (Figure2.9).*

question was open-ended, which means that the participants had the freedom to write their own response. Unfortunately, not all the respondents answered this question, and some of them chose to write "other" as a response. On average 85% of the participants named a STEM discipline, with no significant differences between the two genders ($p > 0.05$). The remainder 15% included those who did not respond to this question, those who responded "other" or a non-STEM discipline (lists of the disciplines grouped into STEM and non-STEM are available in Appendix A.3). Therefore, most people that received HPC related training came from a STEM background.

Similarly, there were 11 courses in 2015, 15 courses in 2016, and 16 courses of 2017, that included the question asking the participants to choose between "academic" and "industry" as their job sector. We correlated the participants' gender with their responses to the question on job sector, to identify any potential differences between participants working in different environments regarding gender. The numbers of respondents who chose "industry" as their answer and of those who did not respond to this question were very small. On average 93% of the participants worked in academia, which suggests that people who work in academia are more likely to attend HPC related training courses, regardless gender ($p > 0.05$).

2.2.3.5 Courses for Women and Life Scientists

As we mentioned in the Data Selection section 2.2.1, so far, our analysis and results did not include the data of two Level 1 courses for women, that took place in 2015, and one Level 1 course for Life scientists, that took place in 2017. Here, we present how our results changed when these data were included in the analysis. We expected that adding these data could change the results for Level 1 courses, the results for the years 2015 and 2017, and all the total results for levels and years when we compared gender with other variables, such as difficulty,discipline and job sector. The numbers of the participants of these "special" courses can be seen in Table 2.9. These numbers suggest that women are more likely to attend courses that are specially targeted to women, where they feel more comfortable (1.1.4.2). These numbers also verify the fact that in biomedical related subjects, there is better gender balance than other STEM disciplines (1.1.1).

Table 2.9 *Numbers of female and male participants at the three Special Courses, Hands on Introduction to HPC (HoI) and Software Carpentry (S.C.) organised by WHPC in 2015, and Hands on Introduction for Life Scientists (HoI Life Science) in 2017.*

Gender	HoI (WHPC)-2015	S.C. (WHPC)-2015	HoI Life Science-2017	Total
Female	11	24	9	44
Male	1	1	8	10

When we added these numbers in the analysis, the total percentages of female and male participants change, as they are shown in Figure 2.14. The percentage of female participants for the year 2015 change from 25% to 35%, and for 2017 from 23% to 26%, which do not show a significant difference. However, when we calculated the average numbers of female and male participants by level (Figure 2.15), the numbers for Level 1 changed, and the difference between female and male is not statistically significant anymore ($p > 0.05$).

There was also a noticeable difference on the percentages of female "users" and "developers", when we considered the data of the "special" courses. Since all the "special" courses were placed in the "user" route in ARCHER's diagram (Figure 2.9), it was expected that the percentages of the female "users" and "developers" to change. These percentages can be seen in Figure 2.16, which shows that female participants were more likely to be HPC "users"; a result which is different from the result of the initial analysis, where female participants did not show significant preference. The result for male participants remained the same; that they were

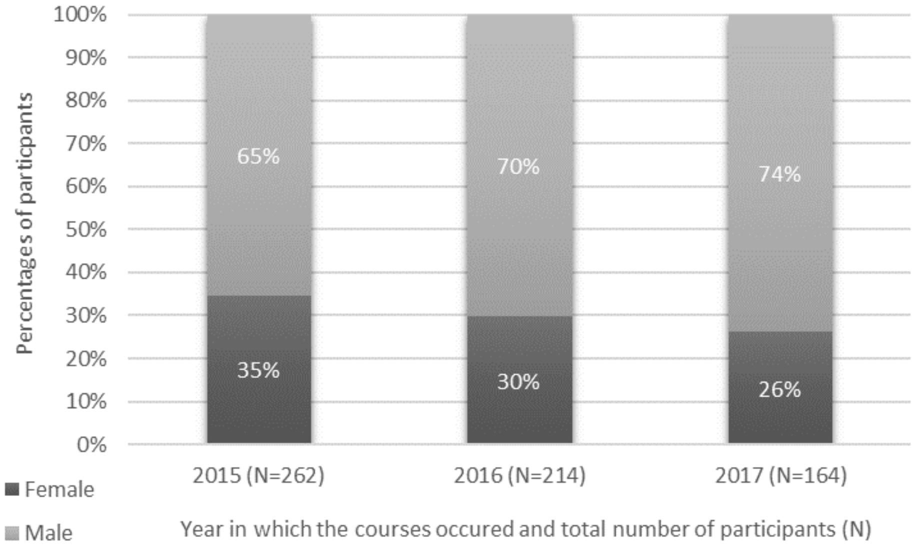


Figure 2.14 *Percentages of female and male participants at the ARCHER training courses of all levels for the years 2015, 2016 and 2017, including the "special" courses' data.*

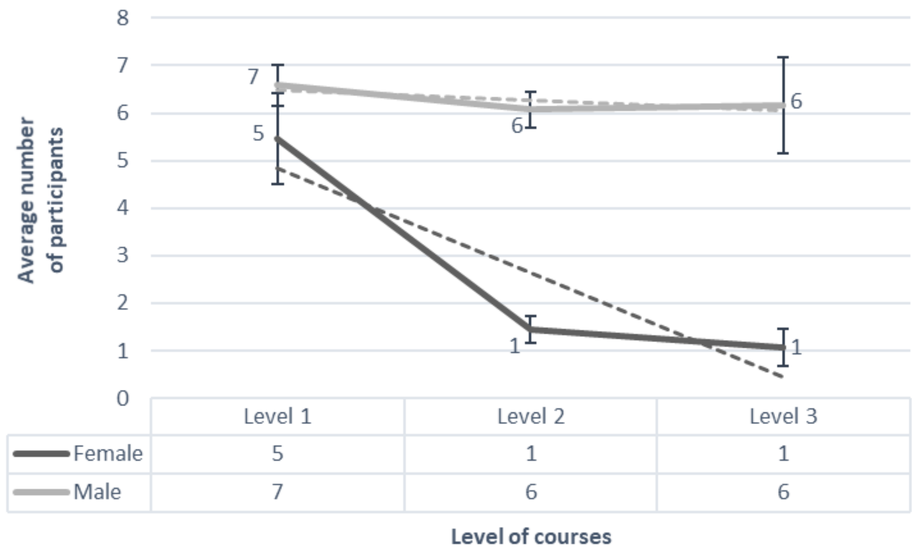


Figure 2.15 *Average numbers of female and male participants of all the courses for the years 2015-2017 by level, including the "special" courses' data.*

more likely to be HPC "developers".

The addition of the "special" courses' data did not change remarkably the rest of the percentages and averages that resulted from the correlation of gender with course difficulty, discipline and sector.

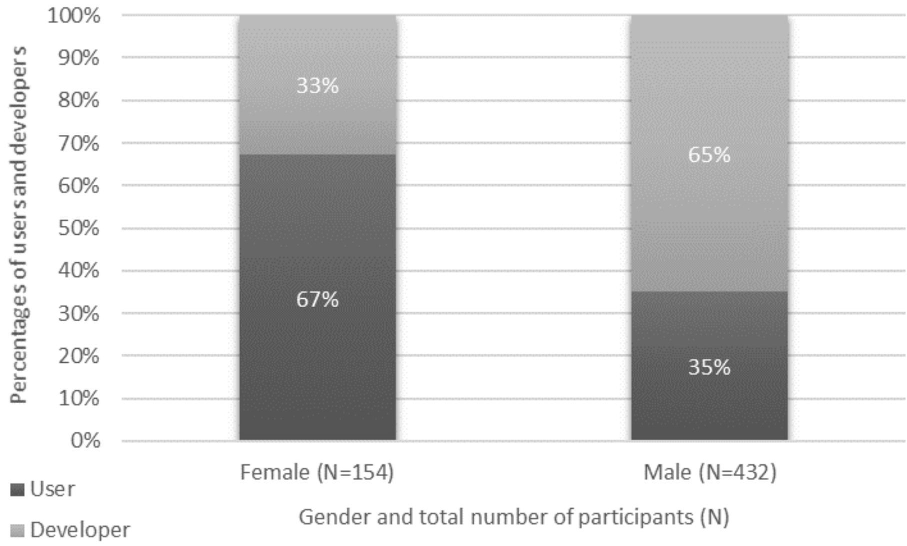


Figure 2.16 *Percentages of female and male participants who use HPC as "users" or "developers" according to the courses which they attended, including the "special" courses, and based on the ARCHER diagram (Figure2.9).*

2.2.4 Summary and Discussion

The main findings from the analysis of the responses to the feedback forms of courses ran by ARCHER are:

- Women were outnumbered three to one by men at all levels and years of the courses we examined.
- Women had a much higher attendance frequency at Level 1 courses (average number of female participants=4) than Level 2 (average number of female participants=1) and Level 3 (average number of female participants=1).
- Women's overall difficulty rating increased gradually with the levels of the courses (2.5 for Level 1, 2.9 for Level 2, 3.2 for Level 3).
- Women were identified more as "users" according to the courses they attended (58%-67% female "users").
- The female participants at courses in 2015 increased from 25% to 35% when we added the participants of the "special" courses, which supports similar finding of the conferences section (2.1.3).

- The majority of the courses' attendees for both genders come from a STEM discipline (85%) and work in academic sector (93%).
- Women attended much fewer advanced level courses (average number of female participants at Level 3=1), into where the majority of MPI related courses were categorised. Adding this finding to the results about women's participation at EuroMPI conferences (2.1.3), made us suspect that the numbers of women are much lower at EuroMPI conference because MPI is considered as an advanced or more specialist topic of the broader HPC field.

In conclusion, women are underrepresented at all levels of training courses, particularly at higher levels, and they seem to identify more as "users" than "developers". In a similar way to our conference results, courses organised by the WHPC Network seem to attract more female attendees.

The data used for this analysis were produced by feedback responses for courses that took place in the UK, and generalised conclusions generated from these results for the worldwide HPC community carry threatened validity. Additionally, we do not know the actual number of the people who attended each course, hence the results and conclusions we drew from this analysis refer only to the participants who filled in the feedback form, and they might not reflect the actual population of participants nor their behaviour.

Chapter 3

Experiences, Views and Behaviours of the HPC Community: Semi-quantitative Approaches

The use of surveys as a research method can produce data for both quantitative and qualitative analysis. The number of times the most popular responses to a question appear can be counted and considered as data for quantitative analysis; additionally, the interpretation of the content of these responses can be part of qualitative analysis. The open-ended questions, unlike the forced-choice or multiple-choice questions, give the opportunity to the respondents to answer with a free text, and they are mainly used for a qualitative explanation of the research questions [132].

With this semi-quantitative approach to our study, we aimed to record experiences of women and men in the workplace and in conferences of the HPC community; to identify potential differences in views and behaviours; and to record best practices to improve the gender balance within the community. To achieve these objectives, we firstly used the responses from a survey that was designed and conducted by the Software Sustainability Institute (SSI), as our first source of both quantitative and qualitative data. This survey was further used as a guide to help us design our own survey. We also designed a short questionnaire for the participants of a "special" HPC course organised by the Women in HPC (WHPC) Network, and used it to pilot test our survey and the forthcoming interview method. We then designed and conducted a survey for the HPC community according to the guidance, results and feedback we obtained from the previous two approaches. Lastly, we designed and carried out a short survey to further examine one of the many and main outcomes of the HPC community survey, which concerned the use of HPC for non-STEM research.

3.1 Ethical Requirements

To ensure that the surveys designed by us were ethically sound, we provided relevant information in the invitation for participation. This included: the nature of our study, the purpose of the surveys, the use of data, and contact information.

In addition, at the beginning of our surveys, we requested an informed consent to state that participation was voluntary [132]. This consent also included brief information about the study, the use of data and privacy policy, and contact details.

3.2 Software Sustainability Institute (SSI) Survey

Software is used everywhere and by everyone in research and academia. However, most of the times it is not really known which type and how much software has been used on a project. Also, it is fairly common for researchers, who do not know how to develop their own software, to ask for the help of the experts, the research software engineers. It would be very helpful for the researchers to know better the software used for their research, to take advantage of its full potential and be aware of its probable limitations [144]. To achieve this, the researchers should get the appropriate training and learn how to develop their own software and adjust it to their research needs.

The Software Sustainability Institute (SSI) issued a survey[1] in 2014 to UK Russell Group[2] University staff on research software usage. In that survey, research software was defined as: *"Software that is used to generate, process or analyse results that you intend to appear in a publication (either in a journal, conference paper, monograph, book or thesis). Research software can be anything from*

[1] `doi.org/10.5281/zenodo.1183562`*Last accessed:October 2018*
[2] `https://russellgroup.ac.uk/about/our-universities/`*Last accessed:October 2018*

a few lines of code written by yourself, to a professionally developed software package. Software that does not generate, process or analyse results - such as word processing software, or the use of a web search - does not count as "research software" for the purposes of this survey." According to the SSI survey, women were more likely to use the software that somebody else developed for their research, as women tended not to develop their own software (30%), and not to receive training (39%) as much as men (63%).

Additionally, from the initial survey analysis by SSI, it became obvious that 92% of the researchers used research software, and that 7 out of 10 researchers in the UK would not be able to conduct their research without software. Also, the survey data showed that 56% of the researchers developed their own code, but unfortunately 21% of them did not receive any training, which makes the software likely unreliable [144]. Another interesting finding of the survey was about the researchers' preference of operating system (OS). According to the data, 98% of Linux users were also research software users, and 90% of them developed their own software. By comparison, Windows users, who were also research software users (88%), tended not to develop their own software (41%) [145].

The respondents of this survey, mostly researchers in STEM subjects, who developed research software, were potential HPC users. For this reason, we decided to use the data from this survey to draw further conclusions on the gender differences in the research software community and their implications for the HPC community, as well as to gather evidence to facilitate designing our survey for the HPC community.

3.2.1 Data Selection

Because of these data being extant, we did not have more information and details on how this survey was designed. The raw data set of the responses and the data analysed by SSI[3] [145] were offered to us for further analysis. According to our research hypotheses, we correlated some of the survey responses with gender, and in some cases, we combined responses of two survey questions with gender. We used the same statistical analysis method that we used in Chapter 2, details of which can be found in the Appendix A.1).

The SSI survey asked 15 questions (11 main questions and 4 extra questions), of

[3] `doi.org/10.5281/zenodo.1183562`*Last accessed:October 2018*

which 13 were forced- or multiple-choice, and one was open-ended. It included a mixture of personal/background questions and questions related to the use of research software. The full survey questions and ordering are available in Appendix B.1.1. The survey request was sent to 15,000 researchers at 15 Russell Group Universities in the UK, during the period from August to November 2014. A total of 417 responses were received, of which 257 self-reported as male, 148 as female, and 12 as *"Other"* or chose the *"Prefer not to say"* option or did not respond.

In our analysis, we did not take into consideration the *"Other"* and *"Prefer not to say"* responses to the question asking about the respondent's gender. This is because our research is mainly focused on women and potential differences between female and male participants. To minimise error from small samples, we also excluded from the analysis the "no response" entries to other questions. Additionally, the responses of those who stated that they did not use software to the respective question (Question 6: *"Do you use research software?"*, Appendix B.1.1) of the survey were not used in the analysis, as we wished to find out more about those who used software as they were more likely to belong to the HPC community. Moreover, the number of those who did not use software was not large enough to provide meaningful insight into their behaviour within the community.

3.2.2 Analysis and Results

In this section, we present the results of the analysis of the survey responses in two groups based on two of our hypotheses:

1 Women are less likely to receive training and to develop software.

2 The majority of men and women who develop and/or use software for their research come from a STEM discipline.

These hypotheses were based on findings from Chapter 2 (section 2.2.3), that fewer women attended the ARCHER training courses, and that according to the courses they attended, the majority of women was labelled as "users" and not "developers". We also found that the majority of the courses' attendees were studying or working at a STEM discipline. We also took into consideration the results of the initial analysis of this survey by SSI (section 3.2.1) [145].

The number of responses for questions with more than three available options is often very small reducing statistical significance. To minimise the impact and facilitate the analysis, groupings were created for the responses to some questions. Discipline options (Question 3 in Appendix B.1.1) were grouped into "STEM" and "non-STEM" (full list of disciplines and classfication are available in Appendix B.1.2). Similarly, we created three groups for the years in research options (Question 5 (*"How many years have you worked in research?"*, Appendix B.1.1): "0-10 years" (including the "less than a year", "1-5 years" and "5-10 years" groups; this group is also called "early-carreer stage" throughout this study), "11-20 years" (including the "11-15 years" and "15-20 years" groups; also called "mid-career stage"), and "more than 20 years" ("late-career stage").

3.2.2.1 Gender Correlation with Software Training and Development

As reported in the initial analysis by the SSI, male respondents were far more likely to develop software (75%) than female respondents (33%). In our analysis we further correlated training, development and gender. We found that most male respondents received some kind of training and developed their own software (59%), and only 6% of male respondents received training and did not develop software. By contrast, most female respondents did not receive training and did not develop software (51%), and only 25% of female respondents were trained and developed software (Figure 3.1). We suggest that in order for women to use HPC for their research and be an active part of the community, they need to receive training and to take part in all software development aspects of the research.

Anecdotally, there is a significant correlation of software training, software development and the operating system (OS) that one uses for research. The SSI initial analysis showed that 90% of Linux users developed their own software. Our correlation of OS with gender made obvious that the majority of women (67% of a total of 121 women) and the majority of men (40% of a total of 210 men) chose Windows as their OS of preference. Yet, as Figure 3.2 illustrates, there is a significant difference ($p < 0.05$) between women and men in the distribution of preference into the three main OS categories: Windows, Linux and Mac OS X. This finding has serious implication for the HPC community, since the majority of HPC facilities are Linux based[4].

We later correlated gender with OS and training likelihood, and then gender

[4] `www.top500.org/statistics/list/` *Last accessed: October 2018*

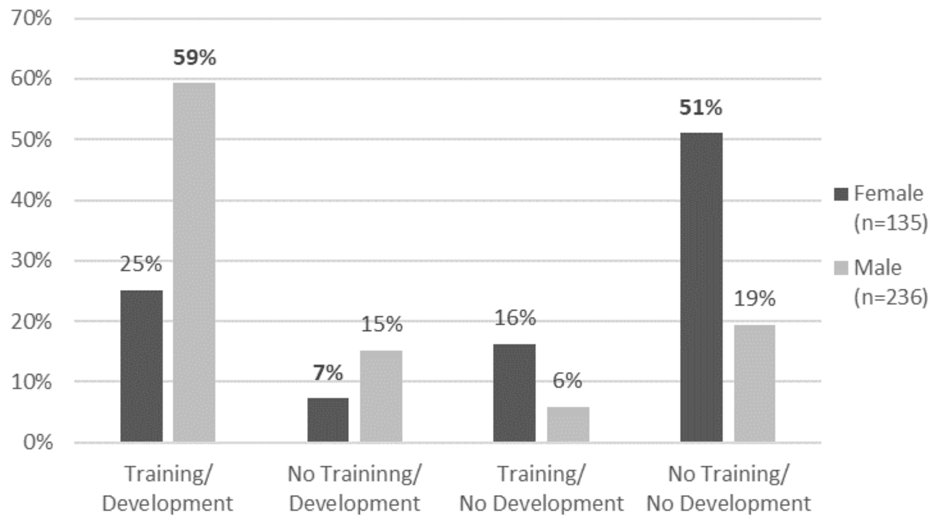


Figure 3.1 *Gender correlation with software training and software development, across all possible combinations. The percentages in bold indicate the significant differences between women and men ($p < 0.05$).*

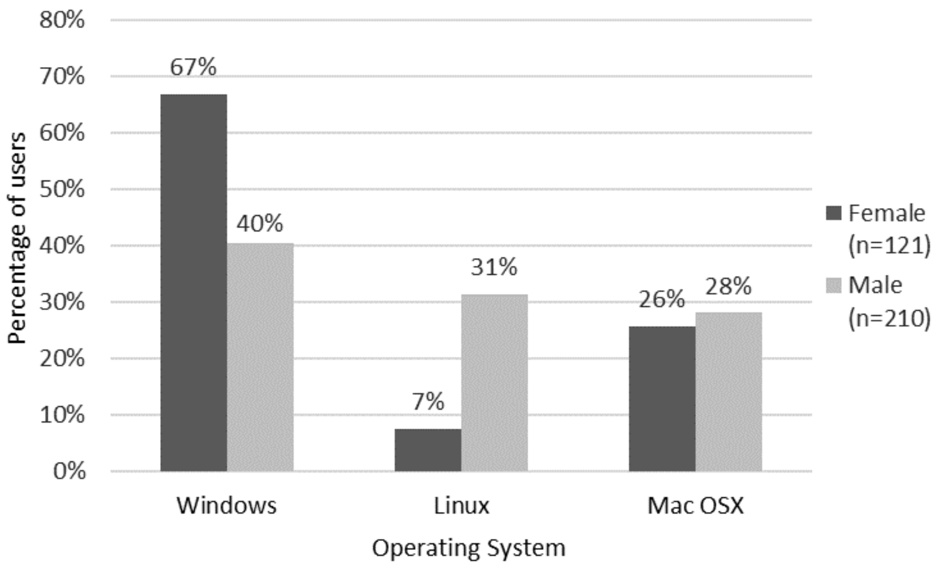


Figure 3.2 *Gender correlation with Operating System (OS) preference. Most women (67%) prefered Windows, whereas men distributed their preference almost equally across the three OS (Windows, Linux, Mac OSX.*

with OS and development likelihood. We found that male respondents received more training than female respondents, whichever operating system they used ($p < 0.05$) (Figure 3.3). Similarly, male respondents developed software more than female respondents regardless of OS preference, and the majority of the respondents (both female and male) who did not develop software were Windows users ($p < 0.05$) (Figure 3.4).

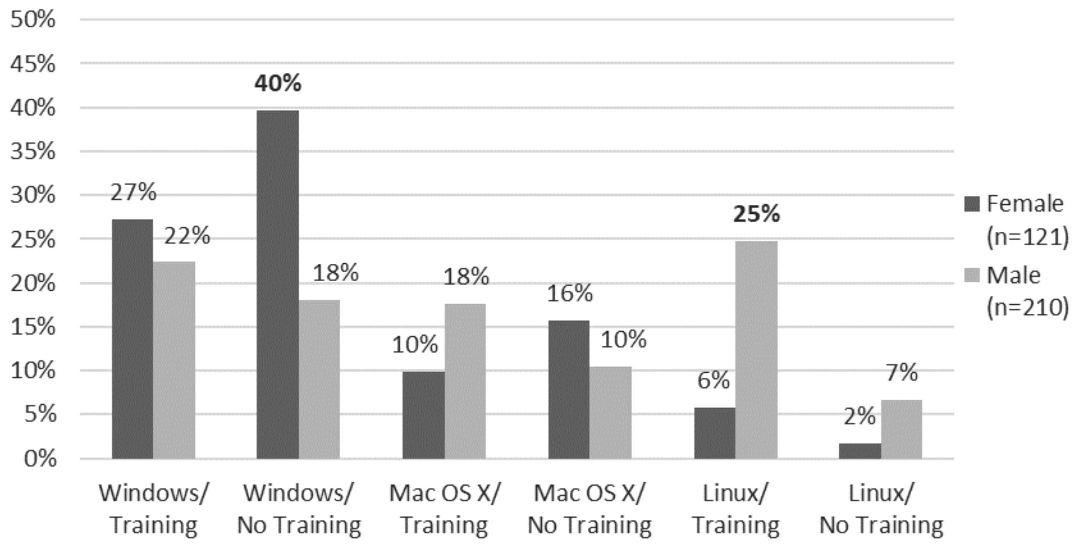


Figure 3.3 *Gender correlation between preferred Operating System (OS) and training. Most women were Windows users and did not receive training (40%), whereas most men were Linux users and received training (25%)*

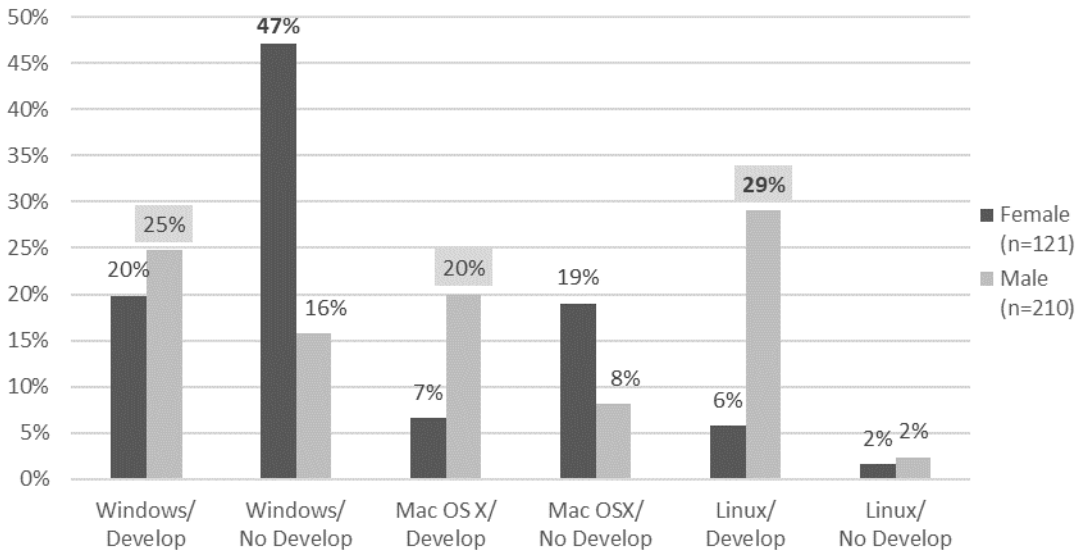


Figure 3.4 *Gender correlation between preferred Operating System (OS) and software development. Men were more likely than women to develop their own software regardless of preferred OS.*

3.2.2.2 Gender Correlation with Discipline and Working in Research

To facilitate the understanding of how discipline could impact on the use of software in research, the following analysis considered the responses to the survey by male and female respondents based on their field of study. The majority (78% of a total of 371) of respondents self-classified as STEM, suggesting that those more likely to use software and therefore to respond to a "software survey" are in STEM subjects. In correlation with gender, the majority of both women and men belonged to STEM group. However, there was a significant difference ($p < 0.05$) between men and women who belonged to non-STEM group, with more women (27%) than men (18%) coming from a non-STEM discipline. This larger number of women from non-STEM disciplines supports the hypothesis that attracting more women form non-STEM background to programming and software development might positively affect the number of women in the HPC community.

Our analysis found that women were less likely to develop software than men in both STEM and non-STEM subjects; however, a far greater proportion of women were in the STEM group and developed software (28% of a total of 135 women) than in the non-STEM group and developed software (4% of a total of 135 women). For both genders fewer received training in software development in non-STEM than in STEM group. We found that 56% of the 236 male respondents were in STEM disciplines and received training. Most of the women, however, belonged to STEM group and did not receive training (41% of a total of 135 female respondents).

We also found a significant difference ($p < 0.05$) in OS choice by gender when compared to discipline. Women in STEM (n=86) and non-STEM (n=35) primarily chose Windows, with Linux being the least preferred OS (Figure 3.5). However, for men there was a clear difference between OS choice for STEM (n= 172) and non-STEM (n=38) male respondents. In the STEM group, men showed little difference in preference for Windows (38%) or Linux (35%), but in the non-STEM group 53% of men prefered Windows, compared to 13% preferring Linux. This could be an indicator of the group of people who were more likely to be HPC users.

When asked if they would be able to conduct their research without software (Question 7: *"What would happen if you could no longer use research software?"*, Appendix B.1.1), a majority for both female (67% of a total of 135) and male

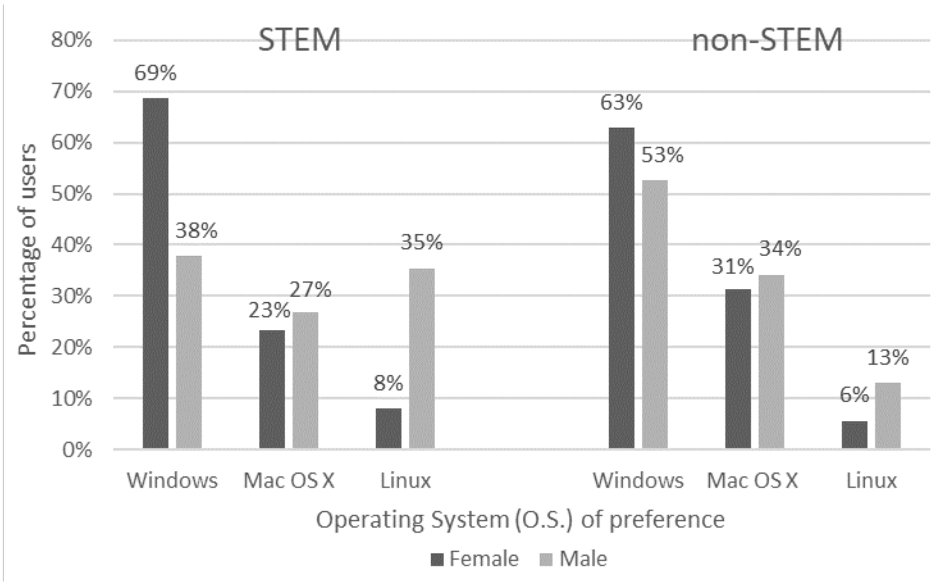


Figure 3.5 *Gender correlation with preferred OS in STEM (left) and non-STEM (right) groups.*

(78% of a total of 236) stated that it would not be practical, implying that for both genders the importance of research software is the same. However, the slightly larger percentage of women who stated that it could be possible to work without software (30% female, 19% male), suggested that either their work had less intrinsic reason to rely on software (and this may provide an explanation for why fewer women also developed their own software); or that women possibly gave a different interpretation of what research software was; or, finally, that women had a better ability to adapt in different work conditions.

From the correlation of gender with the responses to Question 5 (*"How many years have you worked in research?"*, Appendix B.1.1) we found that the majority of respondents for both genders (67% of 135 women and 58% of 236 men) belonged to the early-career stage group (0-10 years). For both genders fewer participants were classified as late-career (> 20 years), but double the proportion of male to female selected this category (19% of male respondents and 10% of female respondents). This finding is consistent with HESA findings[5] that fewer women hold senior academic positions in the UK, which we also discussed in Chapter1 (section 1.1.1).

In the early-career stage group, there was a significant difference ($p < 0.05$) in male and female Linux users (59% male and 5% female of a total of 75 Linux

[5] `https://www.hesa.ac.uk/data-and-analysis/staff` *Last accessed: October 2018*

users of all career stages) (Table 3.1). This has great implications for the future of the software community as well as the HPC community, as the early-career stage respondents are the future of this community, hence fewer Linux users will have a direct impact on those that consider HPC in their careers. We also found that men were more likely than women to have received some form of software training and to develop software at all three career stages. Women in both the early- and mid-career groups were more likely to receive training (30% and 7%, respectively. out of a total of 135 women) than to develop software (22% and 4%, respectively, out of a total of 135 women).

Table 3.1 *Distribution of Linux users (n=75, 100%) into the three career stage groups and by gender.*

Gender	Early-career stage	Mid-career stage	Late-career stage
Female	5%	4%	3%
Male	59%	17%	12%

The analysis of the correlation between the type of employment contract (Extra Question 3: *"Are you employed on a permanent or fixed-term contract?"*, Appendix B.1.1) and gender also found that the majority of the female respondents was employed on a fixed-term contract (63% of a total of 128 women), whereas for men the proportion on fixed-term contracts (48% of a total of 224 men) was almost equal numbers to permanent contracts (52%). We assume that more fixed-term contracts occur during early-career stage, and therefore this finding is in line with the greater proportion of those who were in early-career stage (67% of women and 58% of men).

3.2.3 Summary and Discussion

In this section we presented briefly the initial findings of the survey conducted by SSI. We also presented the results from our analysis of the survey responses. The main findings from our analysis are:

- Most women did not receive training and did not develop software (51%);
- Women chose Windows as their OS of preference (67%) and left Linux as their least prefered OS (7%);
- Men received training and developed software regardless OS;

- Most respondents both female (73%) and male (82%) came from STEM disciplines;
- Both women and men in the non-STEM group received less training than those in the STEM group;
- Linux is the least prefered OS of both women (6%) and men (13%) in the non-STEM group;
- The majority of both women (67%) and men (58%) belonged to the early-career stage, but more men (19%) than women (10%) belonged to the late-career stage.

In conclusion, our analysis of the SSI survey showed that female researchers are less likely to receive programming training, develop their own software, and use Linux.

The survey was exclusively targeted at UK Russell Group Universities, excluding responses from other sectors of the research software community. For this reason, we anticipate that the results of this survey apply only for that specific research group, and any generalisation of the results to the wider academic community might not be valid. We also recognise that the results of the analysis of this survey concern mainly the research software community, and we did not have any evidence that the respondents of this survey were HPC users or not. This survey was primarily used as guidance for designing our HPC community survey. Any assumptions made for the HPC community from the results of this survey are clearly implications and hypotheses, which helped us build the next steps of our study. Finally, we recognise that any qualitative conclusions drawn from the responses of this survey are based on our interpretation of the quantitative data, since this survey did not include any open-ended questions of qualitative nature.

3.3 Short Questionnaire

The SSI survey was not the only guide used for designing our survey. In 2015, we attended the annual "Hands-on introduction to HPC" course organised by the Women in High Performance (WHPC) Network. The purpose of this "special" course is to attract more female participants; we discussed about the impact of this and other similar courses on women's representation at HPC related courses in section 2.2.3.5. We decided to use a questionnaire with pilot questions for the attendees of this course as a guide for building the survey for the HPC community by identifying key responses.

3.3.1 Data Collection

The questionnaire consisted of eight questions, which can be found in Appendix B.2. The order of the questions was designed by starting with personal and factual information about the participants' background and use of HPC, and then moving into more abstract, open-ended questions about the representation of women in HPC. The ordering was chosen as to keep the interest of the respondent by starting with easier questions; to provide a better flow by grouping questions; and to avoid biased answers by asking the main research question at the start [132]. Some of the questions were based on the findings of the SSI survey (3.2).

To keep the procedure short and to be able to approach as many of the participants as possible, we used an iPad where the responses were typed either coded as one-word answer, for example *"Yes"*, *"Physics"*, or by using crucial words and phrases of the participants' answers, for example *"gender stereotypes"*, *"computer science is male dominated"*. However, some quotes were captured by using the device's recorder, when possible and with the respondent's consent.

The numbers of the course attendees (11 female and 1 male) and of the respondents to this questionnaire (8 female and 1 male) were too small to be useful to be used for quantitative analysis. Therefore, these data were only used to support the qualitative approach of this study and to facilitate designing the HPC community survey and the subsequent interview method.

3.3.2 Analysis, Results and Discussion

As mentioned earlier, 8 female and 1 male respondents from a total of 12 course attendees answered the short questionnaire. Here, we present the results in two groups: according to the repondents' background and relation to HPC, and according to their views on women's representation in the community.

3.3.2.1 Respondents' Background and Relation to HPC

Most of the respondents were PhD students in computational or science fields (4 female and 1 male); 3 women were postdoctoral researchers in computational or science fields; and 1 woman was a Research Associate. They were all interested in using HPC for research purposes, such as developing large scale simulations and models, data handling and increasing the efficiency of their research. Also, they were all keen on attending a similar training course in the future. To the question about the operating system preference (Question 2), most of them (6 female, 1 male) responded Linux and 2 of them Mac OS X. Interestingly, two of the female Linux users explained that they prefered this operating system for work purposes and Windows for personal use.

3.3.2.2 Respondents' Views on Representation of Women

The majority of the repondents (8 out of 9) stated that they were aware of the fact that there is an underrepresentation of women in HPC. The only respondent whose response to this question was negative was a woman. She added that according to her personal experience, she had not noticed a gender imbalance and this might depend on the institution. Among the reasons for the women's underrepresentation suggested by the respondents, the most common was the fact that there are fewer women in science subjects in general, and hence there are fewer women in HPC. Other reasons included: lack of role models, gender stereotypes and cultural views, the wrong idea that media and basic education create about science and computers, lack of confidence of women in technical issues and in learning new skills, and general lack of self-confidence.

The most popular suggestions to improve gender balance in STEM and HPC were related to organising communities and events especially for women in HPC, as well as addressing the problem and promoting the work of the communities through emails, newsletters and social media. Other suggested solutions included: mentoring, organising outreach activites at schools, promoting gender balance by conference organisers, educating parents and increasing role models' visibility, majorly changing people's views on science and computers, and creating effective policies that promote gender equality.

We included below some of the respondents' quotes to the last the question (Question 8), where the respondents explained why improving the numbers of women in STEM and HPC was really important.

> "We should change everyone's view on science and make it more attractive to women. Gender equality and diversity are really important, because they help opening different opportunities and make a difference at work."

> "We need to make maths and math-related fields more attractive and approachable to everyone by changing the teaching methods and the idea that maths is difficult. By changing governmental policies and strategies, we must change the position of women in family and society to achieve equality at all levels. An equal society is really important for better understanding and better opportunities and options."

> "The problem starts from education. It is very important to change this and do more outreach activities at school as soon as possible, because as long as the problem lasts, it gets harder to solve it."

As stated earlier, because of the small number of responses, the results of this questionnaire are not considered representative of the community's views. However, they played an important role for designing the next steps of our research. Additionally, the coding of the responses might have been affected by biased selection of keywords and phrases.

3.4 HPC Community Survey

As mentioned previously, the next plan was to design and conduct a survey for the HPC community using the knowledge obtained from background work on related topics, and the conclusions drawn from the SSI survey data analysis, the pilot short questionnaire, and the quantitative results of Chapter 2. The purpose of this survey was to gather evidence that added to the quantitative approach on the underrepresentation of women in the community, and to identify areas for further investigation and improvement in the HPC community.

3.4.1 Data Collection

The survey was designed and produced using the online survey development software "SurveyMonkey"[6]. There were 37 questions including the consent of participation. As explained earlier in the Ethics section 3.1, the first page of the survey stated the terms and conditions of the study, the data management, and contact details of the primary researcher and the supervisor, followed by the request of the participant's consent.

The first group of questions covered the background of the participants and their relation to the HPC community and research software. With the next group of questions, we attempted to identify gender differences on hypotheses related to conference and workplace experiences. This group was followed by questions about the respondents' views on gender related matters in the HPC community. Next, there were questions asked for further personal information, including gender, as well as caring responsibilities questions. The last four questions concerned the participation of the respondents at the two major HPC conferences, International Supercomputing Conference (ISC) and Supercomputing Conference (SC) (details of those conferences were presented in sections 2.1.1.1, 2.1.2.1), aimed to support our findings from the quantitative analysis (Chapter 2).

The order of the groups of questions was designed this way to keep the participant's interest until the end of the survey, combining easy personal and factual questions, with more difficult abstract questions. The terms and conditions of the survey, as well as the complete questions and ordering, are available in the Appendix B.3.2, B.3.3.

The survey included a combination of multiple-choice, forced-choice questions, rating scales and open-ended questions. We included the *"other (please specify)"* option as a response to some of the questions giving the opportunity to the respondents to expand on their answer. The "other" option is also an effective way to avoid forcing the respondents to choose one of the given answers, thus producing false assumptions. It also helps the survey appear more polite and friendly to the participants, and adds more chances of them completing the full survey [132]. In the same way, none but the consent question was mandatory, and respondents were able to skip questions. We counted but we did not take into consideration the "no response" answers during the analysis of each question.

[6] `https://www.surveymonkey.net/` *Last accessed: October 2018*

Even though the purpose of this study was to identify potential differences between female and male in the HPC community, we acknowledged the cases of respondents self-identifying themselves as non-binary. We also respected the right of individuals who wanted to participate in the survey but did not want to share their gender. For this reason, there were four answer options provided to the question about the participant's gender: *"female"*, *"male"*, *"other"*, and *"prefer not to specify"*. However, the responses of those who self-reported as *"other"* and *"prefer not to specify"* were not included in further analysis. Beyond that, it is important to mention that we included racial and ethnic minorities in the questions about discrimination at workplace to avoid arguments and misinterpretation of the survey discussion recognising only gender discrimination. The responses related to racial and ethnic minorities were not included in the analysis of this study; however, they formed a good sample for further work.

The first release of the survey invitation occured through the WHPC website[7] in January 2016, when the survey opened to public. During the second half of February 2016, a first round of invitation emails was sent to individuals, centres, institutions, companies, universities in the UK, media and mailing lists related to HPC. The invitation procedure was monitored by creating a list of the invited contacts with marked dates. This way we tried to keep track of the people we invited to avoid disturbing them by contacting them multiple times during the same period. However, it was hard to monitor the number of times someone was invited to participate, as some of the people with whom we got in touch were willing to circulate the invitation to their contacts, to send it to users' lists or to advertise it on their websites. A second and final round of invitations was sent during the last week of March 2016 to all the contacts we had previously reached, as well as to additional contacts that we found or that we were recommended by others. The survey was open until the first week of May 2016, and the raw data set was exported from the survey website later that month.

According to the data received from the online software we used, the survey was reached by 854 people; however, only 538 were *"complete"*, according to the definition provided by "SurveyMonkey"[8] (a response is considered complete when *"the respondent answered all required questions they saw and clicked Done on the last page of the survey"*). Of the 538 responses examined and analysed,257 self-reported as women and 266 self-reported as men as a response to Question

[7] `https://womeninhpc.org/` *Last accessed: October 2018*

[8] `https://help.surveymonkey.com/articles/en_US/kb/What-is-considered-a-survey-response` *Last accessed: October 2018*

32).

3.4.2 Analysis and Results

For the analysis of the HPC community survey data we used the same analysis procedure as for the SSI survey, where the responses to each question and combinations of responses to two questions were correlated with gender, following our hypotheses and previous results. To facilitate the analysis and the presentation of the HPC community survey results, we created groups of responses of related and corelated questions with gender. The order of the presented results was chosen to connect findings of this survey with the findings of the previous methods of our study so far.

The quantitative analysis was done by calculating the numbers or percentages of women's and men's responses to the survey questions. We also used the same statistical analysis method as in Chapter 2 and for SSI survey data analysis (Appendix A.1). The qualitative analysis was done by identifying the most popular responses, which might have supported or rejected our hypotheses. In order to analyse the open-ended questions qualitatively, I used the method of *coding* [132]. Coding is an analytical process in which data are categorized to facilitate analysis. I followed the manual process (no use of software), by highlighting different concepts by assorted colours, and then organising them into broader theme categories. Then, I tried to reduce the themes to a manageable number, and created a hierarchy of the most popular responses within each category.

Consequently, in this section there are 5 groups of results:

1 Background and career progression in the HPC community

2 Training and Software Development

3 Family Responsibilities and Travelling for Work Purposes

4 Gender Balance and Discrimination in the HPC Community

5 The Impact of Mentoring and Stereotype Threat

3.4.2.1 Background and career progression in the HPC community

First of all, we needed to define the HPC community. In this section, we presented the responses to those questions which informed us on the respondents' background, job description and current role, sector, qualifications and age group. We also compared the perception of success and career progression between women and men in the community.

Question 2 asked the participants to select all those disciplines that best applied to their background and/or current working environment. To facilitate the analysis, we created a STEM group and a non-STEM group (full lists of disciplines per group in Appendix B.3.3). There was also an *"Other"* option, which was picked by 18 women and 10 men. These were not taken into consideration for further analysis.

The vast majority of both women and men belonged to the STEM group forming together 91% of the responses (total responses of women to this question n=348, and of men n=394); 88% of women picked a STEM discipline (30% chose *"Computer Science"* and 29% chose *"Physical and Mathematical Sciences"*) and 12% picked a non-STEM discipline, which is double the percentage of men who belonged to the same group (6% of the men non-STEM, $p < 0.05$). Consequently, 94% of men picked a STEM discipline (37% *"Physical and Mathematical Sciences"* and 35% *"Computer Science"*).

Question 4 asked the participants to choose all those options that best described their work in interaction with the HPC community. There were 11 options including *"Other"* (full list of options in Appendix B.3.3). To facilitate the analysis, we grouped some of the options to reduce the impact of small number of reponses. The final groups were: *"HPC applications/tools development"*, *"Users of HPC facilities and HPC software/applications"*, *"HPC systems administration and validation"*, *"Sales"*, *"Management"*, and *"HPC training"*.

There were small but significant ($p < 0.05$), differences in the HPC work description between women and men. Most women and men belonged to the *"Users"* group (51% of the 539 responses by women, and 43% of the 597 responses by men). The second choice was *"HPC applications/tools development"* for both women and men (18% for women and 25% for men). *"HPC training"* group was the third most popular choice for women with 10%, and the fourth choice for men (10%) after *"HPC systems administration and validation"* with 11% (8% for women). It is worth mentioning that most of the 6% women who picked *"Other"*, specified their work as *"Marketing"* and *"Communications"*; whereas most of the only 2% of men who chose *"Other"*, specified it as *"HPC Hardware development"*.

We correlated the responses to those two questions (2 and 4) with gender, to detect any potential differences in the way people from STEM backgrounds work within the HPC community in comparison to those from non-STEM background. We were particularly interested in the two most popular groups of work that we identified from the analysis of Question 4, *"Users"* and *"HPC applications/tools development"* groups, as well as the almost as popular *"HPC training"* group. The results showed significant differences between the STEM and non-STEM groups, and between the two genders in each group ($p < 0.05$). Figure 3.6 summarises these results, with the majority of both women and men of the STEM group describing themselves as *"Users"* (63% women and 54% men) and more than the non-STEM group (45% women and 46% men), where both women and men appeared to interact with HPC more as trainers (36% women and 29% men).

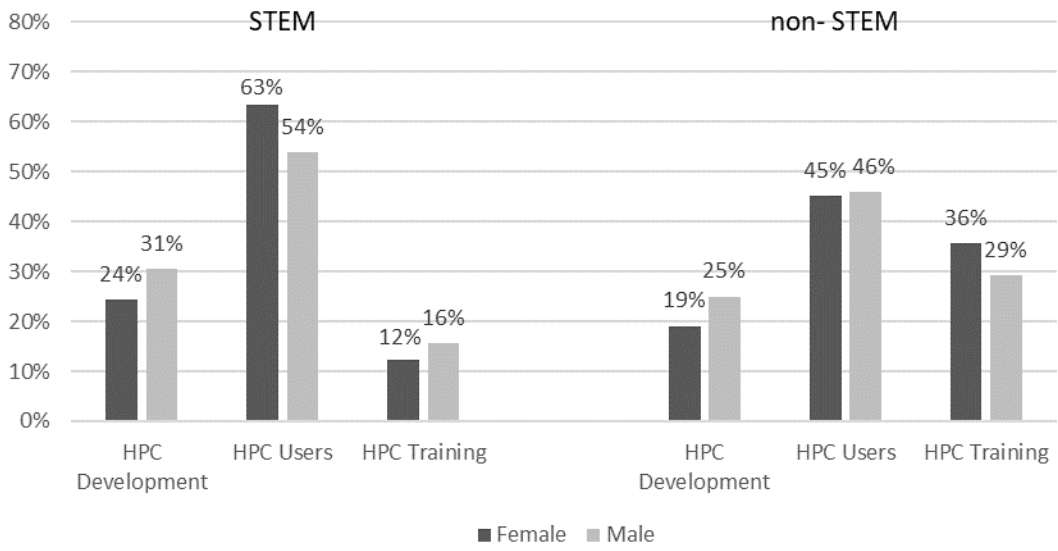


Figure 3.6 *Percentages of women and men from STEM and non-STEM background in the three main work interactions with HPC.*

The results from these two questions supported our hypothesis that the majority of people in the HPC community - especially users and developers - come from STEM background, and the people from non-STEM background are more likely related to HPC in different capacities.

When we asked the participants to rate how successful they thought they were

(Question 11), more than half of both women and men did not respond (58% of 257 women and 63% of 266 men). There was no significant difference in personal success rating between the women and men who responded (109 women and 99 men, $p > 0.05$), with women's average rating at 3.23 and men's at 3.35 on a 1-5 scale, where 1 stands for *"Not at all successful"*, and 5 stands for *"Extremely successful"*.

The question offered the opportunity to the respondents to explain their answers. Of the 124 women, who gave an explanation, almost equally divided proportions of them thought of themselves as being *"successful"*, because of: having been promoted, being in a senior position, being recognised and having many publications; and as being *"reasonably successful"* or *"average"*, because of having achieved some of their goals, still wanting to do more, or not having been given more responsibilities. A few women stated that they were at the start of their career or were still students and they were not able to rate their success yet. Very few were not happy with their work and considered themselves *"unsuccessful"*, because of having been at the same position for many years. Some of their comments included:

> "I'm reasonably successful in my field, but I'm not managing a project or setting technical direction for a big software effort."

> "I'm relatively well-known and have a strong network both in HPC industry and at the national labs."

> "Not as successful as I should have been, mainly due to a lack of mentoring when I needed it most! Also, I am not currently happy to relocate to advance my career due to my children being at a pivotal point in their school education."

Of the 109 men, who gave an explanation, most of them considered themselves *"successful"*, because of being in a senior position, being promoted, getting a good salary, working in a good place, being recognised; or they considered themselves *"relatively successful"*, because of doing well but wanting to do more, being good enough but not exceptional. Very few men appeared to consider themselves as *"not successful"*, having had little or slow progress, not being happy with their job, being still students or at early career stage, which left them unable to rate their success. Some of men's responses were:

> "I have a permanent research position with stable funding, fantastic co-workers, and I am paid very well."

> "I think I have achieved some good things, but nothing truly outstanding."

> "I think I should be paid more by now. I would like to move into a more senior role."

There were no significant differences in the responses of women and men to the next question, which asked them where they wanted to be in 5 years (Question12, $p > 0.05$).The majority of both women and men stated that they wanted to be *"in a more senior role"* (58% of the 254 women and 61% of the 266 men who responded to the question), followed by 17% of women and 19% of men who wanted to stay in their current role. Most of both the 32 women and the 20 men who picked *"Other"* as their response, wanted to see themselves retired in 5 years.

No significant difference was noticed on the sector that the participants worked in, with the majority of both women and men working in academia (70% women and 72% men, $p > 0.05$). This finding agreed with the results of the analysis of training courses feedback reponses (section 2.2.3.4). Finally, almost all of the 27 women and the 28 men who chose *"Other"*, specified it as *"Government"* or *"National Laboratory"*.

No significant differences were found on the highest qualification of the participants (Question 28), with most women and men holding a PhD (48% of the 252 female and 56% of the 263 male respondents to the question). However, there were significant differences on the current role of the female and male respondents ($p < 0.05$). As Figure 3.7 illustrates, there were considerably more men than women having picked *"Research (academia, tenured)"* (14% of the 263 men and 8% of the 253 women); three times more men than women picked *"Systems administrator"* (9% men and 3% women); more women than men picked *"Marketing"* (6% women and 1% men); and women picked *"Other"* twice as much as men (19% women and 8% men), where most of both women and men qualified it as *"non-academia research"* and *"management"*.

Finally, the majority of both women and men self-reported to belong in the *"30-39"* age group (34% of the 257 women and 39% of the 265 men who responded to

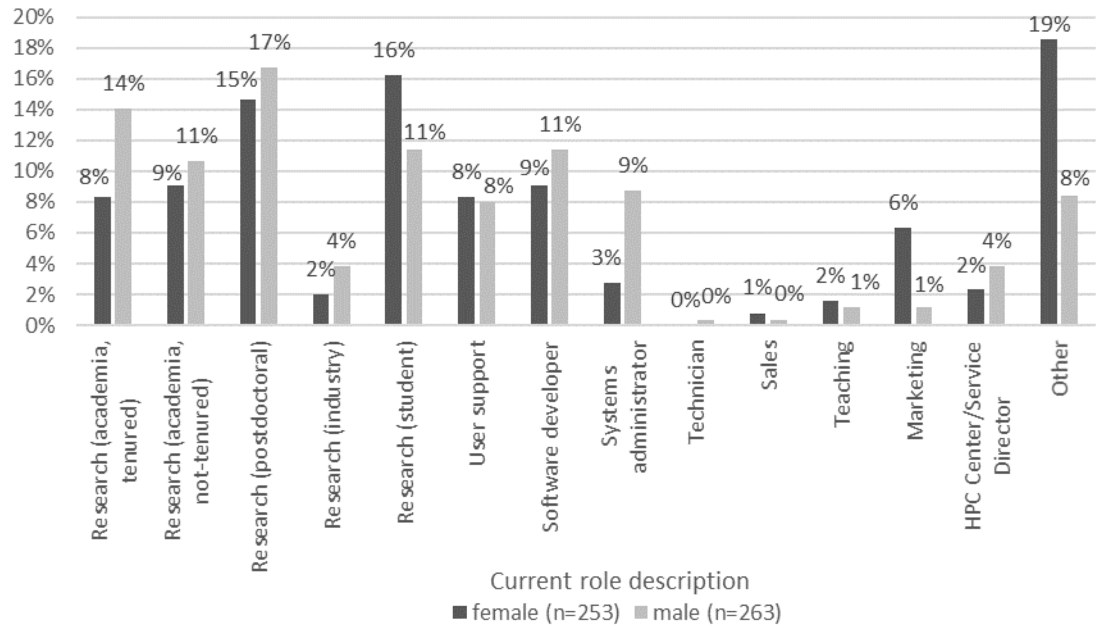


Figure 3.7 *Percentages of women and men who picked the best description for their current role.*

the question). However, more women than men belonged in the *"20-29"* group (28% women and 18% men) and *"50-59"* group (14% women and 8% men), and more men than women belonged in the *"40-49"* group (27% men and 21% women) ($p < 0.05$).

3.4.2.2 Training and Software Development

We discussed earlier the results of the SSI survey data analysis (3.2.2), which showed that women were more likely than men to choose Windows as their prefered Operating System (OS), as well as women being far less likely to receive training and develop their own software. Since these results refered to people that used research software in general, and since OS is a very important factor for developing software and using HPC, we included the same questions in our survey to examine trends in the HPC community.

From the analysis of the responses to Question 5 (*"How familiar/comfortable are you with the following operating systems? 1 is not at all familiar and 5 is very familiar"*), we found that in the HPC community there was no significant difference with OS familliarity between women and men (256 women responded to the question and 6 chose *"Other"*, 266 men responded to the question and 22 chose *"Other"*, $p > 0.05$). However, men appeared to present a slightly higher familiriaty with Linux (4.4) than the other OS (Windows 3.5 and Mac OS X 3.3) and than women (Linux 4, Windows 3.8, Mac OS X 3.2)

When we correlated the OS familiarity with discipline (STEM and non-STEM groups from responses to Question 2, 3.4.2.1), we noticed that women from non-STEM background were by far less familiar with Linux (2.8) than with Windows (4.4), and than women from STEM disciplines (4.2). Whereas, men from non-STEM disciplines seemed to be equally familiar with all three OS ($p > 0.05$, Figure 3.8.

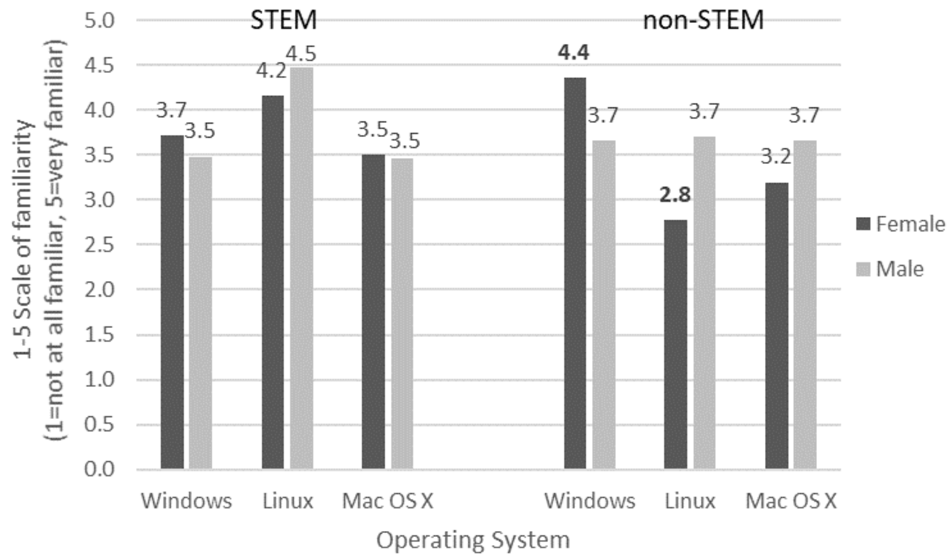


Figure 3.8 *Female and male participant's familiriaty with Windows, Linux and Mac OS X from STEM and non-STEM disciplines, where 1=not at all familiar and 5=very familiar.*

The analysis of the responses to Questions 4 and 6 revealed significant differences between the two genders on software development and training ($p < 0.05$, in both cases). About half of the 254 women, who responded to Question 4, stated that they did not develop their own software (51%), whereas for men this percentage fell to only 25% (of the 264 male respondents to this question). A similar picture was shaped from the reponses to Question 6, with more men (64% of the 264 male respondents to the question) than women having received software training (51% of the 255 femal respondents to the question). There was no remarkable difference on the type of training the respondents received, with the majority of both women and men having attended a course in person (94% of the 131 women and 92% of the 170 men who stated that they had received training, $p > 0.05$).

From the correlation of the Questions 4 and 6 with gender, we found significant differences between women and men ($p < 0.05$, Figure 3.9). The majority of men (54% of 263) received training and developed their own software; whereas the majority of women splitted into two categories: those who received training and developed their own software (31% of 253), and those who did not receive any training and did not develop software (30% of 253). Additionally, there was a remarkable difference between women and men who received training and did not develop software, with women being twice as likely to fall into this category (20% of 253 women and 10% of 263 men).

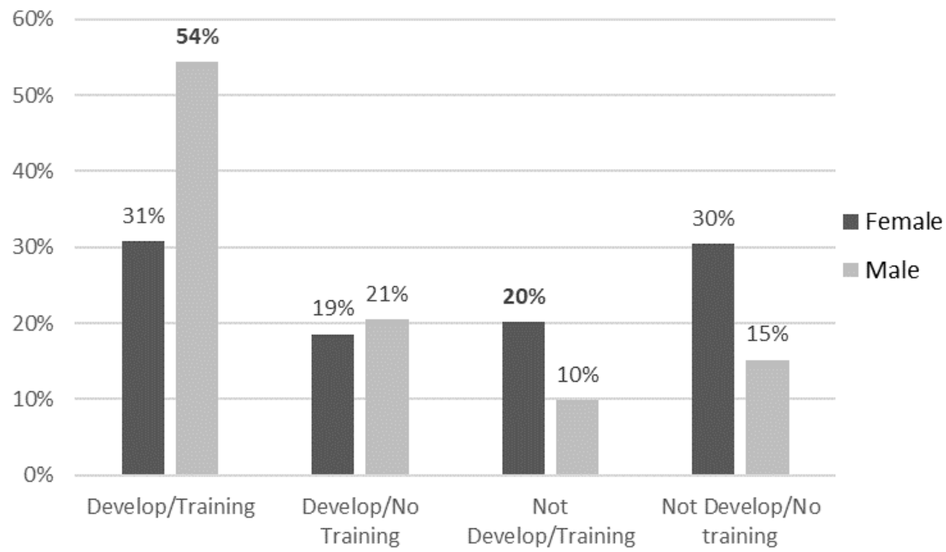


Figure 3.9 *Percentages of women and men in correlation with software training and development.*

We also noticed remarkable differences when we correlated the background of the respondents (STEM and non-STEM groups from Question 2) with Question 4 and 6 respectively ($p < 0.05$). Both women and men from non-STEM disciplines were considerably less likely to receive training, and to develop software than those from STEM background, as Figure 3.10 illustrates. These findings, in addition to the finding about the OS familiarity, suggest that respondents from non-STEM disciplines are more likely to be involved in the HPC community in different capacities than users or developers, as we discussed earier (3.4.2.1).

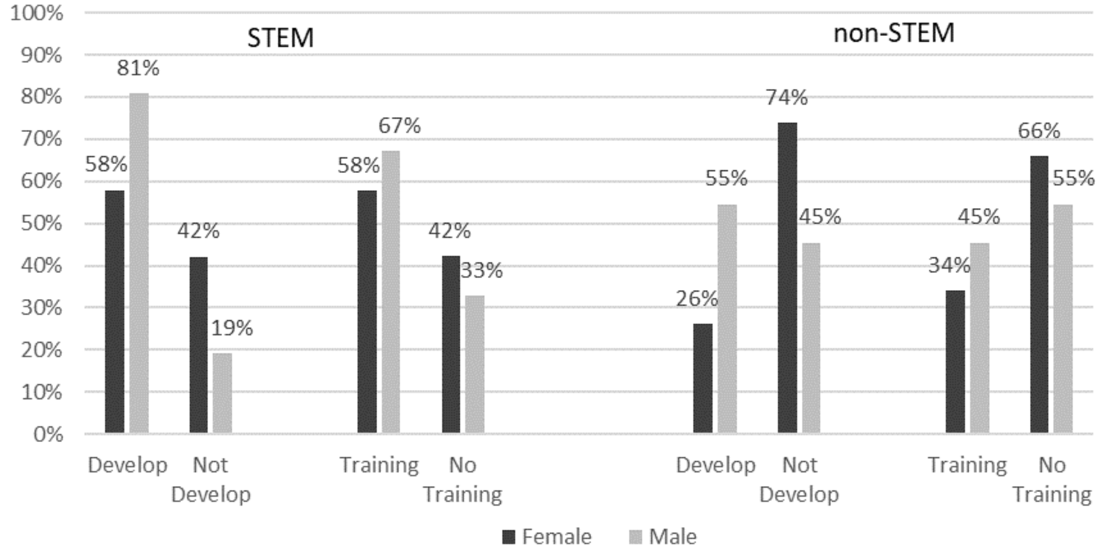


Figure 3.10 *Percentages of women and men from STEM and non-STEM backgrounds in correlation with software training and development.*

3.4.2.3 Family Responsibilities and Travelling for Work Purposes

Our aim in this section was to identify potential gender differences in travelling patterns and care responsibilities, in order to accept or reject our hypotheses that women travel less than men for work and that women are less likely to have family if they want to progress in a STEM career [23–26]. For this purpose, we analysed the responses to questions related to family and travelling for work.

In Question 29, we asked the participants to state if they had children or not. There was not a significant difference between women and men with children, with fewer women than men responding positively (40% of the 248 women who responded to that question, and 46% of the 260 male respondents, $p > 0.05$). However, there was a significant difference between women and men on the impact that they believed having children had or would have had on their career (Questions 30 and 31). The majority of both women and men with children - but more women than men - believed that having children had impacted their careers (75% of the 97 women with children who reponded to that question and 62% of the 114 men with children who responded to that question, $p < 0.05$). Similarly, from the repondents without children, the majority of both women and men believed that having children would have impacted their career (85% of the 146 childless women who responded to that question and 75% of 127 childless men who responded to that question, $p < 0.05$). Figure 3.11 summarises these findings.

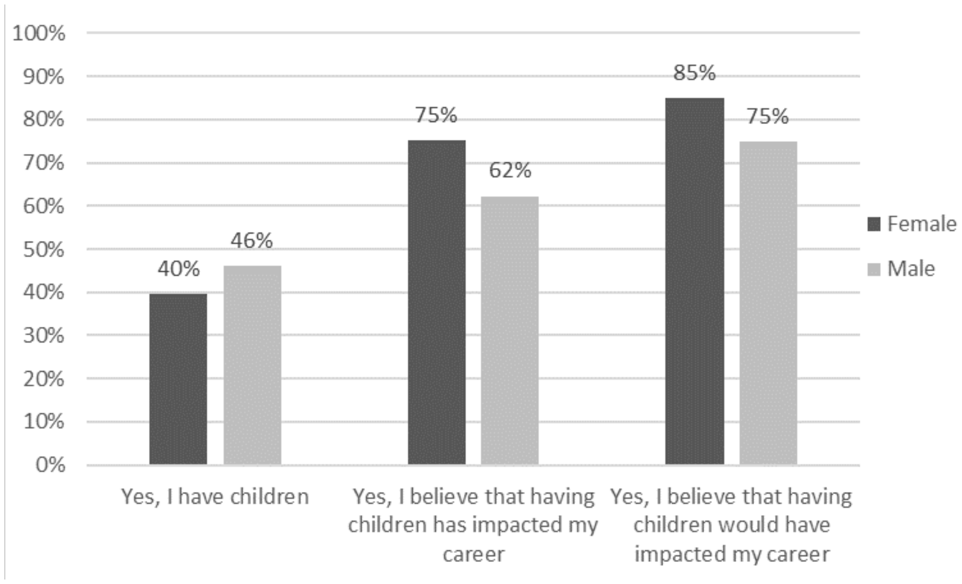


Figure 3.11 *Percentages of: female and male respondents with children (left); female and male respondents with children who believed that having children had impacted their career (middle); and female and male respondents without children who believed that having children would have impacted their career (right).*

Unfortunately, we did not provide an open-text option for explanation of responses to this question to reduce the number of long responses for analysis. However, by omitting this option, we did not know if having children impacted or would have impacted negatively or positively the career of the respondents. Event though, our question primarily implied a negative impact according to our hypotheses, the lack of explanation by the participants threatened the validity of the interpretation of the responses to those questions (Questions 30 and 31), as it is explained in Section 3.4.3.

There was no significant difference between women and men on their annual frequency of travelling to conferences (Question 7, $p > 0.05$). The majority of both women and men splitted into the two middle options: *"Once per year"* (35% of the 255 female and 32% of the 266 male respondents to that question) and *"2-3 times per year"* (39% of both female and male respondents).

However, when we correlated the responses of the female and male participants with children and those without children with their annual frequency of travelling to conferences (Questions 7 and 29), we noticed that there was a remarkable difference between women and men with children, and men with children and men without children, on their travel frequency ($p > 0.05$). As Figure 3.12 shows, the majority of men with children (41% of 120 men) was concentrated into the *"2-3 times per year"* group, with the percentages for the next two groups being very close (26% into the *"Once per year"* and 23% into the *"More than 3 times per year"*). Most women with children stated that they travelled *"2-3 times per year"* and *"Once per year"* (38% and 37% respectively of the 98 women wth children), and only 13% of them stated that they travelled *"More than 3 times per year"*; which was a very similar picture with the one for women without children (33% *"Once per year"*, 39% *"2-3 times per year"*, 14% *"More than 3 times per year"*). Also, interestingly, more men with children (23%) than men without children (13%) appeared to travel *"More than 3 times per year"*, and *"2-3 times per year"* (41% men with children and 38% men without children). Finally, it is interesing that women's travelling pattern is very similar either they have children or not (mainly between *"Once per year"* and *"2-3 times per year"*). Those last two findings do not seem to have a clear explanation, and should trigger further investigation.

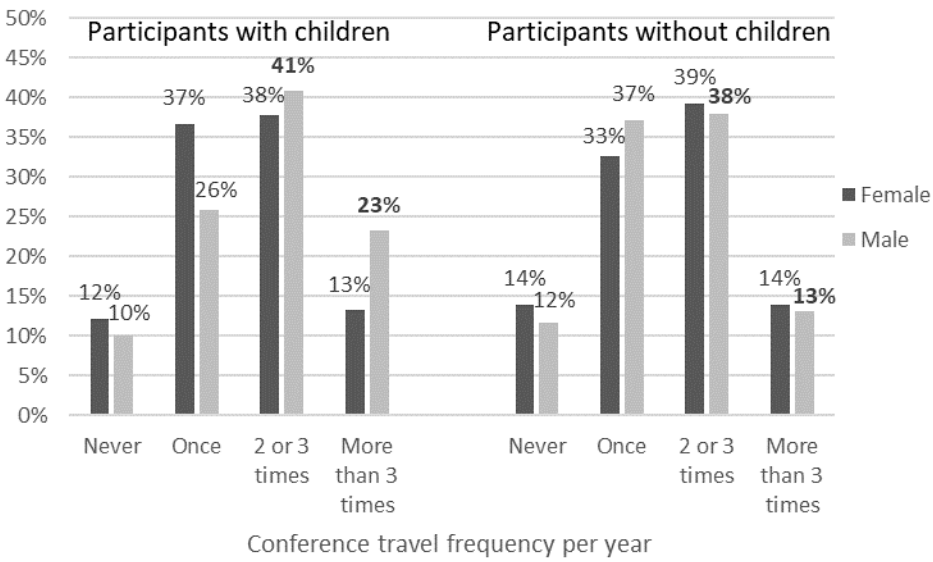


Figure 3.12 *Percentages of female and male participants with and without children in correlation with their annual frequency of travelling to conferences.*

We also asked the participants if they would consider to move their family to a different part of the country or a different country for work purposes (Questions 13 and 14). In both cases, more men than women - however the majority for both women and men - responded positively (63% of the 254 female and 71% of the 264 male respondents to Question 13, 59% of the 254 female and 64% of the 264 male respondents to Question 14, $p > 0.05$ in both cases). Figure 3.13 summarises these results, which might show that family and children do not affect men's travelling pattern for work purposes as much as women's.

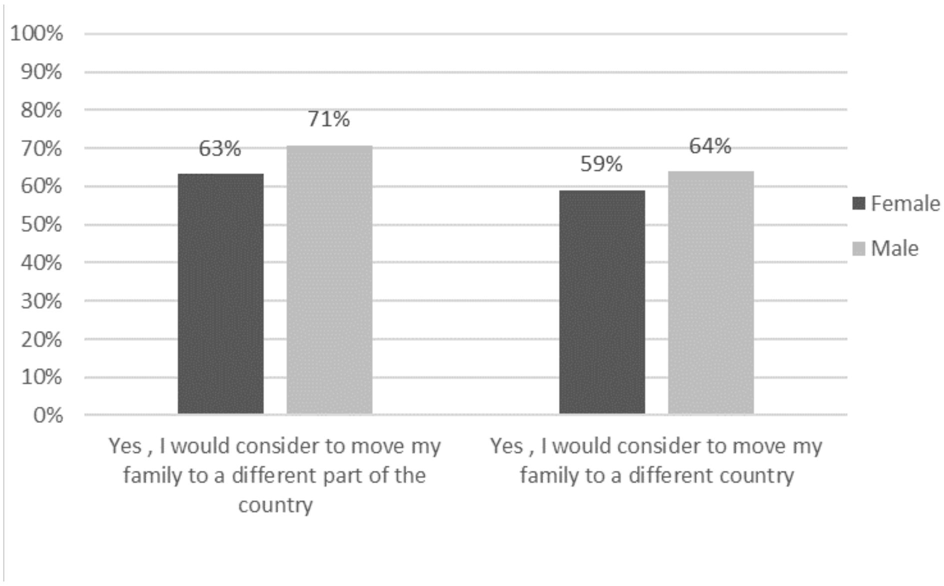


Figure 3.13 *Percentages of female and male participants who would consider to move their family to a different part of the country or to a different country.*

Finally, Questions 34, 35, 36 and 37 asked the participants if they had attended the two major HPC conferences - ISC and SC - the previous year (2015), and other related questions. The purpose of those questions was to gather information on the conference attendance frequency and type of participation of women and men of the HPC community. However, the idea for those questions was conceived before we gathered and analysed the data of the various HPC related conferences, which we discussed in Chapter 2. Additionally, the majority of the participants (81% women and 76% men, Question 34) stated that they had not attended either of the two conferences, and they did not respond to the final three questions related to those conferences. Hence, we did not include here the results from the analysis of those questions, since they did not provide us with insightful information.

3.4.2.4 Gender Balance and Discrimination in the HPC Community

In this section, we analysed and compared the responses of women and men to questions related to gender balance, discrimination at workplace and conference environment, and minorities in the HPC community (Questions 15-23). As mentioned earlier (section 3.4.1), we did not take into consideration responses related to race and ethnic minorities.

Question 15 asked the participants if they agreed or disagreed with statements related to gender-based discrimination in the workplace within the HPC community. We used a 1 to 5 scale, with 1 standing for *"Strongly Disagree"* and 5 for *"Strongly Agree"*, and we calculated the weighted average of the responses. Figure 3.14 shows that in all cases, men (n=263) agreed more than women (n=255) with the statements, which means that men considered more that there was no discrimination and all colleagues were treated equally regardless of gender ($p < 0.05$).

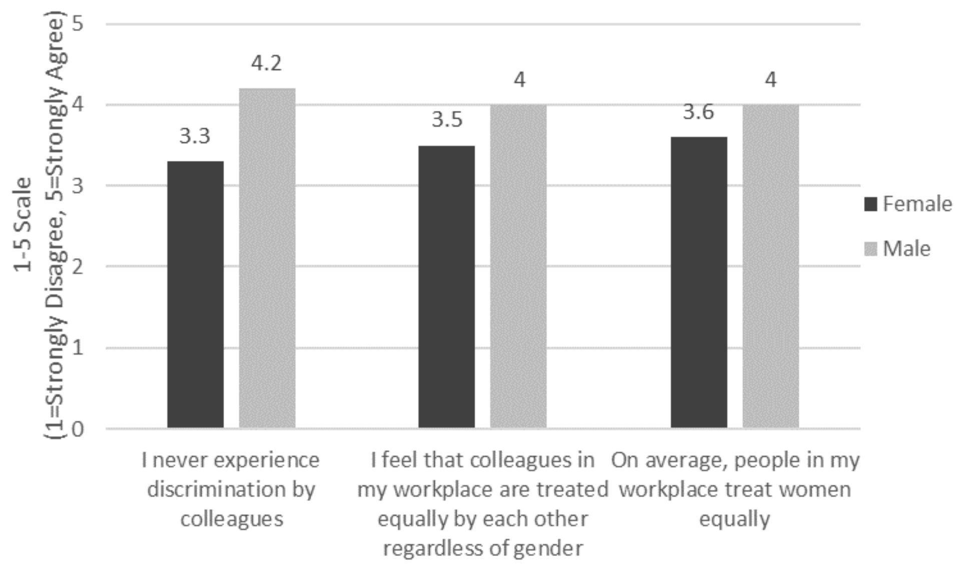


Figure 3.14 *Ratings of women and men on workplace discrimination and gender equality on a scale from 1 to 5 (1=Strongly Disagree to 5=Strongly Agree).*

For those who disagreed with any of the aformentioned statements, Question 16 gave them the opportunity to explain why they disagreed in open-text responses. About 10% of the 89 women who responded to that question expressed their opinion on ethnic minorities and other forms of discrimination. Of the responses related to gender discrimination, the majority focused on salary discrimination and the fact that women needed to work more to prove themselves and be recognized, for less money and fewer opportunities to be promoted. They also emphasised the administrative and social responsibilities that usually were given to women and the assumption that women were *"secretaries"* and that men were always in leader positions. The difficulties to be heard in meetings, and those that pregnant women and mothers could face were also mentioned often. Additionally, the fact that women could not socialise with their colleagues, when the majority was male, and how this could affect their work-life was also mentioned a few times. *Unconscious Bias* or just *"biases"* were given as an explanation for discrimination by a smaller number of female respondents compared to male. Some of the women's responses to that question that led to these findings were:

> "Women are given less work and job duties and less pay I guess."

> "Women always have to do the *small tasks* in our organisation, like making coffee, taking minutes, calling taxis, even if they have a role in a meeting or are not hired to do these tasks. Our management is predominantly male."

> "There are many examples of gender inequality also in my department. Women's statements are not taken as much in consideration as men's one. A woman is often excluded by jokes made between male colleagues. Pregnancy is a huge obstacle for women's carrier."

The majority of the male respondents (45 men responded to this question) appeared to be aware of some kind of discrimination towards female colleagues, but they tended not to give more details on how or why women were discriminated. Interestingly, the words *"differently"* and *"specially"* were used for describing how women were treated, without being more specific on what they meant with those words. Some men suggested *Unconscious Bias* (discussed in section 1.1.2.3) as the reason of discrimination, while other recognised the lack of women and diversity in their groups and how that possibly revealed an issue.

Gender pay gap was also mentioned, as well as the fact that women were given more administrative than technical roles. There was a small number of male respondents who believed that women have *"advantages"* or *"privileges"* or that women were treated more nicely because of their gender. There were a few male respondents who supported the idea that older or senior male members were the ones who discriminated against women. About a quarter of the male respondents to this question shared their opinion on other discrimination matters, such as: ethnic, age, career stage. Some of their responses included:

"I often see examples of implicit (and sadly sometimes explicit) discrimination against women - behaviour which simply wouldn't be practised, far less accepted, with men."

"A senior male obviously treats women differently than men."

"There appears to be some unconscious bias in certain men when listening to women speak."

Question 17 and 18 followed the same pattern with Questions 15 and 16, but in relation to conferences environment. Similarly, men (n=251) agreed more than women (n=248) with all the statements ($p > 0.05$), as it is shown in Figure 3.15. This means that men believed more than women that there is no discrimination at conferences and that all attendees were treated equally regardless of gender.

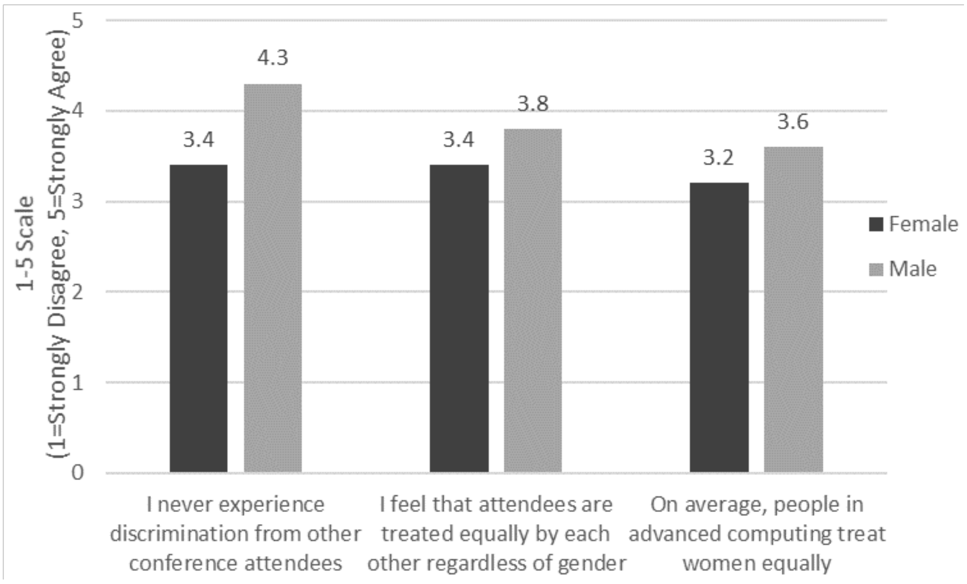


Figure 3.15 *Ratings of women and men on discrimination and gender equality in HPC conference environment on a scale from 1 to 5 (1=Strongly Disagree to 5=Strongly Agree).*

Those who disagreed with any of the statements of Question 17, could give their explanation in an open-text response to Question 18. Of the 72 women who responded, more than half of them explained that men at conferences made very often assumptions for women's position or skills based on their gender, and that women were usually ignored and not taken seriously in a male-dominated environment. Some women mentioned that they did not go to conferences at all for all these reasons. The most popular responses of the second half included: social awkwardness and difficulties of mingling in an environment where men were the majority, as well as sexism with rude behaviour and jokes. Responses to that question included:

> "It does seem women are less heard at conferences - not well represented in talks, often looked over in question time, and certainly social interactions are not always appropriate. Again, I have not noticed problems with race and ethnicity to the same degree."

> "I think people tend to assume that women are at most consumers of HPC, not leaders in the field."

> "I am often the only woman in a group and find that my presence is often ignored. I have also been in vendor meetings where I have felt uncomfortable with sexist comments made."

Similarly, most of the 44 male respondents that explained their views on discrimination at conferences talked about the gender-based assumptions for the female attendees' position and qualifications at the highly male-dominated conferences of the community, as well as the implicit biases. Very few men mentioned that they did not attend conferences. The next most popular responses - almost equally mentioned - sexism, favouritism over women because of their small number at conferences, and the fact that maybe women were not interested in the field and that was not something that needs to be changed. Some of the responses were:

> "I think that HPC still has a very real bias towards assuming women either aren't interested, or aren't capable. This needs training, education and advocacy."

> "I have seen women's research dismissed after their talks. I have seen men make comments about women's bodies and ignore their research. And I have heard many stories of sexual harassment occurring at conferences."

> "I guess at times women are favoured in conferences! Also, not many women show up in highly scientific studies. They are favoured in general. I have observed it!"

Most women and men (81% of 257 female and 71% of 266 male respondents to Question 19) agreed that *"Women are under-represented in HPC and [I think] there should be more women"*. Of those who stated that they did not agree with any of the statements concerning women's underrepresentation in the HPC community (9% women and 12% men), most of them explained why. Most women explained that they did not think the fact that women were underrepresented in HPC was something needed or forced to be changed. The next two most popular answers to this question were: *"I dont know/ I have no idea"*, and that probably women did not like this field. They explained:

> "I don't think it's something that should be forced. People should do what they are interested in regardless of gender..."

> "Women should have the same opportunities, but I think, naturally, this area appeals more to men and opposite shouldn't be forced."

The majority of men specified that they did not know or did not care if women were underrepresented in the HPC community. Then, some men explained that there was no need to force a change on the underrepresentation of women; and few of them mentioned that the reason for women's underrepresentaiton was that women were not interested in this area:

> "Opportunities are equal for men and women, in fact women are favoured, but I cannot guarantee that women are ready to compete in this men-dominated field. Just go and see in bachelor's and master's level classrooms of various universities. I might have taught a total of 250-300 students and the number of girl students was less than five in total!"

> "Women are under-represented; a balance would be desirable however not artificially enforced."

> “I honestly have no idea about the ratio of men to women in HPC, it’s such an international community that I do not know the sex of authors based on their names. In my limited personal experience, roughly half of my co-workers have been female.”

Question 20 was an open-ended question asking how the women’s proportion in the HPC community could be changed. About 2/3 of 172 women, who responded to this question, believed that the best ways to increase the proportion of women in the HPC community were: to focus on the education of the girls since the early school years; to encourage female undergraduate students and women who are already in the field to stay and progress; to provide mentoring; to promote female role models; and to change the image of HPC and the general idea that computing is a male or geeky thing. The rest of the responses included the view that we firstly need to change the proportion of women in the STEM disciplines, as this was regarded as the main reason for the underrepresentation of women in the HPC community; to change policies and to provide more flexible hours to support a work-life balance for women with children; to provide unconscious bias and equality and diversity training to all staff members - especially those who are involved to recruitment and promoting procedures. Their responses included:

> “Provide better computational education for girls so they know it is an option, promote a better culture so that women don’t get pushed out once they’re here, promote better policies around child care so that women aren’t disproportionately forced out due to logistical reasons.”

> “Tackle the under-representation of women in science generally.”

> “More female gender friendly conferences. Mentoring can also help.”

> “Introducing computing to girls at a younger age. Changing the perception that everyone who works within computing is a *geek*.”

A similar picture was obtained from the 158 responses by men. Most popular responses included encouragement and computing education of girls at early ages, active promotion of female role models, outreach, change of HPC image, and change of STEM male-dominated environment which affects HPC. Equality training, change of policies, and more family friendly workplace environment were the next popular responses. These responses were followed by the idea that nothing needs to be done - either because it will change itself gradually or because women are just not interested in this field. Some representative quotes included the following:

> "STEM education in schools has to be more gender-neutral. *Boys culture* in STEM-fields has to be discouraged."

> "Nothing. Working in science, engineering or technology is a mindset. If it interests you, you will pursue a career in it. If you don't, you will do something else."

> "Make the workplace more welcoming; offer more flexibility in schedules."

> "More mentoring and outreach programs for college and high school girls."

A significant difference between women and men was found on how a gender-mixed team would affect work. Most women and men, but more women than men, believed that working in a gender-mixed team would have a positive effect (84% of 252 female and 66% of 262 male respondents to Question 21, $p < 0.05$). A similar significant difference was observed on the effect of a gender balanced HPC community. The majority of both women and men, but still more women than men, believed that a gender-balanced HPC community would have a positive effect (84% of 248 female and 74% of 256 male respondents of Question 22, $p < 0.05$). Figure 3.16 summarises these findings.

Most of the 87 women who explained their answer to Question 21 and the 65 women who explained their response to Question 22, refered to working in a mixed-gender or gender-balanced environment as positive because it could encourage women and it could bring a variety of views and ideas. A few women explained that gender did not matter to productivity and what matters was the personality and the quality of work. Quotes from their responses included:

> "Different genders consider things in different ways, which improves the output of the team."

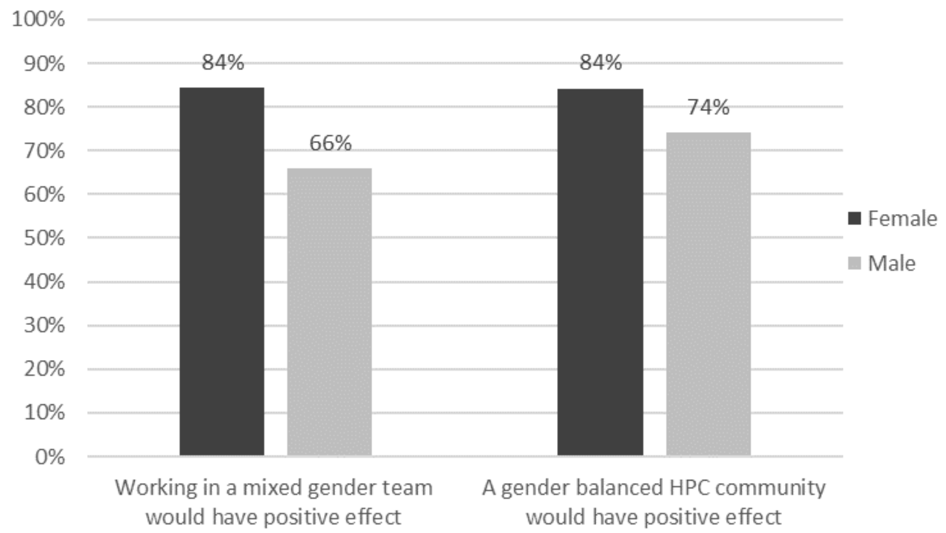


Figure 3.16 *Percentages of women and men who believed that working in mixed gender teams and a gender balanced HPC community would have positive effects.*

> "It is down to individuals in team to work productively, to achieve this, gender balance I don't believe is important."

> "Different approaches and ways of thinking can complement each other."

> "If we have a more gender balanced HPC community there will be more visible role models so this could inspire young girls to get into the field."

More than half of the 88 men who gave an explanation to Question 21 and the 78 men who did the same for Question 22, believed that working in a mixed-gender team was positive because it could bring variety of views and skills, and it could encourage minorities. Some men also mentioned that only personality should be a factor for productivity and not gender, and some men supported the idea that gender balance should not be forced and should not have any effect on the community. A few of them also explained that they did not have enough data or experience to be able to acknoelwdge a difference between mixed and non-mixed-gender teams. Some of their responses were:

> "I think working in a mixed gender team means that, often, a larger diversity of opinions is given, which strengthens the team."

> "I don't believe someone's gender affects their expertise or ability to contribute to a group."

> "A variety of approaches and views is always good in a team and one thing that can help towards this is a gender balance."

> "Gender balanced inherently has no effect, but the mechanisms to achieve it may have other effects on the community."

Finally, remarkably more women than men believed that they belonged to a minority group (Figure 3.17, $p < 0.05$). Half of the 250 women who responded to Question 23, believed that they belong to a minority group. The majority of those who specified (n=93) named the minority that they belonged to using the words *"Woman"*, *"Women"*, *"Female"*, *"Gender"*. Sometimes those words were followed by other words such as *"in STEM"*, *"in tech"*, *"in HPC"*, *"in Computer Science"*, *"with children"*. Few of them refered to their nationality, race and/or religion (for example: *"Jewish"*, *"Latin-American"*, *"Indian"*, *"non-UK"*), and fewer to their sexual orientation (such as: *"Lesbian"*, *"Bisexual"*). Only 16% of the 263 men who responded to Question 23 believed that they belonged to a minority, with the majority of them specifying it as related to race and/or religion (such as: *"Hispanic"*, *"Foreigner"*, *"Jewish"*, *"Non-white"*), and a few of them to their sexual orientation (for example: *"Gay"*, *"LGBT"*).

3.4.2.5 The Impact of Mentoring and Stereotype Threat

Mentoring is considered one of the best ways to retain women and help them develop their career in STEM fields, and research has shown its positive effect on women's academic development and career progresions in STEM fields (section 1.1.4.2). The responses to the mentoring related questions of our survey (Questions 8, 9 and 10) confirmed that mentoring was a positive experience for both women and men, with women being more eager to have a mentor.

Most women (44% of 149 female respondents) and men (42% of 180 male respondents) reported that they were (currently) providing mentoring to the part

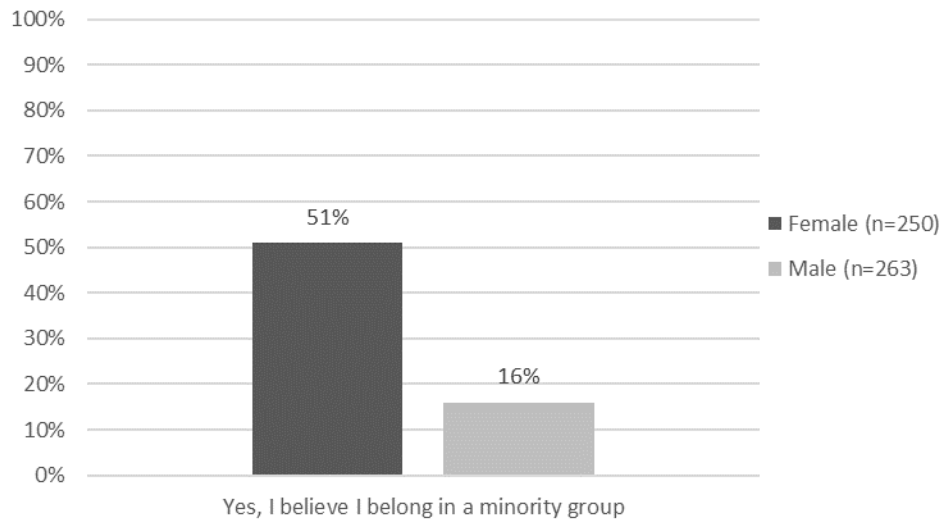


Figure 3.17 *Percentages of female and male respondents who believed that they belonged to a minority group.*

of Question 8 concerned with being a mentor. The least popular response here was the one where the respondent stated that they had never been a mentor and they did not want to (12% of 149 women and 8% of 180 men).

However, there appeared to be a significant difference between women's and men's responses to the part of Question 8 concerned with having a mentor ($p < 0.05$). Women's responses (n=206) were equally spread among the options: currently having a mentor (33%); previously having had a mentor, but not currently (32%); and not having a mentor but wanting to (30%). Whereas most men (40% of 201) reported having had a mentor previously, but not currently. When we compared only the responses of women and men who reported never having a mentor, there was a significant difference ($p < 0.05$) between women and men who would like to have a mentor (86% of 71 women, 55% of 66 men) or not (14% women, 45% men) (Figure 3.18).

The vast majority of both women and men (97% of 138 female and 98% of 161 male respondents to Question 9) stated that mentoring had been a positive experience for them. This question concerned both being a mentor and having a mentor. Subsequently, Question 10 asked the participants to explain in an open-text response *which aspects of mentoring made it a positive or a negative experience*. The views of 96 women, who experienced mentoring positively and responded to that question, were equally divided into two groups: those who

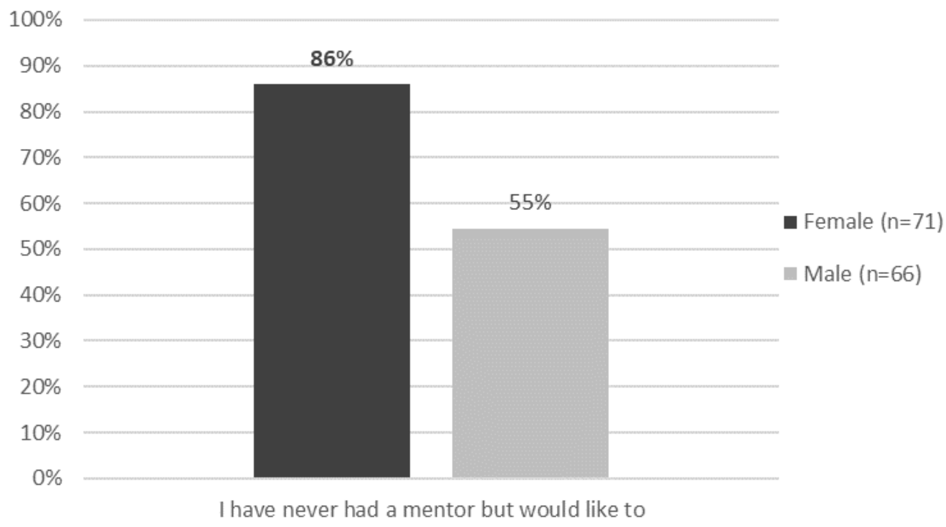


Figure 3.18 *Percentages of female and male respondents who reported not having had a mentor but wanting to have a mentor. There is a significant difference between women and men, with the majority of women wanting to have a mentor.*

considered mentoring as a way of sharing new ideas, passing experience and knowledge, learning and self-developing; and those who saw mentors as people from whom they could seek for support, discussion, help, advice, guidance and feedback. There was a very small portion of women who described mentoring as a negative experience, because of not good match with the mentor or because it was time consuming. Some of the responses of women participants included:

> "[Mentoring] increases motivation, the mentee and the mentor recognise their abilities and limitations, develop communication and organizational skills."

> "In terms of being a mentor, being able to be a resource for someone else was fulfilling. In terms of receiving mentoring, encouragement that I was not alone in the field and having someone outside of my current team with whom to talk about career topics has been great for maintaining a positive mentality in general."

A similar picture was obtained by the 85 men who answered that question and described mentoring as a positive experience. About half of them described mentoring as an opportunity to exchange new ideas and different views, to learn, to share knowledge and experiences, and to self-develop. The rest of them used mentoring for support, guidance, advice, feedback, as well as for engagement with students and new staff members. Very few described mentoring as negative or neutral experience. Responses by men included:

> "Learning from a more experienced mentor saves a lot of time. And inspiring an interest in somebody else is very rewarding."

> "Someone to bounce ideas off, get advice on careers, which is all valuable."

In Chapter 1 (section 1.1.2.2), we mentioned that one of the considered reasons of the underrepresentation of women in STEM is Stereotype Threat (ST). We included two questions (Questions 24 and 25) related to ST to examine if ST is a likely reason of women's underrepresentation in the HPC community as well. The majority of women (63% of 251 women) and men (77% of 265 men), who responded to Question 24 (*"Have you heard of the term 'Stereotype Threat'?"*), had not heard of ST before, however more women than men had ($p < 0.05$).

In Question 25, we provided a definition of ST: *"Stereotype Threat is when a person believes that their performance in a particular situation may confirm other peoples negative stereotypes about them"*. We then asked the respondents if they had ever felt that they confirmed other people's stereotypes. Interestingly, about half women (52% of 248 women) stated that they had felt that they confirmed stereotypes. The percentage of women who had felt it though, is signifiicantly higher than that of men (22% of 261 men) ($p < 0.05$). Figure 3.19 summarises the significant differences between women and men related to ST.

3.4.3 Summary and Discussion

In section 3.4, we described the design process of our HPC community survey, the data collection and their analysis, and we finally presented the results of quantitative and qualitative nature. We summarise below the main findings of this survey:

- The vast majority of both female and male respondents' background or current workplace belonged to the STEM group (88% female and 94% male);

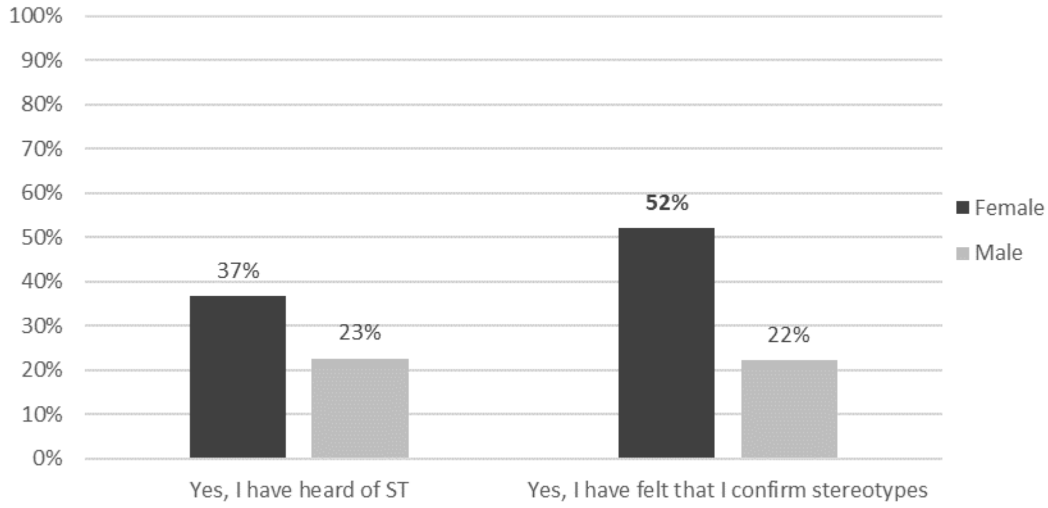


Figure 3.19 *Percentages of female and male respondents who had heard of Stereotype Threat (ST) before (left, 251 female and 265 male respondents), and of those who had felt that they confirmed other people's stereotypes (right, 248 female and 261 male respondents). Significantly more women (in bold) than men stated that they had felt confirming stereotypes.*

- More women and men from the STEM group (63% and 54% respectively) than the non-STEM group (45% of women and 46% of men) appeared to interact with the HPC community as *"Users"*, whereas more women and men from the non-STEM group (36% and 29% respectively) than the STEM group (12% of women and 16% of men) seemed to interact as *"Trainers"*;

- No significant differences between women and men were noticed in: personal success rating (3.23 average for women and 3.35 average for men); where they want to be in years (58% of women and 61% of men want to be *"in a more senior role"*); the workplace sector (70% of women and 72% of men chose academia); and highest qualification (48% of women and 56% of men holding a PhD); which suggest that women in the HPC community are not less qualified, confident or ambitious;

- Significantly more men than women had a tenured academic position (14% of men and 8% of women) and a *"Systems administrator"* job (9% of men and 3% of women); whereas more women than men picked *"Marketing"* as their current role (6% of women and 1% of men) and *"Other"*, which they described as *"non-academic research"* and *"management"* (19% of women and 8% of men);

- No significant difference was noticed in women's and men's familiarity with the examined Operating System (OS), with both women and men being more familiar with Linux (4 average familiarity for women and 4.4 for men). However, women from the non-STEM group were by far less familiar with Linux (2.8) than Windows (4.4), which might be a barrier for them for using HPC facilities;

- Half of the female respondents (51%), but only 25% of male respondents, did not develop their own software;

- More men (64%) than women (51%) had received software training;

- The majority of men received training and developed their own software (54%); whereas women were twice as likely as men to have received training and not develop their own software (20% women and 10% men);

- Both women and men from the non-STEM group were less likely to receive training (34% and 45% respectively), and to develop software (26% of women and 55% of men), than those from STEM disciplines (58% of women and 67% of men received training, 58% of women and 81% of men developed their own software);

- More women than men with children (75% and 62% respectively), and without children (85% of women and 75% of men), believed that having children had or would have impacted their career;

- More men with children appeared to travel to conferences *"more than 3 times per year"* (23%) than women with children (13%) and than men without children (13%);

- More men than women would move to another part of the country (70% men and 60% women) or to another country (64% men and 59% women) for work purposes;

- Men considered more than women that there was no discrimination in the workplace and all colleagues were treated equally;

- Most women believed that women need to work more to prove themselves, for less money and fewer opportunities for promotion than men;

- More men than women believed that there was no discrimination at conferences and that all attendees were treated equally;

- Both women and men who believed that there was discrimination at conferences mentioned gender-based assumptions of women's position or skills as the main case;

- The majority of both women (81%) and men (71%) believed that there is an underrepresentation of women in the HPC community and that there should be more women;

- Most female and male respondents focused their suggested ways to increase the proportion of women in the HPC community on girls' education, female role models and change of the image of HPC;

- More women (84%) than men (66%) believed that working in a gender-mixed team, and that having a gender-balanced HPC community (84% of women and 74% of men), would have a positive effect. The main explanation of both women and men was that there would be a better variety of views, skills, and ideas;

- Remarkably more women (51%) than men (16%) believed that they belonged to a minority, and the majority of them specified it as related to gender in a STEM related field';

- The vast majority of both women (97%) and men (98%) described mentoring as a positive experience, and as an opportunity to share ideas and to self-develop, or to seek support and advice;

- More women (37%) than men (23%) had heard of Stereotype Threat (ST), and more women (52%) than men (22%) had felt confirming stereotypes.

Conclusively, our survey findings indicate that in the HPC community, most women and men come from a STEM background, and believe that there is an underrepresentation of women in the community. Additionally we found that women are less likely than men to receive training and to develop their own software. Women are also more likely to have been affected by parenthood, to travel less often to conferences, to feel discriminated at workplace and conferences, to be interested in having a mentor, and to feel that they belong in a minority group. Finally, there were clear differences in training and software development tendency, and OS familiriaty between the STEM and the non-STEM groups.

Even though with this survey we reached a large number of people from the HPC community, without existing demographics of the community, we cannot be sure that the number of responses is representative of the community. Additionally, the survey reached people globally unlike the other surveys, which were conducted locally (UK), and some of its results may differ due to culture, policies and other reasons.

We recognise that it would have been useful if we had included a question asking for the area where participants worked and lived; however, this question was considered ambiguous and sensitive, and it was ommited in an effort to keep the lenght of the survey short and obtain more important information on the charactersitics of the community. We also accept the threat to validity of the results due to likely miscomprehension of a question or the purpose of the study. We tried to avoid miscomprehension by providing open-text options for the participants to explain and specify. We also took into consideration in the analysis additional feedback on the survey received by email.

Another lesson learnt from designing and conducting this survey is that the length of the survey should have been shorter by not including the questions related to the specific conferences (SC and ISC), as well as some very specific questions which despite being relevant appeared to be out of context or/and leading, for example the questions about Stereotype Threat (ST). Even though we had a very good response to our survey, a shorter survey is more likely to be completed by more people. From analysing the survey data, we also realised the importance of having planned ahead on how the survey questions can correlate to each other and on what conclusions we can draw from these correlations. Even though we conducted the pilot short questionnaire (section 3.3) before the HPC community survey, a trial of the actual survey might have helped to improve its final version.

3.5 The Use of HPC by Non-STEM Disciplines

Our results so far have shown that the majority of the members of the HPC community come from a STEM background (section 2.2.3.4 and section 3.4.2.1). We hypothesised that since there is an underrepresentation of women in STEM (1.1.1), this might be one of the main reasons that there is possibly an underrepresentation of women in the HPC community. This hypothesis was also supported by the participants' responses to the HPC community survey (sectioon 3.4.2.4). Consequently, an effective way to increase the number of women using HPC could be attracting more people from non-STEM disciplines, such as Humanities and Social Science.

There are other studies which also proposed that interdisciplinary research, among STEM and non-STEM disciplines, would benfit innovation and diversity, and it would improve the impact of technology on society [146], as well as creating new career pathways and improve diversity in STEM [147]. Additionally, studies suggested that there is gender inequality in the use of technology in some non-STEM dsiciplines, which has implications not only for the lack of equal opportunities, but also for perpetuating gender stereotypes and the absence of women's participation to progress [148], [149].

An online survey was conducted to further investigate the underrepresentation of individuals with non-STEM background in the HPC community. The main purpose of this survey was to obtain evidence, which would support or reject our hypothesis that HPC is not popular in non-STEM disciplines and its implications for the community's gender balance.

The survey was designed using Google forms[9], and it was open to the public from the end of July 2017 until the end of September 2017. The invitation to participate was distributed via email (available in Appendix B.4) to several non-STEM departments and schools of universities in the UK. The invitation mainly aimed at individuals with late- or mid-career positions and administrative staff, who could circulate it to the members of academic and research staff, and the postgraduate students of the school. We received a total of 26 complete responses, of which 18 were from women and 8 from men.

Due to the small number of responses, we recognise that the results of this survey are not conclusive nor representative of the population. We assumed that one reason for receiving such little response to this survey could have been the biased approach to the topic in the invitation of participation (Appendix B.4) and the different way of the survey's design (starting by asking the main research question, Appendix B.4), which might have caused misunderstanding of the purpose of the study by some of the invited participants (for example, wrong impression that the study concerned STEM disciplines, according to emails received during the period when the survey was being carried out). Another reason for the small number of responses might have been the restricted period of time for which the

[9] `https://www.google.com/forms/about/` *Last accessed: October 2018*

survey was open to the public.

The main conclusions of the analysis of this short survey were: 2/3 of the female and 1/2 of the male respondents did not know what HPC was; none of the respondents used Linux; and the vast majority of the Schools, where the respondents worked or studied, did not offer computational courses. As mentioned earlier, the sample was too small to draw conclusions; however, this snapshot of non-STEM polulation confirmed our hypothesis, and gave us an idea of what we could expect. Further investigation is strongly recommended and encouraged.

Chapter 4

Personal Experiences: Qualitative Approaches

The aim of the qualitative approaches was to gather further evidence from the HPC community on the reasons that might cause the underrepresentation of women and the suggested practices for improvement. We designed and developed these approaches based on the results of the quantitative and semi-quantitative approaches that were discussed in Chapters 2 and 3, and on the background work on related topics, which was discussed on Chapter 1. For the collection of data for qualitative analysis, we conducted interviews and focus group discussions with selected and recommended individuals of the community.

4.1 Ethical Requirements

4.1.1 Ethics Review and Data Management Plan

According to the Policy for ethical research of the University of Edinburgh, parts of the research methods and data of this study were subject to ethical review because of the involvement of human participants. Two ethical review self-assessment forms and a data management plan were submitted to the Ethics and Integrity Officer of the School. The ethical review forms included details of the project and self-assessment questions. The data management plan included information about the purpose of the project, the type of data, data storage and preservation, confidentiality and data sharing. All the documents were reviewed and approved by the Research Ethics Board of the University of Edinburgh.

4.1.2 Consent Documentation

Following the regulations for ethical research, consent documentation was produced and provided to the participants of the interviews and of the focus group discussions conducted for this study. This documentation included an information sheet to fully inform the participants on the terms and conditions of their participation, and a consent form, which was signed by the participants before the data collection. The information sheet covered the purpose and details of the research, the procedure of participation, benefits and risks, terms for withdrawal, strategies for ethical use of the data, and contact details for participants wishing to raise concerns and questions.

The consent documentation was read before the beginning of each interview, it was verbally agreed by the participants who were interviewed either in-person or through a video call, and it was audio recorded together with the interview of each participant. The participants who chose to be interviewed via email received the consent in written form, which they had to sign and send back before they received the interview questions. The participants of the focus group discussions were provided with the consent documentation in written form, which was read and signed by them before the beginning of the discussion. All versions of consent documentation are available in sections C.1 and C.2 of AppendixC .

4.1.3 Working with Personal Data

To protect the participants' identity as stated in the data management plan and the consent documentation, direct and non-direct identifiers of the participants were removed or recoded. Direct identifiers included name and institution or workplace, and non-direct identifiers included occupational details, location, age, sexual orientation, health conditions or any othe presonal information that had been mentioned by the participants during the interview or the focus group discussion. The names of the interviewees and the focus group discussion participants were replaced by the words "Female" or 'Male" and numbers or capital letters; for example, "Female 1","Male E". The names of institutions and workplaces and non-direct identifiers were replaced with generic descriptive words or categorical groups; for example “a British university”, “late-career stage group”, “a STEM subject”. A more detailed explanation of the recoding procedure is provided in the respective analysis sections.

4.2 Interviews

The purpose of interviewing was to gain an in-depth understanding of the experiences of the group of people of our interest. Our plan was to interview men and women from a variety of backgrounds, job positions, age and career stages, that self-identified as part of the HPC community. Our final set of interviewees included people who were working in the HPC community at the moment of the interview, people who had worked in an HPC environment before but they had moved to a different or non-technical career, and students and researchers who were HPC users. Our target number of interviews was maximum 30, but the progress of the research and the interest of the interviewees led us to a final total number of 42 interviews (25 women and 17 men), when we decided that we reached the data saturation point in our study, since there was some repetition and no new information in the interviewees’ responses [150]; we also took into consideration the amount of time needed for the analysis of these data.

We used a structured/semi-structured type of interviewing, which involved a predetermined interview plan, organised around a set of scheduled questions, with the possibility of other questions emerging from the discussion [151, 152]. We developed an initial plan of 16 questions, which included some questions almost identical to the ones of the HPC communtiy survey, and questions which were generated from the results of the survey analysis and from hypotheses developed from studying further resources of related topics. Four more questions (Questions 4, 10, 17 and 18) were added after the first couple of interviews. All the questions asked during the interviews are included in the interview script presented in Appendix C.1.

To open the discussion gently, the first group of questions dealt with personal questions about the interviewee’s background, career route and plans, and their relation to HPC (Questions 1-5). With these questions, the participants presented themselves and explained how they became part of the HPC community. Additionally, the responses to these questions offered all those features that could define the HPC community, such as background studies, research area, reason for HPC use, job title, career route. The next group of questions dealt with the interviewee's and the community's potential challenges (Questions 6-10). The aim of those questions was to identify potential difficulties - technical or social - that one might encounter when joining the HPC field and how to overcome them, and suggestions to make the HPC community more accessible. Here, we included a question about training (Question 4), since training together with software development and specific OS (Unix/Linux) are essential for using HPC facilities, which showed significant differences between women and men in previous chapters (2.2.3, 3.2.2.1, 3.4.2.2). Another important finding of this study was the large difference in the number of people from STEM background and the number of those from non-STEM disciplines who form the HPC community, with STEM being significantly more popular. This matter was also discussed in that part of the interview (Question 10).

The following set of questions concerned some of our hypotheses on potential reasons of the underrepresentation of women in the community, gender related issues, as well as our hypotheses on ways to improve equality and diversity in the community (Questions 11-18). Some of these questions examined parameters that had been examined through the HPC community survey questions (3.4), for example the impact of mentoring or of Stereotype Threat. We also included questions based on popular responses to questions of the HPC community survey related to the reasons of women's underrepresentation such as the lack of role models. Finally, in this set of questions we also included questions that we designed based on findings of related studies on other fields, such as the impact of a team's gender balance on the team's productivity (1.1.3).

The closing question asked their opinion on how to achieve gender balance and diversity in the HPC community (Question 19). By the end of the interview, to gradually conclude the discussion, we asked the interviewee if they had additional comments or questions (Question 20), and we thanked them for their time and their contribution to this study. We anticipated the length of the interview to be 45 minutes to one hour.

The first call of invitations took place in November 2016, and the invitations were firstly sent to a small number of people that we knew in the community or we met at HPC conferences. We reached the rest of the interviewees by the "snowballing" process, in which the participants introduce to the researcher other individuals with similar or relevant characteristics [150, 152]. The interviewing period lasted from December 2016 until June 2017. The invitation (available in section C.1 of Appendix C) included the purpose of the interview and the general topic of the study to be discussed, but no sample of the questions, as we wished the discussion to be spontaneous and the interviewees not to have their answers prepared.

4.2.1 In-person Interviews and Video-call Interviews

Interviews can be conducted in several ways: face-to-face or in-person interview, phone interview, on-line video call or e-interviews. This study aimed to use in-person interviews for the people based in the local area or willing to travel, and on-line call interviews for the individuals located in other areas.

4.2.1.1 Data Collection

I conducted 37 interviews, of which 19 were in-person and 18 via on-line call. For the in-person interviews, I met the interviewees at a quiet place in their work environment, and I used a Dictaphone to record their consent and interview. During the time of the interview, I was keeping notes of important or interesting information given by the interviewee, which helped me lead and follow the conversation. The notes proved very useful also for archiving the interview data and for the analysis.

For the on-line interviews, I used the on-line video call tool "Skype"[1]. The procedure was the same as for the in-person interviews. However, due to technical challenges, there were incidents of call interruption, poor sound quality and lack of video, which sometimes affected the flow of the discussion, the time needed, and the quality of the audio data.

The recorded interviews were transcribed by using online transcription software "VoiceBase"[2], which converts uploaded audio files into text documents (the privacy policy of the data is available on the website provided; additionally, all our files were immediately removed from the online account). The transcripts were then edited manually, based on the original audio data and the notes taken during each interview, and were organised into files for analysis. Data from the

[1] `https://www.skype.com`*Last accessed: October 2018*
[2] `https://www.voicebase.com`*Last accessed: October 2018*

in-person and the on-line interviews were treated as one dataset.

4.2.1.2 Analysis and Results

For the analysis of the interviews, I used the thematic analysis approach, namely by organising individual excerpts and passages from the transcripts into categories with thematic connections to conclude responses to our questions and hypotheses [150]. The categories were based on the themes of the groups of questions explained at the beginning of this section (4.2).

As mentioned in section 4.1, I replaced names, titles, affiliations and other charecteristics considered as indentifiers of the interviewees. The gender of the interview with a number formed the identity (ID) of each interviewee. All other identifiers were replaced by generic words in brackets; for example instead of *"Physics"* we wrote *"[a STEM discipline]"*. Also, words that we added in order to make sentences of the quotes comprehensible were added in brackets; for example *"I had to attend introductory [courses] on HPC"*.

The main aim here was to gather collectively views and thoughts of women and men of the HPC community, and not to compare them between the two genders. For this reason, unless there was a significant difference in the responses of the two genders, I present the main points of the reponses to each topic as a whole; otherwise, the gender is explicitly stated.

4.2.1.2.1 Background and Career Route of the Interviewees Table 4.1, 4.2 and 4.3 display the identity and background characteristics of the 37 interviewees (21 women and 16 men) divided by career stage. The career stage was determined by the number of years of working according to the interviewees' responses. Early-career stage stands for 0-10 years, mid-career stage stands for 11-20 years and late-career stage stands for more than 20 years. The "Gender/ID" is the coded ID with which we refer to each participant in this section, and it consists of the interviewee's gender and a number from 1 to 37. The "Discipline" group informs us if the interviewee has a STEM or a non-STEM background; "Degree" stands for the interviewee's highest qualification.

We managed to have equal numbers of representatives of the three career stages: 12 interviewees were from the early-career stage group, 13 from the mid-career stage group, and 12 from the late-career stage group. However, the majority of women (16 out of the 21 female interviewees) belonged to the two first career groups; whereas almost half of the male interviewees belonged to the late-career stage (7 out of 16). We assumed that the smaller number of female interviewees from the late-career stage was reflected the lack of women in more senior positions or due to the fact that women had joined the community more recently.

Table 4.1 *Identity and background of in-person and on-line interviewees in an early-career stage. (A) stands for academia and (I) stands for industry.*

Gender/ID	Current Position	Previous Position	Discipline	Degree
Female 1	HPC Applications Analyst (I)	PhD Student	STEM	PhD
Female 2	Research Community Manager (I)	Training Leader (A)	**STEM/ non-STEM**	PhD
Male 3	Postdoctoral Researcher (A)	PhD student	STEM	PhD
Female 4	Postdoctoral Researcher(A)	PhD Student	STEM	PhD
Male 5	PhD Student	Student/Intern	STEM	PhD
Female 6	Lecturer (A)	Postdoctoral Researcher	STEM	PhD
Female 7	Research Staff (A)	Statistician (I)	STEM	PhD
Female 8	PhStudent	MSc Student	STEM	MSc
Male 9	MSc Student (A)	Programmer (I)	STEM	BSc
Male 10	Project Management Officer (A)	Administrative Staff (I)	**non-STEM**	MSc
Female 11	Applications Developer (A)	MSc Student	STEM	MSC
Female 12	Data Architect (A)	PhD Student	STEM	PhD

When we mapped the route of the career of our interviewees according to their responses to Question 1, we noticed that the majority of those who were currently working in academia were coming from an already academic background (either as academic staff or postgraduate or undergraduate students, 24 of the 37 interviewees). Fewer interviewees (8) moved from industry to academia. This was a route chosen mostly by individuals of the early- and mid-career stage; for example, few of them had worked at a company after their studies and then moved back to academia. They did not explain the reasons for taking this route. However, based on some of the responses to the questions about their future plans and the reasons they chose to join the HPC community (Questions 2 and 3), we assumed that their choice was related to their view that academia offers larger

Table 4.2 *Identity and background of in-person and on-line interviewees in a mid-career stage. (A) stands for academia and (I) stands for industry.*

Gender/ID	Current Position	Previous Position	Discipline	Degree
Male 13	Chief Information Officer and Professor (A)	Lecturer and Company Founder (A and I)	STEM	PhD
Male 14	Director of Supercomputing (A)	Computing Manager(I)	STEM	PhD
Female 15	Lecturer (A)		STEM	PhD
Female 16	IS Department (A)	Software Developer (I)	STEM	MSc
Female 17	Research Software Develper (A)	Software Engineer (I)	STEM	PhD
Female 18	Research Fellow (A)	Postdoctoral Researcher (A)	STEM	PhD
Female 19	Research Mangement (A)	Postdoctoral Researcher(A)	STEM	PhD
Male 20	Director of Institute (A)	Software Developer (I)	STEM	BSc
Female 21	Research Fellow (A)	Medical Support (I)	**non-STEM/ STEM**	PhD
Female 22	Software Engineer (I)	Technical Staff (I)	STEM	PhD
Male 23	Project Manager/PhD Student (A)	Applications Consultant (A)	STEM	MSc
Male 24	Project Manager (A)	Applications Consultant (A)	STEM	PhD
Female 25	Applications Consultant (A)	PhD Student	STEM	PhD

variety of research projects, better career progression, flexibility, as well as more security.

Additionally, as Tables 4.1, 4.2 and 4.3 display, the majority of the interviewees worked at or/and came from a STEM discipline; Only three of the interviewees moved between the two environments: Female 2, who moved from a STEM to non-STEM; Female 21, who moved from non-STEM to STEM; and Male 10, who joined the HPC community at an administrative position from a non-STEM background. These results agree with our findings, discussed in previous chapters, with the majority of the people who consist the HPC community coming from

Table 4.3 *Identity and background of in-person and on-line interviewees in a late-career stage. (A) stands for academia and (I) stands for industry.*

Gender/ID	Current Position	Previous Position	Discipline	Degree
Female 26	Research Software Consultant (A)	Research Fellow (A)	STEM	PhD
Male 27	Lead Scientist (I)	Head of HPC Service	STEM	BSc
Male 28	Head of Research (A)	Software Developer/ Postdoctoral Reasercher (A)	STEM	PhD
Female 29	IT Project Management (I)	IT Library (I)	STEM	BSc
Male 30	Associate Professor (A)	Lecturer(A)	STEM	PhD
Female 31	Director of IT (A)	Manager IT (A)	STEM	MSc
Female 32	Head of Department (A)	Senior Lecturer (A)	STEM	PhD
Male 33	Software Architect (A)	Training Officer (A)	STEM	PhD
Female 34	Applications Consultant (A)	Programmer (A)	STEM	PhD
Male 35	Applications Consultant (A)		STEM	PhD
Male 36	Software Architect (A)	Applications Consultant (A)	STEM	PhD
Male 37	Group Project Manager (A)	Programmer (A)	STEM	PhD

a STEM background, where there is evidence of underrepresentation of women (1.1.1)

Most of the interviewees expressed interest in staying in their current position and/or continuing research and using HPC. Very few were thinking that they would like to move to a more senior or less technical and more managerial role, and only a couple of them stated that they were not interested in academic career and/or they were open to other opportunities. There was no difference between the two genders, as these views were equally reported by both men and women.

> "I think I'd like to see how far I can go [with research]. I'm actually trying to convince people that [HPC] is the only way to do (research); they shouldn't be working on single servers [...] But I find that people are quite reluctant, because there is quite of a learning curve."

(Male 3)

> "I'm always looking out for opportunities and I suggest everybody does that, to be honest, it doesn't hurt to keep your CV up-to-date. But I think because I have a reasonably young family now, staying at one place is probably more important at the moment than finding new career opportunities."

(Male 24)

> "I don't find this job currently very intellectually stimulating [...] I'm not convinced managing people in academia is a good way to go."

(Female 26)

> "At the moment, I've no real desire to go looking elsewhere; I have a fairly interesting job that pays me a decent amount of money, it's not too stressful, I enjoy most of what I do, and the things that drive me mad are not enough to make me leave."

(Female 34)

In the responses of the female and male interviewees to Question 3 (*"Why did you choose an HPC career or to use HPC facilities?"*), there were two equally reported main reasons for starting using HPC and joining the HPC community: 1. they had to use HPC for their research, otherwise they cannot do their work; and 2. they were interested in computing or they prefered computational research than lab work or theory. The fact that HPC is a very good tool for advanced research and it should be the only direction for those who want to do *"good"* research was also mentioned often in the responses of both female and male interviewees.

> "[Using HPC] is mostly something that just kind of happened [...] Throughout my early degree, I suppose from lack of knowledge and ability to actually program, I avoided interacting with computers very strongly [...] But to be able to do research which other people appreciate and find meaningful and useful, you need a computer; so

> I pick up those skills by necessity [...] I enjoy it more now, and there are a lot more opportunities outside, if you have some background in it."

(Male 5)

> "I definitely didn't want to do like proper [science] and I was more interested in that computational side of things."

(Female 11)

> "During my degree I didn't particularly enjoy the subjects that had anything to do with it [HPC] but then in the job that I had in industry, we were doing data processing, so HPC started happening there; I kept on working on it and I just became more involved, and now I like it."

(Female 16)

> "I have always been interested in computers, [...] but at university I didn't just want to do a pure computer science degree but rather something that actually had applications. That was kind of how I came about doing computer simulations."

(Male 23)

From those who were asked the question related to training courses (Question 4, added later) very few interviewees stated that they had attended programming training courses. Most intereviewees (more female than male) mentioned that usually there was no formal training, and they learnt programming and/or using HPC while working either by themselves or with help by others. Additionally, more men than women stated that they were interested in programming since early years of their life. Finally, the majority of those who moved to a non-technical or a managerial position explained that they had to attend courses not related to HPC or programming.

> "I hardly knew how to open a computer, frankly speaking. So, I was really thoroughly guided by my supervisor [...] I had definitely no

background in computers or computer science [...] of course I attended several different courses at the faculty of computer science."

(Female 8)

"I didn't get training. I got training on the job yes. But, I did feel that at the beginning there were some things I didn't know, because I didn't have an informatics or a computer science degree; people were talking about some things that I wasn't familiar with [...] there's lots to learn; and I don't think there are courses that would teach me things."

(Female 12)

"When there were specific problems, there was a helpdesk that I could email, and they would help look into the problems; that was the kind of support they would give. They didn't train people in how to use it (HPC), they would solve specific issues if there were. So it's not quite the same thing."

(Female 22)

"I did very little [programming] in my undergrad. I did more as a kid at home. So, I knew how to program computers before I started my degree; but it wasn't until my PhD that I was introduced to parallel computing, which is the basis of HPC [...] The ability to program was useful [...] It was all hands on and we were learning together with the group.

(Male 37)

The interviewees were also asked what they would do, if for some reason they could not use HPC for their work (Question 5). Most women and men would choose to do different nature of research which would require using smaller scale and slower computers or they would choose an IT/computational job, rather than a pure science related job. More women than men stated that without HPC, they would not be able to do their job or their job would not exist, pointing out the importance of HPC for some jobs or their personal interest in using HPC for their job.

> "Without HPC I can't actually do anything at all or what I can do is very limited [...] I think I'd drop out [...]."

(Female 15)

> "My core interest is in the computing side of things, so I suspect if I wasn't in HPC, I'd be in some kind of commercial software development or some IT; Rather than the primary thing being science for me."

(Male 23)

> "If it's going to be a day [without HPC access] or two it's not that big of a deal [...] I tend to neglect paperwork, because I like doing that (using HPC) [...] If it was to go on for more than a few days, I would be struggling, because there are deadlines to meet. [If there wasn't any access to HPC] then my job wouldn't exist."

(Female 19)

> "If HPC suddenly went away, I would have to change the focus of what I do; and a lot of the experiences and expertise and the reason that I am known internationally would have effectively gone away."

(Male 30)

Very few interviewees (mostly men) stated that they would do something completely different, not related to research or computing. Interstingly, a few of the male interviewees explained that they had had completely different studies and career aspirations, with no relation to computing; but due to the fact that they found programming easy or/and they had difficulties to follow their dreams, they ended up having jobs related to HPC. They also mentioned that they prefered flexible working hours or part-time jobs, so they could dedicate time to their other interests, and if they could not work in the HPC community, they would try to pursue a career related to their interests. Finally, few of the female interviewees interpreted the question in terms of not being able to use HPC because of lack of resources and funding difficulties, which would not allow them access to HPC facilities. As this seemed to be an issue with which they had dealt before in order to do their work, further research on the effective ways to tackle this issue is recommended.

> "My research [and] my career [...] have been fundamentally built on the applications of supercomputing to solving problems which are worth solving. What I'd do otherwise? I'd go and write books."

(Male 37)

> "Some of the things I do, I wouldn't really be able to do [without HPC]. But for the majority of the type of work I do I would say that I need it [...] There is always a way to find access to the resources, because now the resources exist; so if it's something that's necessary you will find a way to have access to it in order to do the work."

(Female 21)

4.2.1.2.2 Challenges in the HPC Community The main challenges that the majority of the interviewees, both female and male, met when they first entered the HPC community (Question 6) was their non-computational background and lack of knowledge in programming, using Linux/Unix or using HPC facilities, and the lack of respective formal training, documentation and support. These factors led most of the interviewees having had technical difficulties and a slow start in using HPC for their work by being mostly self-taught.

> "Since I don't have a background specifically in computing, I find things like judging how much memory you need or run times and the resource side of the computer, like how much power and time I would need for a particular type of analysis, quite difficult. The programming itself, I didn't find it as difficult, because I've done that before on a smaller scale."

(Female 7)

> "There was a little bit of a learning curve, but it wasn't very steep. I think mainly because I was already a very experienced Unix user; so I think for most people and my experience with training people to use it, their main problem in HPC, it's actually the Unix [...] once you have that then HPC is actually not that difficult to handle."

(Female 21)

A few of them mentioned that the only way to deal with those difficulties was receiving help by other HPC users or software developers. Very few people mentioned other challenges during their early time in HPC, such as difficulties in getting funding or lack of diversity in the community.

> "Just understanding the language of how to access these things, what it was all about, it was incredibly complicated. And basically, without being helped by a lot of the people who were already in the system, I might have not been able to go anywhere [...] it took me about six months trying to understand things and in the end I got some access, but it was really complicated."

(Male 30)

> "Sometimes you don't feel that you belong; but then I discovered that there are more people than you would think who come from different backgrounds. There are several actually from [non-STEM]. But I've always felt not included."

(Male 10)

> "Socially, it was not anything different than being in science, being in the technical world. I mean, you were not necessarily faced with blatant misogynism, but there were definitely underlying assumptions. I personally don't mind that much surrounded by men and and I collaborate [with] no problem, but I can see other people being intimidated by it."

(Female 17)

The majority of the responses to the next two questions (Questions 7 and 8), asking the interviewees if they were facing any current challenges and what they would like to change in the HPC community, mainly overlapped. This was due to the fact that most of the interviewees extended their responses into discussing issues of the general community that they would like to see changed, rather than just personal challenges. The most popular response to these questions was that the HPC community is not easily accessible and not particularly appealing to many researchers who have not used it before. The main explanation for this was that the HPC providers do not really interact wiht their users and the users' needs, which leads the HPC facilities not to be developed in a user-focused way, but more in a speed/power-focused or money-focused way. This also seems to be hugely connected with grant applications, as it was mentioned (mostly by women) that only applications which were written in a certain way or for specific projects usually succeed.

> "The problem I'm having at the moment is securing grant funding. I think that's partly because what we are proposing to do with our software development is extremely new and people are very nervous with new ideas. Resources into something that's not tried and tested is a very risky thing to do."

(Female 15)

> "Sometimes I worry HPC is driven for HPC's sake; [...] I think most people understand that it's a tool for scientists but of course, there's a game you have to play around getting funding, persuading people to sponsor you [...] [the] genuine reason for doing supercomputing is lost [...] so you end up doing stuff that you think people want you to do rather than things that you think are the best things to do, which can be a little bit frustrating."

(Male 24)

> "I think it's possible in the future HPC providers will have to think again about how they interact with the users and the facilities that

> they provide. And basically be much more adaptable to the needs and wishes of their users and make it very much simpler for users."

(Male 30)

It was also suggested that there needs to be a fine balance on how HPC and related programming are taught, so they are appealing and understood by both scientists/researchers (of all disciplines) and software developers. Some interviewees (more women than men) explained that the use of jargon, and the lack of training and of well-made documentation are not helpful to people who join the HPC community from non-computational background, but sometimes even from computational background.

> "The community of HPC I think sometimes speaks in terms that [are] very hard to understand [...] [Also] people always assume that you know a lot about complicated scientific problems, so when I go to the training courses they start with a scientific problem that I don't understand and everyone else does, because I'm a software engineer and they are researchers who have done more maths or physics. And for me that's challenging."

(Female 16)

> "The understanding that people in this field are not all software engineers [...] that there are some fundamental things I don't know, I haven't learnt, and I'm not expected to know."

(Female 17)

> "The university runs these courses every maybe twice a year, but then if you're starting your project now and have four or five months doing it, you can't afford to wait for the course [...] the training could be maybe more frequent or it would be good [if there was] some sort of like online training. There is the documentation, but it's not OK for someone who doesn't know anything."

(Female 21)

Other challenges and concerns that were mentioned often by some of the interviewees included: the lack of diversity in the community, and the lack of clear career paths and progress. More women than men mentioned the lack of diversity (not only regarding gender) as a problem of the HPC community's present and future, since the currently existing barriers lead to shortage of talent and skills. Some of the interviewees suggested as solutions to the diversity problem programming courses and training to be added in earlier school years or as part of all university degrees, and to keep addressing the lack of diversity and its effect in the community, so everybody is aware and willing to change it.

> "It feels male dominated. It feels like I am an outsider. It would be nice if there were more women. It would be nice to see how other women have done as reference for the future. I think that the HPC community is aware that there is an imbalance, but it seems like only women are pushing for change, I don't know how everybody else is pushing for equality and diversity. I have come across people who think that there is no problem.

(Female 1)

> "There is some ground work in that area, as you see young kids doing some coding, but perhaps making it more accessible and more visible to the wider public with online tutorials. Even if this starts off at the primary or even high school level, I think it gets students thinking; so when they get to university this is something that is not new and abstract."

(Female 6)

> "It would be nice if it (HPC) was part of their first degree or maybe like initial PhD stuff; like when you're inducted into the university, here's the admin staff here's the library and here's the computer cluster."

(Female 19)

> "There's certainly a discriminatory behaviour in HPC as a whole. There is a dominance of certain kinds of people in HPC and particularly in the leadership positions of HPC. [There is an] imbalance of diversity across a whole range of different attributes, but clearly gender is by far the easiest to focus on."

(Male 27)

> "I think it's one of the challenges we have within our universities, that we tend to be quite binary in the way we think about our staff, [in a British University] we have professional services staff and we have academic staff and of course when a lot of people working in HPC and other areas of research technology, they're very much borderline. And I think to really grow this community within the UK, we should be paying more attention to this [borderline] area"

(Female 31)

When the interviewees were asked if they believed that HPC was well promoted and what needed to be done to promote it (Question9), most of them pointed out similar aspects of the HPC community as the ones they wished to be changed, which were dicsussed earlier (responses to Questions 7 and 8). The vast majority of them believed that HPC is not well promoted and that it is a very closed community. They mentioned that HPC should be integrated in various university degrees and that programming courses should start earlier in school. Outreach was often mentioned as a way of promoting HPC.

> "it's not integrated into PhD programs as much [...] all of the things you want to do, the training and the access, are there, but they're kind of external of the things that people maybe engage with or of their institutional programs."

(Male 23)

> "I think we've got a deeper problem further back; it's hard enough getting people interested in doing computational science or science as a whole, and HPC is a niche within it. When we get to the position

where every single undergraduate student who does science, is also taking at least one module on computational science, then we can worry about HPC."

(Male 27)

Another popular response to this question was that HPC is not a popular tool to researchers from other than the *"traditional"* disciplines (STEM), which was a very good way to connect with the next question (Question 10) about the use of HPC for non-STEM research. Most interviewees would like HPC to be promoted and used by non-STEM researchers. However, some of the interviewees were not sure if the nature of non-STEM research demands the use of HPC, but most of them were very positive. A few of the interviewees explained that getting more people from the better gender-balanced non-STEM disciplines engaged in the HPC community might improve the gender balance of the community; which was the main reason that this question was added in the interview.

> "I imagine that disciplines like physics, engineering, chemistry, naturally use computers, because of the nature of the equations they are involved; sooner or later they'll end up using HPC. But other disciplines, maybe life sciences, biology, even social sciences, the people to whom it is not immediately obvious that they can get something out of HPC, they don't even think about it. And I don't think HPC community does much about it.

(Female 11)

> "There is a growing demand for HPC in those kind of fields (non-STEM). I think because of the nature of their research there is always going to be a smaller proportion of that field [in need of HPC] than if you go to a physics department [where] probably 50% of all the people might want to use it [...] there is certainly room for those communities to come and use HPC and as a result presumably improve the gender balance; but it's not going to solve it."

(Male 23)

> "If you spoke to most researchers in the university, they wouldn't have heard of HPC, or wouldn't know the difference between computing on a desktop computer and computing on a cluster or a supercomputer [...] I don't think supercomputing or HPC is important in every field. There are good examples where HPC is important but they're quite rare."

(Male 24)

4.2.1.2.3 Underrepresentation of Women in the HPC Community and Potential Reasons Only three of the 37 interviewees said that that they were not sure if there is an underrepresentation of women in the HPC communtiy (Question 11). The rest of them believed that there are fewer women than men in the community, and the main explanation they gave was the smaller number of women in the STEM subjects, where most of the HPC community's people come from. Most of the interviewees struggled to answer why there is an underrepresentation of women in STEM fields. The most popular explanation was the early exposure of children to the socio-cultural stereotypes about what women are and are not supposed to do, and the stereotypes treating science and technlogy as somehting scary, not connected with real life, and as something that only white *"nerdy"* men do.

> "The common thing to say is that men are more interested in these things (computers) [...] I have a friend who does a PhD in the same group with me; we have exactly the same background. But when I was a teenager, I was interested in something else than building computers, and of course now he is more experienced in actually how the hardware works; That was his hobby when he was a kid; my hobby was something else. But I don't believe that men in general are just more interested in it, but maybe they might be more interested in it in a certain age [...] and it's more advertised for them in a sense."

(Female 8)

> "Probably the way the society presents science is just in a slightly misleading and wrong way. And because men and women think

differently, it might affect children and the way they perceive world and a potential role in it and whether what they can or cannot do with their lives."

(Female 11)

"I don't know whether HPC is any different to computing in general or even science and engineering in general. I think that there's underrepresentation of women in all [these fields], maybe not biology so much. I think there's a reason starting from childhood that we're just the way we were brought up with the stereotypes of a very very young age and women often are pushed towards the more traditional women subjects [...] if you don't feel confident enough to go down to computing, you might as well feel less confident to go down to a specialised computing field as HPC."

(Female 22)

Some interviewees suggested that an important problem causing the underrepresentation is that programming and supercomputing are introduced very late in one's education or career; so, people (mostly women) who have to use HPC for the first time without having been exposed to it before, find it difficult and unattractive. It is even harder, especially for women, when there is no formal training and support. This concern was also discussed earlier as something that needs to be changed in the HPC community.

"The STEM subjects as a whole in most cases have an underrepresentation of women in the first place; and HPC is a niche within STEM [...] It's hard to find out that HPC exists when you're early in your career [...] it's not proactively advertised to a broad range of people; only [to] people of already self selected science degrees, which have also got a lower uptake amongst the underrepresented groups. Why don't women choose HPC? I don't know the answer to that at all. There are clearly behaviours exhibited within the world in general, but specifically within the HPC community, that make HPC an unattractive place."

(Male 27)

> “There are two points: we are not educating enough the female students and there is underrepresentation of female students in computing and other areas that feed in the HPC ; and second there are not undergraduate courses in HPC to increase the understanding of this area, there are not courses embedded in the curricula that promote the focus on HPC, somehow we stubble on HPC rather than plan to use it.”

(Female 32)

The fact that HPC is a *“club”* of people that supports certain groups of people, specific projects and STEM fields was another common response to why there is an underrepresentation of women in the HPC community. It was explained that this makes women, who are already a minority in these fields, less confident in wanting to be part of the community and using HPC. Another important point made by some of the interviewees was the lack of women in managerial and decision-making positions, which leads to fewer women being hired, promoted or funded. Those inerviewees expressed their concern in men dominating leadership positions, which might mean that senior men are more likely to support someone who looks like them and to continue cultivating the stereotypes of the community. Also, fewer women in senior positions means fewer role models for women in early- and mid-career stage, which we discuss extensively later in this section.

> “If you don’t know the jargon, if you don’t know how things work, it’s really hard to get in [the HPC community]. It took my whole PhD [time] to really be confident in what I was doing, and even when I finished my PhD, I still didn’t feel confident enough [...] Women don’t have people that they can actually go to for help, whether it’s because they don’t feel comfortable doing so or they feel intimidated.”

(Male 3)

> “In my collaborative team, there’s me and four or five men, every one of them is so absolutely vital to me every day. But whenever the grant from our group is submitted with their names on it, I think it always does better than when it has my name on it. I think in general women have to achieve more to get the same amount.”

(Female 15)

> "I think it is part of the fact there are no women at leadership, because they were actively discouraged ten-fifteen years ago. So, we haven't got the role models in there, because they've got rid of them."

(Female 26)

> "Yes, there is [an underrepresentation of women in the HPC community]. The meetings I go to, either in the broader European context or the university context or at all points in between, it's all just middle aged white guys. I think the representation at the senior management level seems to be worse than the technical level. Also, there aren't as many female applicants for entry level programming jobs."

(Male 37)

The majority of the interviewees, both women and men, did not have a clear opinion on the effect of gender balance and diversity on a team's productivity, mainly because of lack of experience in working with balanced teams (Question 12). However, many of them believed that diversity of opinions and views of people from different backgrounds could definitely have a positive effect; whereas a dominating group of people might affect negatively the minorities' behaviour and consequently the team's productivity. Some of the interviewees mentioned that they were aware of a study that has shown the differences in productivity between gender balanced teams and gender imbalanced teams (1.1.3). A few interviewees believed that gender or any other characteristic should not play a role in a team's collaboration, and it is only the individuals' personality and skills that matter. Finally, very few interviewees had experienced the difference between teams and explained that having more women in a team had a positive impact on the team as a whole.

> "If you get a dozen white males of elderly age of the same career background, in the same room, then the chances are [that] you got no different experiences and no different viewpoints to bring to anyway."

(Male 27)

> "By any content of measure, the groups that I led that had both male and female in, have outperformed the groups which were just women or just men. So, making sure that there's a diverse, healthy supply of people, men and women, but also with a range of different skills, is absolutely critically important. Simply because you can do better engineering, better science, if you have both women and men."

(Male 13)

> "I do find that when I' m in meetings with project teams that have more women in them, the flow of the discussion is more balanced, more people speak, which probably does improve things. Sometimes it is one of the senior man that just takes over the whole thing [...] I know studies that say that having a more balanced team does produce better results."

(Female 19)

> "I just don't look at people in terms of gender. I think it's more related to the personality of people rather than whether is a man or woman or old or young [...] But I do see how someone might be affected [...] But, I don't think that's something that should be forced; putting people based on their gender forcefully into a team to create some sort of artificial balance, that's a big no."

(Female 11)

Most of the female and male interviewees believed that female role models and leaders are really important for women to follow and pursue a certain career (Question 13). However, most of them could not think of an example or did not have a personal role model; which conclusively was a result of the lack of female role models in the HPC or even the broader computing community; As mentioned earlier, this might play a role to the underrepresentation of women in the community. A small group of interviewees supported that if someone really wants to do something, role models are not essential, or that the gender of the role models should not matter. Also, it was mentioned that even though role models are important, due to the very small number of women in the HPC community, some of those women are tired of playing this role and of being forced to participate at every conference panel or outreach activity; which emphasises even more the need for more women in the community.

> "I have evidence that female role models are important; if I organise a conference and on the list of speakers there are women, then I get more applications from women to give talks rather than posters. And I think if people see a woman doing a particular thing they do think: oh, I could do that; it's not a conscious thing, it's a subconscious thing."

(Female 15)

> "I've always wanted to do science, since I was little and having women around doing it or not I don't think it affected my decision that much. But that's personal [...] I think that a female role model will be important especially [for] younger ages, PhD students; or maybe even later, seeing a female professor, it does make a difference."

(Female 17)

> "I think role models are really quite important in terms of giving people something to aspire to. I suspect that actually what would really help is having the right people and potentially female role models at a lower level, so whilst you can see the high profile people who've made it to the top of their professions, what you really want to be able to see is people who are happily going about their careers and who are just two steps ahead of you, [which] can give you encouragement."

(Male 20)

The majority of the interviewees, either with children or not, believed that parenthood affects more women's career than men's (Question 14). Many of them shared their personal stories, some of which were described as more successful because of shared parental responsibilities or flexible and parent-friendly workplace policies. Most of the interviewees admitted that working fewer hours or part-time after coming back from parental leave is certainly a career obstacle and leads to slower career progression. Lack of mobility and difficulties in renewing or finding new contracts or part-time jobs, which could help new parents, were also mentioned as results of parenthood that could affect one's career.

> "My thoughts are not just for HPC, but in general for scientific and technical jobs; in my opinion there are not family friendly working conditions, so women get turned off. For example with technical jobs, it's very difficult to find part time jobs, which would help, coming back from maternity leave[...] Men can start a family much later, so they can wait until they reach a certain point in their career when they think it's a good time now, women cannot wait that long."

(Female 17)

> "Certainly was a factor in choosing to go for IT services rather than academic track, because I just don't have the energy to put in the hours that are required for doing that; I've got major respect for people who do that. If I really wanted to go for an academic career, I might not have had a second child [...] It has limited me in terms of being able to go for a promotion; I could have done that a year before I did."

(Female 19)

> "Once I headed back to work, I work three days a week and honestly that has an impact on my career. I do few hours, consequently less work. And there are fewer opportunities, like maybe meetings I'd like to go to, which are on a day that I'm not working. Surely I do manage to go to them, if I could arrange childcare [...] it has impacted my husband's work in the sense that he doesn't travel so much in order to spend more time with our child, he has flexible working arrangements. But in terms of career, I don't know how much impact that would have in his career, I think less."

(Female 22)

> "I didn't come back till after six months [of maternity leave], but I came back full time; and not long after I got back, I got promoted. With my husband, we were equal parents [...] But I think if you choose to go part-time, it's more difficult to be recognised."

(Female 29)

Few of the male interviewees expressed no particular career progression concern, because their wives were willingly stay-at-home mothers. However, they could understand how motherhood can have a bigger impact on a woman's career. On the other hand, a few women and men described programming and computing jobs as ideal for parenthood, because a lot of work can be done remotely or because it is easy to catch up with any programming updates and new projects, when you already have fundamental knowledge.

> "Parenthood constrains my hours of work. I probably used to work longer hours, whereas now I have to go to get the kids from school or drop them off at school or take them to football training. It's probably better for me actually; I probably spend less time working. Although ironically, I don't think that I do less work. I suspect I do more work; I just work more efficiently."

(Male 24)

> "It didn't affect my career very much, except the ability to work flexibly from home for instance [...] it's certainly affected my wife's career. I take full advantage of the ability to do it flexibly and I was able to stop having to travel [...] My career within a university made family life easier. Now, whether if we didn't have children, would my wife want to work full time? It's hard to imagine."

(Male 37)

> "As a person using HPC, I think the beauty of it is that we're able to work remotely. Obviously my teaching might suffer. But at least I can still work from home because I'm just accessing the computer or going on the server, so I don't see it as being too much of a disruption.

> But I can see it probably leading to delayed response in terms of my research publications."

(Female 6)

4.2.1.2.4 How to Achieve Equality and Diversity in the HPC Community

Fewer than half of all the interviewees (more women than men) stated that they had had a mentor currently or at some point in their career; whereas almost half of all the interviewees had never had a mentor (Question 15). Those who had a mentoring experience described it as a very positive and useful experience. Most of them needed a mentor for getting advice on career development matters. The majority of those who did not have a mentoring experience also found mentoring a good idea.

> "It (having a mentor) was valuable [...] a mentor is someone who's not just there to give you support, but someone that you can always feel comfortable, just even bouncing ideas off, even if they are negative ones, you don't want a mentor to always be supportive."

(Female 6)

> "I liked the fact that my mentor was not in the same organization. So that they were able to give a bit of objectivity, and also I could place a little bit more trust in them because they didn't have a conflict of interest. So, that was really useful. I think particularly I appreciated their ability to talk through similar situations that they had seen happen and give advice to me on that."

(Male 20)

A few of those (more women) who had never had a mentor mentioned that they had expressed interest in having a mentor at their workplace, but they hadn't got any response. They explained that the mentoring schemes of their workplaces (which included academia and industry) did not operate properly or were not publicised enough for people to join and become mentors or mentees. However, the majority of those who had a mentoring experience, had an "informal" mentor, not a mentor throught a formal mentoring scheme of their workplace. The "informal" mentor was in some cases a former PhD supervisor or a former senior colleague. Additionally, all the interviewees believe that a formal mentoring scheme which makes you aware and gives you the option and oportunity to have a mentor is a very good idea and it is needed in all workplaces; but a formal mentoring scheme which is mandatory and puts pressure on people to become mentors or have mentors is a waste of time and a *" box ticking"* task.

> "I think it's a good thing for both sides not only for the person being mentored but for the mentor as well. I don't think it should be forced to do that, then it's pointless. If you're forced and you're really just doing it because you need to tick a box somewhere, that defeats the purpose of the whole thing. And I think you should be able to choose the person. But I definitely think that everyone should know it's an option."

(Female 11)

> "I think mentoring should be encouraged and the potential advantages should be well publicized. I don't think the mentoring network in the universities are well publicised [...] I don't think it should be compulsory, but it would certainly be better if it was given more importance.

(Male 24)

> "I would have liked to have had one but it has never happened. A while ago I expressed some interest and nothing came out of it, but I don't even remember the details anymore, I'm afraid."

(Female 25)

> "Recently I've become aware of the the way that mentoring is used seriously in non-academic organizations. They have good programs and it's expected to happen, that's far from being the case in academia; I could probably dig out a document somewhere that says

> that every new member of staff to be assigned a mentor, but does it happen? It happens more from personal effort, mostly of the mentee, and not because it's part of the DNA of the organisation."

(Male 30)

Finally, some of the interviewees had been mentors to others, mainly in an informal capacity, and a few interviewees expressed interest in becoming mentors in the future. The main concerns of most of them were: the time that one needs to dedicate as a mentor, the unawareness of the way to become a formal mentor, and the experience and skills needed to become a mentor.

> "I was quite fortunate, because I asked for time from [my workplace] to do mentoring which isn't long, about two hours a month. It wasn't a very big time commitment, but you do need permission to do that; unless you're going to do it in your spare time"

(Male 24)

> "I think if you want a mentor you can go and find one. If you want to mentor somebody or offering yourself as a mentor I don't know how you do that, because there's no mentoring matchmaking service for academics online, and I have no idea if maybe the university runs a mentoring service."

(Male 35)

All the interviewees were aware of the term "Unconscious Bias" (UB), and they were able to describe what it means (Question 16). The majority of them also had heard and knew the meaning of "Impostor Syndrome" (IS). Fewer people had heard the term "Stereotype Threat" (ST), and most of them could not describe what it means. Interestingly, this showed that probably those people had not filled in the HPC community survey we conducted the year before, where we included two questions related to ST and its definition (3.4.2.5). "In-group - Out-group Bias" (IGOGB) was the least heard term with only nine intreviewees having heard of it. However, most of the interviewees could understand what it means.

After giving the definition of all the terms[3], the majority of all interviewees, but more women than men, identified themselves having suffered from IS. Many of them explained that they have found ways of dealing with it, such as mentoring. Also, most of them admitted that most likely they had UB, since they believed that all people have some kind of implicit bias due to socio-cultural effects and stereotypes. Very few recognised themselves with cases of ST and IGOGB.

> "I remember seeing women staff around assuming that they are clerical staff; until I found out that there were highly respected academics in my field. But I never made the same assumption for men, actually it was the opposite."

(Male 5)

> "I do feel like I don't know as much as people think I know [...] I think that there are two key things for getting over it: a formal process, and support from a supervisor or a mentor. A formal process is like going for a promotion or the annual reviews, where you have to actually look back at what you did over the last year and numerate it, and think: *yes, I actually did accomplish quite a lot and I do deserve to be here.*"

(Female 19)

> "I'm aware that I have identifiable biases. I identify with Impostor Syndrome in some aspects of what I do, but actually not in a lot of others. I used to suffer from it an awful lot, but I've learned ways of coping with it."

(Male 20)

[3]Unconscious Bias: prejudice in favour of or against a thing, a person, or a group compared with another, that we are anaware of; Impostor Syndrome: inability to believe that one's success is deserved, and fear of being exposed as a "fraud"; Stereotype Threat: belief that one's performance in a particular situation may confirm other people's negative stereotypes about them; In-group - Out-group Bias: a pattern of favouring members of one's in-group over out-group members.

> "The in-group - out-group [bias] thing, I believe, is a big issue in a community like HPC, where there is a fairly small group of people and you don't get exposed to a lot of new people. You see that people don't want to work with other people, and often projects can suffer as a result of that. "

(Male 24)

> "I don't have the technical skills as many of my peers, but I have other skills. I think there is still a bit of a mentality in the HPC community that technical skills trump everything; that used to bother me a great deal, and I used to carry a certain amount of Impostor Syndrome with me as a result of that. One of the things that was helpful for me was mentoring."

(Female 31)

The vast majority of the interviewees had an Equality and Diversity (E&D) training (Question 17). For most of them, the training was an online one which they were obliged to do because of their involvement either with teaching or with interviewing and recruiting. Most of the interviewees described that type of training as not particularly helpful and well-made, as a *"box ticking"* activity, and they explained that the content was mostly *"common sense"*, without interesting real-life examples and case studies.

> "I think what's really important is when there is training but it's of high standard [...] the people who have done a good E&D course, they say that it's eye-opening; but I think if you do a bad one, that's worse than not having done one at all."

(Female 15)

> "There is some online training but [...] it's sort of common sense, and it's the sort of training that was done because they have to say they provide it. So, it's box ticking rather than genuine learning; it doesn't really stand out in memory."

(Female 22)

Even though most of them believed that an actual training would be better, they were concerned that they would not have had the time to attend it. However, almost all the interviewees agreed that such a training is very important and it should be available, and maybe even compulsory, for all the members of staff, not only those with specific responsibilities; because it raises awareness of biases and stereotypes, and provides ways of dealing with various incidents and people from diverse backgrounds. A few interviewees mentioned that they had been or heard of an UB training, which was also very helpful in identifying implicit biases and ways of dealing with them.

> "Within HPC, where there are small groups, less than 10 [people], we tend to stereotype people; HPC is more natural to fall into mistakes. I think we need more awareness, a more tailored for HPC specific E&D training."

(Male 14)

> "I think it's good to have these things as being compulsory [...] so, if there's an email comes saying: here's your online E&D training, please go and do this, I'm sure there are quite a few who would say that's just another email from HR and delete it."

(Male 23)

> "The online one was not quick, it takes a couple of hours to complete. So to be fair, you might as well spend half day in a real room with people. I guess it's a cost issue probably [...] I think that would work somewhat better than an online training [...]for the more human-aspect traits and behaviors, which you don't really get from short videos and multiple-choice questions."

(Male 24)

> "It would be better if it (the E&D training) was mandatory for everybody. Because what's the difference between interacting with

> a student and interacting with a post-graduate at a conference or a member of the public when you do your everyday job? We all have to interact with people to a degree of formality."

(Male 36)

At most of the interviewees' workplaces, there were some social activities (Question 18). However, most of them were not formally organised, but based on peoples' initiative, such as someone arranging drinks after work. The most common formally organised social events were the Christmas dinners or some weekend activities. Most of the interviewees with children admitted that after having had children, they did not attend after work events, except the formal dinners.

> "There are informal Friday afternoon pints, if someone sends an email; but that's not as popular anymore because most people work from home on a Friday. There always has to be someone who takes the responsibility to organise something. If I am free and I don't have anything to do the next day, I join."

(Female 34)

> "I think there was that initiative couple of years ago when they had coffee in the mornings and cake, that was paid by the management. So, people were meeting at a common room and talk about what they're doing [...] it was a good initiative, but we didn't carry it out"

(Female 32)

Also, most of the interviewees (more men than women) believed that socialising with other members of stuff is important and it's a good way of identifying people with common interests or work ideas, and getting to talk with people from other departments and more senior positions; which might lead to reduction or elimination of biases and stereotypes within the community. A few interviewees (mostly women) found social activities, such as going out for drinks, not important and not something that they like doing with people from work.

> "It's very important to interact with other people of different levels and talk about science, and know that professors are actually humans."

(Female 8)

> "I prefer working in a workplace that's more sociable [...] You might be stressed or fed up or upset when you're at work and if you dont have anyone that you could tell you're upset, then you're kind of stuck."

(Male 24)

> "I think [it's important] having a place in the work to encourage people to socialise, like a coffee-room [...] you might end up talking about work, but even if you don't, you're still getting to know people across your whole department who you may not talk with otherwise. I think that's better in general at work, because everybody feels in the group."

(Male 36)

> "I must say that since I had my children, I started to opt out of those kind of activities, because they are ususally in the evening and this is the moment when you have time for them (the children)."

(Female 25)

> "I think it's nice (having social activities at work), but it's not crucial in the sense that if you don't have it, it prevents you from doing your job. I'll probably not go to things (social activities) [that] I'm not interested in just because other people are going and I feel like I should socialise."

(Female 11)

Finally, the main suggestions of the interviewees on how to achieve better E&D in the HPC community (Question 19) summarised the discussions and included:

earlier exposure to programming and HPC facilities in all degrees and disciplines (not only STEM), programming courses in schools, awareness and better training on E&D matters and relevant demographics (as WHPC network does for women of the community), increase of outreach and role models visibility, more regular training courses, use of less technical language in training and job advertisements, and change of the stereotypical perception of STEM and HPC as *"geeky"* or too hard and not related to real life.

> "Changing culture seems to me impossible and it will take hundreds of years, but that's what is required. I think role models are good [...] and things like increasing the number of speakers. I mean you don't have to go up to something that's not realistic, but even maybe just having 20% women speakers, and of different cultures, at a conference."

(Female 2)

> "I would like to see published data of how many women and how many men applied for grants or submitted papers, and the percentages of them that were successful, so people [can] realise [that] there is a problem."

(Female 18)

> "I really do think that HPC needs to think about hiring more people from outside its comfort area even if those people need to be trained up more [...] more people coming in from the social sciences and not just the traditional ones; to give us a bit of objective viewpoint and a new set of skills."

(Male 20)

> "I think exposure is important. Perhaps if I didn't have friends who were into geeky computer stuff, maybe I would never have found something that I'm actually quite good at and really enjoy."

(Female 21)

> "In the short term, it's about convincing the senior people that this is what we need, which is quite difficult [...] In the longer term, I think it's more about educating both young girls and boys why diversity is important and encouraging young girls to not be put off."

(Female 22)

> "I think achieving even a 25% representation of women in my group was difficult, given the field of the applicants we got [...] if you make a really ambitious sounding job advert, women might be less likely to put themselves forward for a more ambitious job advert, I toned down the language on that."

(Male 28)

> "All the sort of outreach, and making it so that people can understand what HPC is, how interesting it is, what rewarding career it is. Then I think it's important that those people who fund or employ people, are aware of E&D issues."

(Female 29)

> "Pretty much any job that you might want to think about in the research technology context involves people, collaboration, project planning, organization, creativity; so, perhaps emphasizing that these skills are as important as the hard technical skills, which we associate to have historically seen as things that men are more dominated, might help."

(Female 31)

> "We need to link science with everyday life, so people can relate to. Basically, everything we use is based on science and engineering. The mobile phones people use so much nowadays, science has built them for you! That's what software engineers do, if someone's ever wondered."

(Male 36)

4.2.2 Interviews by Email

Even though this study primarily aimed to conduct only in-person and video call interviews, 5 of the 42 participants were interviewed by written questions and responses sent by email. The main reasons for using this technique were either participant's preference (for personal reasons) or inability to find a suitable place or time for the interview to take place.

This approach differs from the in-person or video call interviews, because it has the form of a long questionnaire and the participants' responses are not spontaneous and timed, but more well thought through and investigated. This was taken into consideration for the analysis of the responses of those who participated in the study this way, and these data and their results were kept separately from the in-person and video call interviews data.

4.2.2.1 Data Collection

For this approach, we edited and sent the consent in the form of a document to the interviewee to sign it prior to the interview (available in Appendix C.1). As soon as the consent was signed and sent back by the participant, a document with the questions was sent to them to be completed in an agreed period of time.

4.2.2.2 Analysis and Results

For the analysis of the interviews by email, we followed the same thematic approach as for the recorded interviews (4.2.1.2). The identity (ID) of each interviewee was formed here by using the interviewee's gender and a capital letter instead of a number.

4.2.2.2.1 Background and Career Route of the Interviewees Table 4.4 displays the identity and background characteristics of the 5 interviewees via email, which follows the same pattern we used for Tables 4.1, 4.2, 4.3 at Section 4.2.1.2.

Table 4.4 *Identity and background of interviewees by email .*

Gender/ID	Current Position	Previous Position	Discipline	Career Stage	Degree
Female A	User support/Analyst (Academia)	User support/ Web admin (Industry)	STEM	Early	MSc
Female B	Applications Consultant (Academia)	Programmer/ Developer (Academia)	STEM	Mid	MSc
Female C	PhD Candidate	Student	STEM	Early	MSc
Female D	Applications Developer (Academia)	Computational Scientist (Industry)	STEM	Early	PhD
Male E	Chief Technology Officer (Industry)	Technical Director (Industry)	STEM	Mid	PhD

All the interviewees were planning to progress in their current roles or they had applied for more senior positions. The main route which all the interviewees followed to join the HPC community was some kind of involvement with computers and computational research during their degree or interest to use HPC for a certain project. They also chose HPC because of the variety of research it offers or for the technical challenge of it.

> "I started working with computers in research at the end of my [STEM] degree, went on to do a PhD in computational [STEM] and then [...] I didn't directly return to research after the lure of working at the cutting edge of computing performance."

(Male E)

Training was essential for all the interviewees, and it was offered to them at an introductory level, as well as later to keep themselves up-to-date. If the interviewees could not work at an HPC environment, they would like to do a job related to computers and programming, or/and they would like to have similar roles in science or technology related workplaces.

> "I had to attend introductory [courses] on HPC and [other] similar [courses], once I had started. In principle, I can attend any courses I wish to attend."

(Female D)

> "[If I couldn't work in HPC], I would certainly not abandon my computer, even if that would mean that I would stick with writing sequential codes. However, I am very interested in simulation of the real world and I like to contribute to a real-life problem in a scientific way by either programming or developing numerical methods for simulation."

(Female C)

4.2.2.2.2 Challenges in the HPC Community The steep learning curve was one of the main obstacles that most of the interviewees met, when they first entered the HPC community. However, some of them described it as *"challenging, but fun"*. Other challenges mentioned were: short-term contracts and funding. Most of the interviewees were not facing any significant obstacles at the moment of the interview. Some of the obstacles mentioned were: working on too many parallel projects, isolation when not working within a team, and policy-related and funding issues.

> "[When I first entered HPC], there was a steep learning curve to run code on the supercomputer at all. I had only limited experience of programming in a [specific] language, and had a lot of trouble configuring how to run [specific] languages together."

(Female B)

Some of the interviewees explained that due to being unaware of the HPC community as a whole or having joined the community recently, they were not able to suggest any changes in the community. Those who expressed opinion on this matter, suggested lowering the barriers of participation in the HPC community and more diversity of those working in the HPC community to allow better variety of views.

> "[What I would like to change in the HPC community] could be more integration between systems, people and software developers. There should also be no barriers to participation in HPC."

(Female D)

To promote HPC, our interviewees have been trying to make researchers - mainly from non-STEM background - aware of the uses of HPC; they have been participating and encouraging outreach activities and engagement of girls and young women in programming and HPC research. All the interviewees agreed that HPC is mainly used for STEM research and at the moment the use of HPC by non-STEM research was limited. Additionally, they all agreed that it is important to promote the use of HPC and its benefits to *"non-traditional"* HPC fields.

> "[To promote HPC] I answer questions from people wondering if our facilities are suitable for their work and let them know how to use them. We occasionally have a stall at open days and have started having fortnightly drop-in sessions where people can ask us things in person."

(Female A)

> "In my opinion [HPC] is still way too much unknown (and not yet accepted) by humanities and social scientists, and I am totally convinced that these fields could benefit a lot from simulations and computing in general."

(Female C)

> "HPC is available to many non-STEM areas and it's a growing area. Knowledge of what HPC can do in those fields is more limited and requires a different approach. However, there are still some lowering of barriers needed."

(Male E)

4.2.2.2.3 Underrepresentation of Women in the HPC Community and Potential Reasons All the interviewees believe that there is an underrepresentation of women in the HPC community, and the main suggested reason was the underrepresentation of women in STEM subjects, since that is the main way for people joining the HPC community. Female B mentioned that the number of women at her workplace was good, because of the place being *"open to hiring people from a variety of backgrounds"*.

> "Yes [I believe there is an under-representation of women in HPC, and] I think it follows on from the low numbers of women choosing Computing Science degrees."

(Female B)

Most of the interviewees did not have experience of working in a gender balanced team and they could not express an opinion on how this could affect productivity. However, some of them believed that there might be some negative impact of a dominated group in a team. A few of them were aware of research that had found higher productivity as a result of gender balanced teams. Others mentioned that there should not be any relation of gender with productivity or that there was no need for absolute balance in a team.

> "In my own experience, I have seen that "negative" team characteristics can emerge when teams are dominated by one group, and this may produce a less productive team (e.g. some people feel more inhibited to share ideas)."

(Female D)

Female role models were considered very important by all the interviewees, since lack of visibility can lead to normalising the absence and lack of ability of women in certain fields. Additionally, promoting women and meeting succesful women and female leaders could be encouraging and increase the sense of fitting in.

> "[Female role models and female leaders are] quite important. I feel more of a sense of belonging given that I'm working with other women. Also, it's been encouraging to see women I've worked with on the same pay grade being promoted."

(Female B)

Parenthood was regarded as having an impact mostly on women's career, since usually the point in their life that women are supposed to become mothers is critical for the progression of their career. The case of most workplace policies not including shared parental leave was also mentioned as having an impact on women's career. Also, working part-time and returning after parental leave were mentioned as factors that could affect negatively one's career.

> "I know women who have sacrificed progression in their careers to raise children. Part-time parents also have to work harder to prove they are as productive as their full-time colleagues."

(Female D)

> "Many organisations are adopting the shared parental leave model. However, more can be done in creating opportunities for women returning to the workplace after having families (flexible working, childcare etc)."

(Male E)

4.2.2.2.4 How to Achieve Equality and Diversity in the HPC Community

Some of the interviewees did not have experience of mentoring or had a wrong impression of what mentoring was. Those who had some mentoring experience believed that it was very helpful for their career and that it should be promoted and established as a formal practice in organisations.

> "[Mentoring] hasn't been something I've been involved in on either side. We don't do academic research, so [we] don't have PhD students coming through working with us. I think [it] is important in making people more comfortable in a new place and showing what jobs are possible, but don't have any experience with it."

(Female A)

> "I think it is important that organisations facilitate mentoring. This may require simply formalising what already goes on in the

> organisation, but making sure that everyone has access to the benefits of mentoring, not just those who know the *right* people. The idea of technical mentors should be promoted [as well]."

(Female D)

All the interviewees were aware of the terms (UB, IS, ST, IGOGB) they were asked, pointing out that UB and IS were the most popular and the ones that most commonly affected them and others. Most of the interviewees had attended an on-line E&D training, and there were no other comments on it.

> "Impostor Syndrome is common among people prone to self-analysis, so is high among geeks and academics in general and then on top of that higher among women. Everyone has biases. I have occasional impostor syndrome - when entirely new things come along - but not very much; I've been in my job long enough to know what I am doing and know I do it well!"

(Female A)

Those who were asked about social activities in the workplace, explained that there were some social activities, but they were not usually part of them. However, they mentioned that social activities could help people to feel part of the group and *"eliminate biases and stereotypes"*.

> "There are [social activities in my workplace], and I typically don't do them a lot. However, many staff [members] do them, which is good. However, there needs to be care if it's focussed on sports that are typically one gender (football etc)."

(Male E)

Finally, there were suggestions of 4 of the 5 interviewees on how E&D could be achieved in the HPC community (Female A did not provide a response to this question). The main points from those suggestions included: making HPC accessible to more people from different backgrounds, educating girls from early school years, keeping addressing the gender imbalance issue, and focusing on E&D trainining.

> "[To achieve equality and diversity in HPC], make an effort to recruit from non-traditional (e.g. pure Computing Science) backgrounds. Try not to make recruitment too much of a checklist of technologies, which may put someone capable off applying."

(Female B)

> "I think it is very important that especially young girls are not directed by society into the classical girl-themes. I don't know if there is more that can be done, especially [...] [without] becoming unfair to the other people."

(Female C)

> "[To achieve equality and diversity in HPC, everyone should] be educated on equality and diversity, not just through learning in a course; but through practical, real-life cases and scenarios, where people are challenged to face their biases. For stereotypes to be challenged through integration activities and having diverse leaders and role models. But especially by engaging the young and removing barriers to participation."

(Female D)

> "I think [the] Women in HPC [network] really helps the field [to achieve equality and diversity] but needs involvement from all genders to increase the impact. As a society we need to address early education issues around gender as a priority. I think it is very important that the HPC community continues to focus on gender and diversity in general [...] This is very much culture change and we all have a role to play."

(Male E)

4.2.3 Summary and Discussion

In this section, we gathered and analysed data by interviewing female and male inidviduals from the three career stages and from various backgrounds, positions and work environments of the HPC community. Our aim was to collect evidence from within the community, that supports our previous hypotheses and results on the underrepresentation of women in HPC community, the main reasons and likely ways of improvement.

The main findings from the analysis of the interview material were:

- Of 42 interviewees, all but three had a STEM background;
- Most interviewees were self-taught in programming and using HPC facilities;
- It was often mentioned that there is lack of formal training provided, which makes it hard to use HPC facilities and join the community;
- Most interviewees stated that they would not be able to do their job without HPC, and if they could not use HPC, they would still prefer to do computational research rather than theoretical or lab-based research;
- More women than men reported difficulties in getting funding or having their grant applications accepted in order to use HPC facilities for their research;
- The main challenge reported for joining the HPC community was the learning curve and the lack of formal training, which are results of the lack of earlier exposure to programming;
- The main current community challenge reported was that HPC's development is not science- or user-focused, which makes the community unattractive and inaccessible;
- Other serious challenges mentioned included lack of diversity, lack of women in senior positions, and lack of clear career paths;
- HPC was described as a closed community and not adequately promoted to researchers, especially to non-STEM disciplines (where there is a better gender balance). It was suggested that programming needs to be integrated into all disciplines and degrees for easier access to HPC;

- All interviewees agreed that there is an underrepresentation of women in the HPC community, with most popular explanation the underrepresentation of women in STEM subjects. The most common reason suggested for the underrepresentation of women in STEM was gender- and STEM-related stereotypes;

- Even though most interviewees did not have experience in working in gender-balanced teams, most of them believed that gender-balanced teams would have better performance;

- Parenthood was considered as a career obstacle, mostly for women, and when there is no shared parental leve policy at the workplace. However, this should not be an issue for HPC and programming related jobs, since work can be done from home of part-time/flexibly without consequences on the career progression;

- Mentoring was described as very important and helpful, and it should be an option for everybody, but not as a forced *"box-ticking"* activity;

- Most interviewees were not happy with the E&D training they had, and they suggested that it should be better made and available or compulsory for everyone;

- Socialising with people from other backgrounds and positions in the workplace was considered important by most of the interviewees for reducing stereotypes and biases;

- The main suggestions for imporvement of the community were: programming to schools and all disciplines, outreach, female role models, change of perception of STEM and HPC as hard and not related to life.

Taken together, these findings suggest that the reasons for women's underrepresentation in the HPC community are similar to those that cause women's underrepresentation in STEM subjects, which mainly feed in the HPC community. However, the most serious HPC-specific challenge is that the HPC community and its development seem to be focused more on the power and the size of the machines rather than the science they are used for and the users' needs. This image makes the community fairly unattractive to some, and mostly to women. Additionally, the lack of formal training and earlier exposure to HPC are major barriers for entering the HPC community, especially for people coming from non-computational backgrounds. The best ways suggested by our interviewees, to tackle these barriers and increase the number of women in the HPC community, are earlier exposure to programming, outreach, E&D training, promotion of female role models, and change of STEM- and HPC-related stereotypes.

We acknowledge that the people who were interviewed for the purposes of this study, either targeted or recommended, had some interest in the study's topic. Consequently, their views might be biased and reflect the views of a group of people who have an interest in gender balance and its effects in the HPC community. Additionally, we recognise any potential biases that might have affected the analysis and interpretation of the interview material.

4.3 Focus Group Discussions

The difference between interviews and focus groups discussions is mainly the dynamics of the discussion. In focus groups, the collaboration and interaction of the participants may result in the production of data that no other method can do [153]. We decided to conduct focus group discussions, in addition to interviews, hoping to interpret previously obtained data and to stimulate new ideas or hypotheses for future work.

4.3.1 Data Collection

The first step of designing focus group discussions is the development of the questions to be discussed. Following the same pattern with the interview questions, as a basis of the focus group discussion topics, we used hypotheses based on our previous results and results from other similar studies. The main hypotheses were:

- The HPC community recognises that there is an underrepresentation of women
- There are not enough female leaders and female role models in HPC community, and that has a negative effect on the representation of women

- Parental care has a more significant negative impact on women's career than men
- Mentoring can help women to enter, stay and progress in the HPC community
- Gender-balanced teams produce better results and higher performance
- Providing managers and employees with awareness training can positively address the negative impacts of Unconscious Bias, Stereotype Threat, Impostor Syndrome, In-group - Out-group Bias

The next step was to identify potential participants or sample of our target group. Research shows that there are differences in the interaction of men and women associated with the gender composition of the group [153]. It is believed that heterogeneous groups are generally more effective than homogeneous; however, both types of groups can provide with different but complementary information. Other factors which can affect the dynamics of the groups and the results of the discussion are: age, career stage, job position, if they know each other or are strangers.

In our case, we were interested in the views of different people in the HPC community and not just specific group of people, for example only early career stage women. One of our goals was to examine how the gender factor works when the topic is gender related, and therefore we needed three types of groups: one all-female, one all-male and a mixed-gender one. We also needed men and women from all three career stages - early, mid and late- because this way we could gather information about the progress, as well as the different obstacles and ways of dealing with them, in all the career stages. Also, different career stages would have probably meant different age groups. Additionally, variety of disciplines and job positions would have been very useful for understanding the likely differences. It would have been good to have also participants of both genders who are parents, to gather information about existed policies and likely differences in the impact of children on their career. The ideal would have been to have people from different environments (academia, industry), job titles (professors, technical staff, students), and participants coming from different geographic areas or background, because this could have brought additional information on the topic, like cultural differences. For a successful focus group discussion with results of better significance and validity, the number of people for each group should be about 6-8, and the number of the groups can be adjusted according to how well the plan for the first group will work [154].

After having designed the questions and having identified the target group of the sample, we then needed to invite people to participate. The first round of invitations was sent in November 2016 to Edinburgh based HPC facilities users. Unfortunately, the response was limited. With six people having registered their interest, we set up a poll for them to pick the most suitable among likely dates of the discussion to take place using Doodle[4], an on-line tool for creating polls. This way we could screen the eligibility of the interested individuals and arrange the groups. However, finding a date suitable for everyone was difficult, and with some people having abandoned the effort, we did not have enough people to arrange focus group discussions. A second round of invitation was sent to the same group of people in the first half of January. Once again, we did not have enough participants for our discussion. After this second attempt, we quitted the effort and decided not to use this method for our research. The few people who had registered interest to participate were invited to be interviewed instead.

However, an opportunity arose when WHPC network ran a workshop for the HPC community in collaboration with the University of Leeds, in April 2017. With the help of the organiser from the university of Leeds, we invited by email the people who registered to attend the workshop to also register their interest in participating in our focus group discussion. Eight people (four women and four men) confirmed their attendance, and soon after they were asked to choose their preferred time. Our plan was to create two mixed groups: one with female majority (3 female, 1 male), and one with male majority (3 male, 1 female), to observe and record any differences in the behaviour and the dynamics of each group. When the preferred times were decided, and the room was booked, the two groups were set up, and the participants of each group received an email with all the information and details for the focus group discussion. They also received a reminder notice email, two days before the event.

The two focus group discussions took place on the 5th of April at the University of Leeds. Two people, one male and one female did not attend, so the groups were formed into one group with two men and one woman, and one group with two women and one man. Before the start of each focus group discussion, the participants had to read and sign the written consent provided by the moderator

[4] `https://doodle.com/` *Last accessed: October 2018*

(myself) of the discussion. The consent included information about the purpose of the study, the procedures, the benefits and risks, terms and conditions and data management. It included also contact details for participants' enquiries about the study. When all participants agreed to participate, the procedure started with introductions.

To facilitate the discussion, I created a presentation with the topics, supported by some data (percentages and quotes) adopted by our previous approaches and results. The discussion lasted not more than an hour and thirty mintutes, and the topics were ordered from more general to more specific to the research topic. All the groups were asked to discuss the same topics and to respond to the same questions, otherwise the comparison across groups would be more complicated. However, questions might have been revised, modified or rephrased according to the needs of each group. The questions were not survey-like and structured to provide potential responses or to be answered by "yes" or "no" [154]. At the end of each focus group, the participants received a small gift for their time and contribution to the study. All the documents related to the procedure of organising and conducting the focus group discussions are available in section C.2 of Appendix C.

The audio data of the two focus group discussions were transcribed using the same on-line tool ("VoiceBase") we used to generate the interviews' transcriptions and the transcriptions were then edited manually based on thematic analysis, similar to the one used for the interviews analysis.

4.3.2 Analysis, Results and Discussion

We acknowledge that the small number of participants made it difficult to examine if their gender had an impact on the discussions' dynamics. However, we noticed that the female participant in the first group was not intimidated by the two male participants and she spoke more than them. This is likely due to her personal interest in the discussion topic. We also observed that the male participant of the second group was quiet at the beginning of the discussion, but then he seemed more comfortable in contributing in the conversation. We also focused on other characteristics, such as career stage/age, which might have played a role in the discussion flow. Indeed, in both groups the two participants (one male and one female) of the late-career stage group (> 20 years of working experience) were more skeptical, they spoke less than the other participants, and they often made comparisons with the past. Whereas, the early- and mid-career stage groups participants in both groups, had more interaction with each other and talked more about the future. Those of the lattest groups were also more involved to outreach and public engagement activities than the late-career stage participants; which we suspected that it might be related to time availability and differences in responsibilities and priorities.

We also recognise that the small number of groups (2) and of their participants (3 in each group) reduced the statistical significance of this approach. However, in this section we present a summary of the main points raised during the two focus group discussions of this study, as supported material of our earlier findings and as reference for future work.

4.3.2.1 First Focus Group Discussion

The first group consisted of three participants: an early-career female (F), a late-career male (M1) and an early-career male (M2). They all had STEM background and they first started using HPC for their PhD research. None of them had experience in both gender-balanced and not gender-balanced teams, with F having worked only in gender-balanced teams because of the nature of her research; so, none of them had evidence that gender-balanced teams result in better performance and productivity. They were all sure that there is an underrepresentation of women in the HPC community, with both M1 anad M2 pointing out that women more often do pure science than computational related research.

F shared her experience in joining the HPC community, emphasising that she uses HPC as a tool, she was not formally trained and all the technically able people that had helped her with HPC were men. However, the first person who supported her with learning to use Linux was a female friend of hers. When M2 mentioned that the HPC users meetings are definitely male dominated, F added that she had never been to any of those meetings and that she found them terrifying. She specifically said:

> "It's quite a terrifying community, if you come from a non-computational background [...] I still have fundamental knowledge. I always feel that I ask silly questions, so going to [HPC] users meetings is terrifying."

Nevertheless, she mentioned that she now prefers computational than lab based research and she applies for projects with at least 50% computational work. In the discussion on likely reasons for women's underrepresentation in the HPC community, F suggested that it is a result of the women's underrepresentation in STEM subjects, which starts since the early school years. She particularly mentioned that even though her parents were supportive with her decision to study a STEM subject, overall the school was not supportive and they were trying to push girls into specific subjects. All the participants agreed that school teachers influence children more than parents.

Interestingly, F seemed more optimistic believing that the representation of women in STEM and HPC is improving, whereas both men believed that changing the early stage social conditioning and the stereotypes that have been reinforced for years will need decades, and it will not happen if we do not keep addressing the issue. We assumed that the reason that F might have been more optimistic was that she worked at a discipline with an overall better gender balance than the disciplines where the male participants came from; where as she mentioned later, they put a big effort in organising public engagement and outreach activities.

More public engagement and outreach activities were the main suggestions by all the participants for improving the gender balance and changing the stereotypes that define STEM as hard and detached from the real world. Other suggestions included introducing programming courses at schools and increasing female role models visibility. Finally, all three participants were not sure how useful HPC could be for non-STEM research, with main concern the even bigger gaps in programming and computational skills of the people coming from non-STEM backgrounds.

4.3.2.2 Second Focus Group Discussion

The second group consisted of three participants: a mid-career female (F1), a late-career female (F2) and a mid-career male (M). They all had STEM background with M doing currently non-STEM research with the use of HPC. M had no experience in working in both gender-balanced and not gender-balanced teams, F1 had worked only in gender-balanced teams due to the better gender balance of the specific discipline, and F2 had mostly the opposite experience. However, F2 mentioned that the few times she worked with women she enjoyed and valued it, but she was not sure if her performance was better. Both women thought that the personality of each person of the team is probably more important than their gender for the team to perform better.

In the discussion on women's underrepresentation in the HPC community, F1 suggested that it is easier to join a community where everybody uses HPC and there are people who can help, which was based in her own experience working at a discipline with better gender balance. She specifically said:

> "HPC is unusual: we use it as a tool without being experts. The people that do the core technical stuff are all male. The users probably have a better gender balance."

During the same discussion, F2 pointed out that the hard-core engineering and computational STEM subjects, which feed in the HPC community, are heavily male dominated; whereas M, who worked in a gender-balanced non-STEM research group, mentioned that due to his isolated enagement with HPC he had not experienced women's underrepresentation. The main reason for the underrepresentation that F1 suggested was the lack of understanding of what HPC is and its late introduction in postgraduate degrees. The lack of role models was also mentioned as a likely reason by both F1 and F2, which led to a discussion about how the already male dominated environment and the perception of HPC (and computing in general) as *"geeky"* could made HPC unattractive to women. Later, all the participants agreed that the image of HPC as something big and powerful, which is more appeling to men, must be replaced with an image that promotes the importance of HPC in solving problems and making research and life easier, which might be more appealing to women.

Better parental leave policies and an encouraging system for (parental leave) returners were also mentioned by M and F1, as an important part of the solution to the gender issue. However, they both believed that HPC should be an ideal environment for such cases, since part of work can be done remotely (from home) and the skills set for using HPC is very stable. Other suggestions for improving the gender balance in the HPC community included: addition of computing courses in schools, outreach activities which promote the broadth of careers with use of computers and HPC, promotion of HPC in non-STEM disciplines and introduction in programming and basic computing skills to all disciplines.

Conclusively, the views of the focus group participants are in agreement with the majority of the views of the interviewees and our previous findings.

Chapter 5

Conclusions and Future Work

One of the main focuses of this study was to quantify the gender demographics of the High Performance Computing (HPC) community, and accept or reject the hypothesis that there is an underrepresentation of women in the community. In order to obtain a first picture of the representation of women, we gathered historical data of lists of HPC-related conferences participants' names, which we gender-analysed based on their first name and by using a special software. To gather further evidence of the participation of women in the community, we analysed the attendees' responses to the feedback forms of HPC related training courses, where the respondents self-reported their gender. The analysis of these data sets composed the quantitative approaches to this study (Chapter 2).

However, numbers only tell one part of the story; to understand the current gender status of the HPC community, evidence showing likely reasons for the formation of those numbers was essential. In order to obtain such evidence, we designed and conducted a survey which reached more than 800 individuals from the global HPC community (Chapter 3). The survey aimed to accept or reject our hypotheses based on previous relevant studies discussed in Chapter 1, and related to what causes the underrepresentation of women in the community.

In order to support our survey results, we considered important to hear more experiences and views from within the community. For this reason, we interviewed women and men who shared their personal stories. They also shared their ideas on how to create a more gender-balanced HPC community, which is more diverse and inclusive, and encourages women to join and be an essential visible part of the community (Chapter 4). The data obtained from the HPC community survey and the interviews mostly formed the qualitative approaches to this study.

In this Chapter, we summarise conclusions of combining the most important findings of the quantitative and qualitative approaches in three main sections: the underrepresentation of women in the HPC community, the causes of the underrepresentation, and the practices to tackle it. Finally, we discuss suggestions for areas that could be further researched in the future.

5.1 Summary and Discussion

5.1.1 Are Women Underrepresented in the HPC Community?

The HPC community is hard to define since it has a very multi-disciplinary nature, spanning across various STEM, and a few non-STEM disciplines, always in combination with programming and computing. In order to identify if there is an underrepresentation of women, we thought that it was appropriate to attempt to define the HPC community through the participants to our study, by recognising certain charactersitics, such as background, work sector, career stage/age, qualifications, programming skills, current role, and relation to HPC.

Even though there is a current shift of the use of HPC by non-traditional fields and sectors, through our quantitative and qualitative research, we confirmed that the vast majority of the individuals who form the HPC community have a STEM background or work in a STEM environment and in academia. According to the survey responses, the most popular STEM disciplines picked by both women and men as their background or current working environment were *"Physical and Mathematical Sciences"* and *"Computer Science"*. Nevertheless, female respondents appeared to come from a non-STEM background twice as much as men. This is in line with the statistics showing that in non-STEM disciplines there is either a gender balance or a higher proportion of women compared to men (1.1.1). Similarly, all but three of our 42 interviewees had a STEM background. Also, the highest qualification of the majority of the survey respondents and interviewees, both female and male, was doctorate (PhD).

Most of the participants to training courses, who responded to the feedback forms, work in academia. This suggests that people from academia are more likely to receive programming training. Additionally, most of the respondents to our survey and the interviewees also work in academia, with no differences between women and men. From the analysis of the interviewees' additional reponses, we concluded that a career in academia is preferred because it offers better variety of projects and career progression, more flexibility and security, than in industry. We also hypothesised that partcipants to our study were more likely to come from academia, probably because they might have been more interested in such a study.

The majority of female survey respondents were 20-39 years old, whereas most of the male respondents were 30-49 years old. This could be either due to the fact that women do not progress as often or quickly as men in their field or because they quit their STEM related jobs after this age group; or due to the fact that younger women might have been more interested in the survey topic, which could be promising for the future of the women of the field. Also, interestingly more female than male respondents were of the age of 50-59 years, which might show the lack of interest of senior men in such studies. It is, however, crucial for senior men to be aware of the underrepresentation of women and the barriers they meet at all stages, as senior men tend to hold decision-making positions and could influence changing and tackling those barriers to create a more diverse and inclusive working environment for all.

A similar picture was obtained regarding the interviewees career stage, where most women belonged to early- to mid-career stages (0-20 years of working). Again, this might reflect the lack of women in more senior positions or the fact that younger women were interested in participating in our study. On the contrary, half of the male interviewees belonged to the late-career stage (more than 20 years of working), which made their responses really valuable for our study, since views on the underrepresentation of women in STEM-related subjects by senior men are not often heard.

Knowledge and use of Linux as Operating System (OS) is a significant factor for using HPC facilities, as most of them are Linux-based. Additionally, programming skills and the ability to develop software are also considered important for expert use of HPC; hence, attending training courses should be essential for those who wish to join and progress in the HPC community.

Our survey results indicated that women and men from the HPC community are almost equally familiar with Linux. This is not in agreement with the findings from our analysis of the Software Sustainability Insitute (SSI) survey (which we used as a guide for designing our survey, 3.2), where Linux was women's least preferred OS. However, it was an expected result from within the HPC community, since as we mentioned Linux is essential for the use of HPC.

According to ARCHER's description of the training courses (section 2.2.3.3), the participants were identified as *"users"* or *"developers"*, when they attended specific courses. We found that female course participants were identified more as *"users"*, and attended mainly introductory courses, which are mainly targeted at *"users"*. By contrast most men identified as *"developers"*, and they attended the various levels of courses at a similar regularity. This could be interpreted as implying that most women who use HPC facilities, do not develop software, and do not advance their programming skills and HPC knowledge as much as men do. We discuss this finding further in section 5.1.2. We also found that the majority of both women and men - especially from the STEM group - interacted with HPC as *"Users"*, according to their responses to our survey. Additionally, the survey results also showed that more men than women held a tenured academic position (in agreement with HESA statistics 1.1.1), and women were more likely to describe their role as non-academic or related to marketing and management.

Presumably, the fact that the majority of the HPC community belongs to the STEM group and works in academia supported our hypothesis that women must be underrepresented in the HPC community; since statistics show that they are underrepresented in most of the STEM subjects and academic positions (1.1.1). This also led us to examine if existing evidence-based reasons for women's underrepresentation in STEM subjects and practices could apply in the HPC community.

The quantitative approach to our study revealed a clear underrepresentation of women at conferences and courses of the HPC community. We found that women were fewer than men in all the examined categories of conference participation (paper and poster authors and presenters, invited speakers, workshop and tutorial presenters). We also found that women were outnumbered three to one by men at all levels and years of the courses we examined. However, as mentioned earlier in this thesis, the number of women participating at conferences and courses might not be representative of the community, as research has shown that women are fewer even at conferences of subjects with better gender balance [61–63]. Since we are not aware of an existing report of the gender ratio in the HPC community, our findings compose the first picture of women's representation in the community.

The aformentioned findings of female representation in the HPC community were also supported by the views of the participants obtained from our qualitative approaches. Most female and male respondents to the HPC community survey agreed that women are underrepresented in the HPC community and that gender balance issues sould be tackled. Additionally, most of the interviewees believed that women are underrepresented in the community, and many of them provided personal evidence to support their views. Finally, remarkably more female than male survey respondents believed that they belonged to a minority group of the HPC community, with most of them referring to a gender-related minority group.

5.1.2 Understanding the causes of women's underrepresentation in the HPC community

As mentioned earlier, since the HPC community consists of people mainly from STEM subjects, it is very likely that the reasons for which women are underrepresented in the HPC community are similar to those for women's underrepresentation in STEM subjects. Indeed, the most popular cause for the underrepresantation of women in the HPC community mentioned by women and men from the community was the underrerpresentation of women in STEM. We present here the conclusions of our findings in correlation and comparison with the reasons suggested by previous studies and presented in section 1.1.2, which also formed the basis of our research.

5.1.2.1 Biological Differences and Life-style Choices

Our results showed that more female than male partcipants either use HPC for their research or they have a less or not at all technical job, such as a job related to marketing, management or education/public engegement. This might have been a result of either personal choice or of institutional barriers and career route. Some of the female participants in our study who currently have a managerial position explained that they were not particularly happy with their job, and others who have more technical or research related positions admitted that they were happy with that. Similar views were shared by male participants; consequently, a general assumption that women have different preferences than men (1.1.2.1) was not obvious through our results.

However, indeed more women than men (mainly interviewees) work in a biomedical related working environment, which is in agreement with the studies showing that women are more interested in life- and people- related occupations than things-related ones (1.1.2.1), and with the higher proportions of women in health-related disciplines according to the statistics (1.1.1). We also noticed in Chapter 2 that the number of female course participants increased when we added the participants of the "special" course for Life scientists (2.2.3.5), which again shows that a large number of women in HPC might come from those disciplines.

Also, related to preferences and life choices is parenthood and its impact on one's career. We found that a higher proportion of both women with children and women without children believed that having children had or would have an impact in their professional life, compared to men. Additionally, more men than women were positive about moving their family to another part of the country or to another country for work purposes. Both these findings show that men are less affected than women by parenthood. Whereas for women parenthood seems to be still an issue. On top of that, our results show that interestingly the annual pattern of travelling to conferences of women with children and women without children seem to be very similar (1-3 times per year). It is not clear why this happens, however one explanation could be that apart from caring responsibilities, women might have other reasons for not travelling, such as lack of funding, discrimination in presenting conference papers and posters, avoidance of conferences due to small number of women and socially awkward behaviours. These were only some of our female participants' statements, and the very small percentages of female participants in the conferences we examined in Chapter 2 could confirm this travelling frequency.

Also, it is not clear why men with children appear to travel more often than men without children and women, which was also shown in our results. Some of our interviwees mentioned that they were much less affected by having children because their wives/partners were willing to be the stay-at-home parents. However, this still does not explain why they travel more often than men without children. This case undoubtely seeks further investigation, as it might provide evidence of favouritism towards men with children, as previous reserach has shown [31].

Additionally, our results provide evidence of how shared parental leave, flexible working hours and shared caring responsibilities between the two parents can lead to succesful career progression for women, and men in some cases. On the other hand, there is evidence that working part-time after maternal leave or having a break for a long period of time can have a negative impact on women's career progression. However, it was very often mentioned that programming and HPC-related jobs should be ideal for women with children and caring responsibilities, since such jobs can offer the flexibility to work remotely, and to be quickly updated with any software changes, when you already have fundamental knowledge. Presumably, motherhood should not be a barrier for women in the HPC community.

Oversimplified theories on reasons for the underrepresentation of women in STEM, such as "women are not interested in, or good enough at, STEM subjects", have been rejected by our results. Most of our female participants, not only have studied a STEM subject, but they have reached the highest degree of qualification (PhD), and this was entirely their choice. The majority of them also mentioned that since they started using HPC, they prefered computational research to theory- or lab-based research. Most of them added that if they were not able to use HPC, which they consider as the best tool for advanced scientific research, they would still prefer to do computational research on a smaller scale. This rejects the theory that women are not interested in or good at programming and computing (1.1.2.1, 1.1.2.2).

The problem seems to be that owing to lack of previous exposure to programming, Linux and HPC, more women found it harder to start using it, and probably advance in it. They also seemed to find advanced training courses more difficult compared to men, and their attendance frequency to training courses, especially of higher level, is very low. They also tend not to develop their own software, even when they receive training. Similarly, it was noticed that women do not attend very specialised courses and conferences, such as MPI-related ones, which might also be due to confidence issues, as a result of the gender differences in visibility and exposure, or due to Stereotype Threat (ST), as we discuss in the next section 5.1.2.2.

5.1.2.2 Stereotypes and Confidence

As we mentioned earlier, reasons for the low attendance of women to training courses might be lack of confidence, lack of previous exposure and programming knowledge, or Stereotype Threat. Many women in our study talked about how their non-computational background affected their relationship with the HPC facilities and community. Some of them still believe that they do not know enough in order to attend more advanced courses or users meetings, and that they do not receive the support they need in using HPC, either because they feel intimitated to ask questions, or because they often receive gender-based stereotypical behaviour from the male-dominated technical staff.

Also, we found that women's contribution at HPC-related conference was much lower than men's, which is in line with other studies on STEM subjects [62, 63]. Moreover, we found that the proportion of female poster presenters and authors was slightly higher than that of female paper presenters and authors; which is again in agreement with previous studies [61]. The existing research has shown that this is due to either gender discrimination in the selection of presenters and authors, or due to lack of women's cofidence to submit and present papers at male-dominated conferences.

Gender-based assumptions were often mentioned in relation to women's job or position at conferences. Female participants shared personal stories of having been assumed to be secretaries or other administrative staff by men, just because of gender stereotypes. Some women also expressed their complaints about being systematically given more administrative responsibilities than men, which was again due to sterotypical behaviour towards women and what women should do.

Socio-cultural stereotypes about what women should study and do as a job, as well, as stereotypes about STEM subjects being hard and not related to real life, that affect girls' and women's life decisions were mentioned as main reasons for the underrepresentation of women in STEM and HPC. Particularly for HPC, the words *"nerdy"* and *"geeky"* were used regularly by both women and men, to describe the stereotypical idea of HPC (and computing in general) that people have.

We examined if women and men from the HPC community were aware of Stereotype Threat (ST) and its effect. More women had heard of it, and significantly more women stated that at some point in their life they had felt confirming stereotypes believed for women. In combination with the previously mentioned findings, stereotypes and ST seem to be main reasons for women's lack of confidence, and consequently for their underrepresentation in STEM and HPC.

We also examined if our interviewees were aware of the Impostor Syndrome (IS), which, according to studies, seems to be a reason that holds both women and men (but more often women) from progressing in their career. Most of female and male interviewees knew what IS was and admitted having suffered from it at some point in their career. Many of them also explained that a good way of dealing with it was having a mentor.

Furthermore, our survey findings showed that women feel as successful as men, and most of them aspire to be in a more senior position. This shows that women are not less ambitious than men; however, they might have a different perception of success, since more men than women considered themselves successful because of a high salary, whereas very few women throughout our study referred to money, and if so, mostly as another case of gender discrimination (when women need to work harder for less money and promotion opportunities).

Finally, our study showed strong evidence of how women-focused courses and conference sessions can increase the number of female participants. We clearly demonstrated the difference in the numbers of women who attended and participated in courses and workshops/tutorials organised by the Women in HPC Network (WHPC). This suggests that women feel more confident when there are other women, as previous research has also shown [100].

5.1.2.3 Biases and Discrimination

Biases and more specifically "Unconscious Bias" were mentioned often by our participants, especially by men, as reasons for the underrepresentation of women in STEM subjects and consequently in the HPC community. From the analysis of their responses, we hypothesise that maybe biases that *"we all have"* and *"are difficult to deal with or get rid off"* are an easy explanation given by men, who are not affected by the underrepresentation of women and do not think or care about the problem's causes. "Unconscious Bias" was indeed the most popular term of the ones our interviewees were asked (4.2.1.2.4); and the majority of them, both women and men, admitted that they had some kind of bias due to effects of their sociocultural background.

In addition to that, we found that men from the HPC community considered more than women that there is no discrimintation in the workplace and that all colleagues are treated equally regardless of gender. Similarly, more men than women believed that there is no discrimination at conferences and that all attendees are treated equally regardless of gender. Also, a small portion of men talked on favouritism over women at workplace and conference, because of them being minority. This last finding is an example of men's lack of awareness of women's barriers in the community, as well as a likely case of misintepretation of affirmative actions aiming to increase the number of women in the field. We discuss this further in section 5.2.

As previously mentioned, biases and stereotypes are closely connected, and they all lead to various forms of discrimintation. We discussed earlier about how gender discrimination in the selection process of speakers and authors at conferences might be a reason for the small numbers of women at HPC-related conferences. Additionally, some of our female participants mentioned that they had been discrimintated in the selection process to receive funds or to be promoted; probably because the majority of the senior decision-making positions were held by men, who were more likely to choose someone who looked like them.

5.1.2.4 HPC-specific Causes of Gender Imbalance

As we discussed so far, some of the main reasons mentioned for the under-representation of women in the HPC community are the same that cause the underrepresentation of women in STEM subjects, since STEM subjects are the main source of people who form the HPC community. Those reasons were: gender-based biases and stereotypes, which lead to discrimination in various forms and lowering of women's confidence level, opportunities, and exposure to STEM and HPC subjects; and maternity leave and caring responsibilities.

However, the participants in our study mentioned some reasons which were more exclusively related to the HPC community. The fact that HPC community is a very closed "club" of people from certain backgrounds and inaccessible from people with non-computational background was one of the main explanation given by both female and male participants. HPC community is considered as a niche area of computing, which makes it seem even *"geekier"* and *"scarier"* than computing in general, to women and people who come from various STEM disciplines, and even worse for those from non-STEM disciplines.

Our results showed that women from a non-STEM background were by far less familiar with Linux than women from a STEM background. Additionally, we found that both women and men from non-STEM backgrounds were less likely than those from STEM to receive training and to develop software. These findings suggest that people, mostly women, from a non-STEM background were more likely to interact with the HPC community in other capacities rather than users, developers or technical staff. This could undoubtely have an impact on the underrepresentation of women in the HPC community, since most non-STEM disciplines demonstrate better gender balance.

Nevertheless, the majority of the female interviewees from biomedical backgrounds explained that their lack of previous exposure to Linux and programming was a huge barrier for learning and using HPC for their research. They added that the frequency and quality of available training, and the supporting documentation and staff were not enough to help them overcome that barrier. Most of those interviewees were self-taught or had received help from other users and colleagues. This shows that the lack of skills for computational research might not be an issue only at non-STEM disciplines, and it seeks further investigation.

Consequently, HPC community is difficult to be accessed by people from "non-traditional" disciplines. It was also often mentioned that the need or option to use HPC appears very late in one's career. For people with no previous programming experience, using HPC might seem daunting and unattractive, and often it might be rejected by researchers who have not used it before. Additionally, HPC was considered by our study's participants as a not well promoted research tool, which is only known and used by specific groups of people and specific projects within STEM. This is probably the reason why "In-group - Out-group Bias" might appear regularly in the HPC community, as some participants explained.

Despite the efforts of the community to promote historical female role models such as Ada Lovelace and Grace Hopper (section 1.2), the participants to our study were not able to name any historical or current HPC female role model. The lack of female role models, and especially of women in senior positions, in the HPC community was also often mentioned as a reason for the underrepresentation of women in the HPC community. If early-career women do not have examples of other women who have reached higher positions and developed within the community, they could lack motivation, confidence, and mentoring opportunities. Additionally, as mentioned earlier, it is very likely that there is gender-based discrimination in hiring, promoting and funding procedures, when the majority of the decision-making positions are held by men.

Finally, an important finding of our study, which triggers further investigation and action, is that people from within the community are worried that the HPC community is more focused on the power and speed of the machines and is more money-driven, than focused on the users and research needs. This is significant not only for the future of the HPC in general, but also in relation to our study, because this image of HPC might be unattractive to women. If HPC community promotes more the impotance of HPC in solving problems and making research and life easier, it might attract more women.

5.1.3 Ways to change the gender imbalance in the HPC community

The ways to change the gender imbalance in the HPC community should be similar to the ways suggested and, in some cases, applied to tackle the underrepresentation of women in STEM subjects. Before we discuss what our study's participants and results suggest in order to attract and retain more women in the HPC community, we present what the participants think about gender balance and why it is important. We believe that understanding why gender balance is important constitutes the basis for achieving it.

The majority of both female and male survey respondents believed that working in a gender-mixed team would have a positive effect on team's productivity. Similarly, the majority of both female and male respondents believed that gender balance would have a positive effect on the HPC community. Clearly, most women and men would like to work in teams and a community which have better gender balance, even though a few of them explained that gender should not matter to productivity and collaboration, and gender balance should not be forced. The intreviewees' views supported our survey findings, since most of them expressed a positive opinion on more diverse and gender-balanced teams, where there is a wider variety of ideas and experiences. However, the majority of them did not have experience of working in such teams and did not have any evidence to provide.

Since gender-based stereotypes and biases are considered to be main causes of the underrepresentation of women in STEM subjects and the HPC community, ways to eliminate stereotypes and reduce biases were the most popular suggested solutions of the gender imbalance by our study's participants. Addressing the issue and promoting women in STEM subjects through outreach and public engagement were suggested by many participants, focusing mainly on early school years and young students.

Mentoring was also a frequently suggested way to encourage young women towards studying and following a career in STEM subjects, and to support women who already study and work in such environments. According to our results, the majority of both women and men stated that mentoring had been a positive experience for them, and a very helpful one. However, we also found that more women who did not have a mentor, would like to have one. Particularly, some of them reported that even though they had expressed this interest, for some reason they never managed to be assigned to a mentor. These cases triggered a conversation on formal mentoring (i.e., a scheme organised by the workplace) and informal mentoring. With most of the participants having had an informal mentor, formal mentoring seemed like a *"box-ticking"* task which might not have the desirable effect.

Another practice considered to play an important role on reducing stereotypes and biases is the Equality and Diversity (E&D) training. Our study showed that the vast majority of the interviewees had an online E&D training, which they described as not helpful, not adequately informative, lacking of real life cases, a *"box-ticking"* exercise. Some interviewees suggested that owing to the variety of backgrounds of the people who form the HPC community, the community needs a more tailored E&D training, which presents examples and case studies from within the community. Additionally, some interviewees believed that if the training is well-made with appropriate content and not online, it should be compulsory to every member of staff and it should be offered during working hours.

Finally, a good and relatively easy way to reduce stereotypes and biases discussed with the participants of our study was social activities within the workplace environment, which are open to all members of staff and are not gender-based, such as going to a pub after work. Indeed some participants who were parents mentioned that socialising after working hours is not possible for people with children or caring responsibilities. For this reason, social activities within working hours and the working environment, such as organised lunch and coffee breaks, were suggested as ideal for members of staff from various departments and seniority levels in order to mingle and share views and ideas, and hopefully to reduce stereotypes and biases.

Some ways to change the gender balance more specifically within the HPC community were also suggested in our study. First of all and more importantly, changing the image of HPC as *"nerdy"* and *"geeky"*, and as very technical, difficult, and not related to real-life, could attract more people from non-computational backgrounds, who would like to use HPC for their research. These people might come from non-STEM disciplines, or STEM disciplines, with better gender balance, and this might result in attracting more women in the HPC community. Best media to achieve this image change are outreach and public engagement activities, which should be targeted to school children, university students and early-career researchers. In relation to that, it was also passionately suggested by our participants that programming courses should be added in schools, and in all disciplines and degree levels, and that HPC should somehow be integrated earlier to the curricula.

According to our results, the fact that HPC and programming related jobs could be ideal for mothers and people with caring responsibilities should also be promoted. Since many women in STEM, academia and research, worry about balancing family and career, being able to work remotely and flexibly should be appealing to most women.

Additionally, more regular and better quality training courses for all levels of programming and HPC skills are needed to cover the demand and the variety of users backgrounds. It was also mentioned that the language used in these courses should be less technical, and HPC jargon should be avoided. The same was suggested for job adverts, and generally for the broader community.

Finally, we discussed earlier the impact of the Women in HPC network on the numbers of female participants at courses and conferences. Our results also demonstrated a similar impact of diversity committees established by international HPC conferences (SC and ISC) to address the gender imbalance issue, to measure and provide evidence of the impact of their efforts to increase women's representation. It is very important for such inititatives to continue raising awareness, and promoting and supporting women, because in combination with all the aformentioned practices, they can lead to faster and more long-lasting effects.

5.2 Future Work

If the HPC community wants to reach its full potential, it needs to attract more women and more people from various and diverse backgrouds, who can offer a broader spectrum of views, experiences and research opportunities to improve humanity and life through their research. It is obvious through the results of our study, that HPC is affected by the barriers and the gender imbalance of the STEM subjects which mainly feed in the HPC community.

Several practices mentioned in section 5.1.3 have been applied and attempted to solve the gender imbalance issue in various STEM subjects. However, none has been particulalry effective so far, since the issue is still present and women still face the same barriers. Thus, in order to achieve better diversity and gender balance, the suggested practices need to be constantly and cooperatively applied.

Thoroughout our study, we noticed that some people from within the community made clear that gender balance should not be forced. In most cases, what they meant was that placing more women suddenly in a workplace, by favouring them or offering them advantages, in order to create "artificial" gender balance, will not solve the problem. In fact, it could create negative attitude and trigger further biases and discrimination rather than eliminating them. This is why we believe that numbers are just indicators of underlying reasons; and to change the numbers, we firstly need to identify the reasons and tackle the barriers they create. Therefore, policies whose target is only to increase the number of women without deeper sociocultural changes of the workplace, tend not to be effective or to be fully embraced by the communtiy.

It is also essential to be able to measure the impact of the various initiatives and practices by gathering feedback and quantitative data. For example, it is not enought to carry out E&D training sessions without receiving feedback which provides information on how useful and effective the training is. Similarly, informal mentoring might have been more popular among our study's participants, however we believe that formal mentoring offers the opportunity to detect the progress of the mentoring relationship, measure its impact, and identify areas of improvement. The same logic should apply to other practices, such as outreach and public engagement activities, social activities in workplace, published demographics, work of diversity committees, and so on.

Despite the recent efforts to lower the barriers of using HPC facilities by providing user interfaces and other means that do not require programming skills, our study showed that not knowing how to program is a disadvantage, as well as one of the main barriers, for researchers to use HPC. Consequently, another important area which requires further investigation is the examination of the level of computational research in various STEM and non-STEM disciplines, in comparison to other forms of research, such as theory and lab-based. This research could provide gender-related quantitative evidence, as well as qualitative evidence on the reasons why women and people from non-STEM backgrounds do not choose computational research. These reasons might include the lack of programming skills and of promotion of such type of research, and results of such an investigation could have great implications for the HPC community.

Finally, future study on this topic should hopefully receive more response and feedback from the industrial and commercial areas that HPC community has been recently moving into, such as Artificial Intelligence (AI) and Data Science, since these fields might offer different insight into the community's gender balance and women's experiences.

The main lesson to be learnt from this study is that the HPC community has been actively missing out talent and progress opportunities. Listening to the people who constitute the community is the best way to identify the community's challenges and the best practices to tackle them. This study set a basis and offered a guideline on where the HPC community should focus on in order to become more attractive, accessible and useful to everyone.

Appendix A

Quantitative Approaches and Statistical Methods

A.1 Statistical Analysis and Tests of Statistical Significance

For the analysis of the data in Chapter 2 and Chapter 3, we used Microsoft Excel and we mainly used "descriptive" statistics [132] indicating certain frequencies and patterns of our samples with tables and graphs. In the cases where we correlated or compared data, we used "inferential" statistics with tests of statistical significance, which can inform us if the sample findings reflect the population as a whole, and if our results are due to unexpected factors or random fluctuation [132], [133]. Tthe standard error of averages was marked with "SE" and $\pm number$.

When we compared responses to one question or two questions with gender, we used the Chi-square test of association to prove significance of the results or not by accepting or rejecting our null hypothesis. To run this test, there must be at least 20 observations in the sample with at least 5 in each category. We firstly calculated the expected values/counts by using this formula on the observed data: *(row total*column total)/grand total.*To find if our hypothesis was accepted or rejected, and if the results were significant or not, we calculated the p value running a Chi-square test with using the formula: $CHISQ.TEST(actualrange, expectedrange)$. If the result of the Chi-square test for these values was smaller than the critical value 0.05, that we had set for our analysis, it meant that our null hypothesis was rejected and the results were statistically significant. The critical value we set for our analysis means that there are 5 in 100 chances that our results being due to unexpected factors or random fluctuation.

In the case of 1-5 scale of preference/familiarity/agreement (Likert-type scales), we calculated the weighted average for each variable, and compared them regarding gender. To find if our null hypothesis in this case was accepted or rejected, and if the result of the comparison was significant or not, we used T-test. We also used T-test in the cases where we compared averages and small samples. The formula for T-test in Excel is: $T.TEST(array1, array2, tails, type)$, where array1 and array2 are the values we wanted to compare, tails shows if we were sure in which direction a significant result would run or not, and type shows if we know that the variances are equal or not. When the result of the T-test was bigger than the critical value that we had set for this analysis (0.05) meant that our results were not statistically significant, and the null hypothesis was accepted.

A.2 Conference Data Selection and Final Dataset

For the SC conferences we adopted the data from the history section of the exclusive website for the SC Conference series (www.supercomp.org/history.php). The final dataset was formed by 13,888 names, of which only 55 were unidentified.

For the ISC we found online (ref. www.isc-hpc.com/all-years.html) only the programmes of the years 2012-2017, named as"Agenda Planner". We collected our data from each years agenda "Speakers" section. We identified the gender of a total of 2,371 names.

The data for ICS were found at the individual websites for each year's conference. A total of 1,361 names were gender analysed, and only four of them were unidentified.

To retrieve data for CUG conference, we used the "Author Index" section of all the publicly available conference programmes (except for the years 2016 and 2017) and the proceedings. For the programmes of 2016 and 2017, we adopted the names from the detailed description of each session of the on-line available conference booklets. We gender-analysed 2,943 names with five of them being unidentified.

For PRACEdays' data, we used the PRACE official website (http://www.prace-ri.eu/pracedays/). Three of the four programmes (2015-2017) for PRACEdays included only the names of the presenters. The programme for 2015 was the same as the printed programme that we used in combination with the participant observation method in section 2.1.1. We identified the gender of 251 out of the 252 names of the PRACEdays' dataset.

For the PGAS conferences, we adopted the programmes for the exclusive website for each year's conference, and the proceedings from the ACM digital library website (https://dl.acm.org). A total of 761 names formed the final dataset for gender analysis.

For EuroMPI we adopted the programmes from the exclusive websites of each year's conference, and the proceedings from the website of "Springer" publications. From the final dataset for EuroMPI conferences, we analysed the gender of a total of 2,630 names, except of only four names.

The data of ParCo programmes were found at the exclusive website of each year's conference, and of the proceedings on the "IOS Press" website (https://www.iospress.nl/). Overall, 2,927 names were analysed with 12 of them marked as unidentified.

For EASC we found available the programmes for 2013 and 2015, and the proceedings for 2016 on the exclusive websites of each year's conference. Totally we determined the gender for 373 names.

A.3 Disciplines of Course Participants

The disciplines named by the respondents of the courses' feedback form and grouped into the STEM group inluded:

- Physics&Astronomy
- Computer Science
- Biology&Bioinformatics
- Chemistry

- Mathematics
- Geology&Oceanography
- Genetics
- Engineering
- Medicine
- Neuroscience

The disciplines named by the respondents of the courses' feedback form and grouped into the non-STEM group inluded:

- Linguistics
- Geography
- Transport Studies
- Political Studies
- Telecommunications

Appendix B

Semi-quantitative Approaches

B.1 Software Sustainability Institute Survey

B.1.1 Survey Questions

Here, we include the full list of questions (and the available response options) in the order they were asked in the SSI survey:

Question 1: *In which institution are you based?* Bristol/ Birmingham/ Cambridge/ Cardiff/ Edinburgh/ Glasgow/ Imperial/ Leeds/ Liverpool/ Newcastle/ Nottingham/ QUB/ Queen Mary/ Sheffield/ UCL/ Hand filled if not found;

Question 2: *Which of the following roles apply to you?* I conduct research/ I support researchers/ I manage researchers or research projects/ I develop software for researchers;

Question 3: *Which of the following disciplines best applies to your work?* (See section B.1.2)

Question 4: *Which funder currently provides the majority of your funding?* AHRC - Arts and Humanities Research Council/ BBSRC - Biotechnology and Biological Sciences Research Council/ Cancer Research UK/ EPSRC - Engineering and Physical Sciences Research Council/ ESRC - Economic and Social Research Council/ Leverhulme Trust/ MRC - Medical Research

Council/ NERC - Natural Environment Research Council/ STFC - Science and Technology Facilities Council/ University central funds/ Wellcome Trust/ Don't know/ Other;

Question 5: *How many years have you worked in research?* 1-5 years/ 5-10 years/ 11-15 years/ 15-20 years/ more than 20 years;

Question 6: *Do you use research software?* Yes/ No;

Question 7: *What would happen if you could no longer use research software?* It would make no significant difference to my work/ It would not be practical to conduct my work without software/ My work would require more effort, but it would still be possible;

Question 8: *Do you develop your own research software?* Yes/ No;

Question 9: *Have you received any training in software development?* Yes, self-taught/ Yes, I attended a course/ No;

Question 10: *Have you ever included costs for software development in a funding proposal?* No, I'm not involved in bid writing/ No, but we expected to develop software as part of the project/ Yes, I'm involved in bid writing;

Question 11: *Please provide the name(s) of the main research software you use*;

Extra question 1: *What is your job title?* Lecturer/ PhD/ Postdoc/ Principal Investigator/ Professor/ Reader/ Research Assistant/ Research Associate/ Research Fellow/ Research Manager/ Researcher/ Technician/ Not provided/ Other;

Extra question 2: *What is your gender?* Male/ Female/ Other/ Prefer not to say;

Extra question 3: *Are you employed on a permanent or fixed-term contract?* Fixed term/ Permanent/ Prefer not to say;

Extra question 4: *What is your preferred Operating System?* Microsoft Windows/ Mac OS X/ Linux/ Don't have a preference/ Other;

B.1.2 STEM & non-STEM groups

The STEM group includes the following disciplines available as responses to Question 3 (B.1.1) of the SSI survey:

Biological, Mathematical & Physical sciences

Agriculture, Forestry & Veterinary science

Engineering & Technology

Medicine, Dentistry & Health

The non-STEM group includes the following disciplines available as responses to Question 3 (B.1.1) of the SSI survey:

Administrative & Business studies

Architecture & Planning

Design, Creative & Performing arts

Education

Humanities & Language based studies & Archaeology

Social studies

B.2 Short Questionnaire

The questions of the short questionnaire in the order they were asked:

1 *What is your current job position?*

2 *Which operating system do you use?*

3 *Why are you using HPC?*

4 *Are you planning to attend a similar course [specially organisede for women]?*

5 *Are you aware that there is an underrepresentation of women in the HPC community?*

6 *What do you think is the reason (of women's underrepresentation in the HPC community)?*

7 *How do you think this (the underrepresentation of women in the HPC community) can be solved?*

8 *How important is it (to solve the underrepresentation of women in the HPC community)?*

B.3 HPC Community Survey

The invitation to partipation, the terms and conditions of the survey, and the full list of the survey questions are available in ths section.

B.3.1 Invitation to Participation

The following invitation was sent by email to potential survey participants.

Ms Athina Frantzana a.frantzana@sms.ed.ac.uk EPCC, School of Physics and Astronomy University of Edinburgh

RE: Research project on the underrepresentation of women in HPC

Dear Sir/Madam,

My name is Athina Frantzana and I am a PhD student in EPCC (Edinburgh Parallel Computer Centre) at the University of Edinburgh, UK. I am also a member of the WHPC (Women in High Performance Computing) network. I am currently working on understanding the obstacles facing women in the HPC community and how to improve equality. The main objective of my research is to quantify the demographics based on gender and other data, identify the reasons of the underrepresentation of women in HPC and suggest the best practices in engaging more women in the HPC community. I am supervised by Dr Toni Collis.

One of the methods I use to collect data for my research is questionnaires and surveys. I am looking for people of all genders and of every possible background, *that are currently working in any HPC field/sector and position. I believe your company/university is ideal to participate in my research by responding to our new survey. I would like to finish collecting data until the end of the spring of 2016, so I have enough time to analyse it. If you agree to participate in my research, I would like to ask your assistance in circulating this email or the survey link to your staff members. The link to the survey is:* `https://www.surveymonkey.co.uk/r/HPC-community`[1].

This research is still a work in progress, thus details of the methods and research questions may still evolve. Parts of my research results might be published as scientific papers or articles. I can assure you that the research will be conducted in a professional manner and that any confidential information will be treated anonymously. If you have any questions or concerns, please feel free to contact me on the above details, or my supervisor, Dr. Toni Collis at acollis@epcc.ed.ac.uk.

Yours sincerely, Athina Frantzana

B.3.2 Terms and Conditions

The following paragraph of the terms and conditions of the survey was provided at the beginning of the survey. In Question 1, the participants were asked if they agree with the terms and conditions, and only if they answered *"Yes"*, they consented to participate and continued with the survey questions.

This survey is being conducted to understand the demographics of the HPC community and will be used in research by EPCC and the Women in HPC network. Participation in this study is voluntary and you are under no obligation to complete the survey.

The survey takes approximately 10 minutes to complete.

All data collected is anonymous and will not be shared in its original form with anyone outside the immediate research team and Survey Monkey who retain data according to their privacy policy. Results from the survey will only be published in aggregated form.

If you have any questions about this survey please contact Athina Frantzana or Toni Collis, EPCC, University of Edinburgh, James Clerk Maxwell Building,

[1] The survey closed in May 2016

Peter Guthrie Tait Road, Edinburgh, EH9 3FD, UK.

B.3.3 Survey Questions

Here we include the 37 questions of the survey (and the available response options), as well as the grouping of responses we did to facilitate the analysis.

1 *Do you agree to the above terms?* Yes/ No

2 *Please select your discipline/background (select all that apply)* Physical and Mathematical sciences/ Computer Science/ Engineering and Technology/ Biological Science/ Medical and Health Science/ Humanities and Arts/ Administrative and Business Studies/ Social and Political Science/ Education/ Other (please specify)

3 *How would you describe your work in the context of your interaction with HPC? (Please select all that apply)* HPC applications/tools development/ User of HPC facilities/ User of HPC software/applications/ System administrator or HPC services/ Sales/ Management, HPC services/ Management, HPC center/ Management, HPC industry/commercial/ HPC Training and/or Support/ HPC systems validation/ Other (please specify)

4 *Do you develop your own software?* Yes/ No

5 *How familiar/comfortable are you with the following operating systems? 1 is not at all familiar and 5 is very familiar* Windows/ Linux/Unix/ Mac OS X/ Other (please specify the operating system)

6 *Have you received any software development training?* Yes attending an in-person course/ Yes completed online training/ No self taught but with no structured training/ No I do not develop software

7 *How often do you travel to conferences?* Never / Once per year/ 2 or 3 times per year/ More than 3 times per year

8 *Have you ever had a mentor or provided mentoring (please select all that apply)?* Yes: I am currently providing mentoring/ Yes: I currently have a mentor/ Yes: I have previously had a mentor, but not currently/ Yes: I have previously been a mentor, but not currently/ No: I have never had a mentor but would like to/ No: I have never had a mentor and I do not

want a mentor/ No: I have never been a mentor but would like to/ No: I have never been a mentor and I do not want to be a mentor/ Other (please specify)

9 *Has mentoring been a positive experience for you?* Yes/ No

10 *What aspects of your experience with mentoring made it a positive or negative experience?* (Open-Ended Response)

11 *On a scale of 1-5 please rate the following (1 is not at all successful, 5 is extremely successful) - How successful do you think you are in your current career? - Please explain your answer*

12 *Where do you want to be in five years?* In my current or similar role/ In a more senior role in the same area/ In a different role/ Other (please specify)

13 *Would you consider moving your family to a different part of the country for your work?* Yes/ No

14 *Would you consider moving your family to a different country for your work?* Yes/ No

15 *When considering your workplace and colleagues you interact with, to what extent do you agree with the following statements? - I never experience discrimination by colleagues - I feel that colleagues in my workplace are treated equally by each other regardless of gender - I feel that colleagues in my workplace are treated equally by each other regardless of race or ethnicity - On average, people in my workplace treat women equally - On average, people in my workplace treat racial and ethnic minorities equally* Strongly Disagree/ Disagree/ Neither agree nor disagree/ Agree / Strongly agree

16 *If you disagreed with any of the statements above please provide more information to explain your answer* Open-Ended Response

17 *When considering your experience at conferences and the attendees you interact with, to what extent do you agree with the following statements? - I never experience discrimination from other conference attendees - I feel that attendees are treated equally by each other regardless of gender - I feel that attendees are treated equally by each other regardless of race or ethnicity - On average, people in advanced computing treat women equally - On average, people in advanced computing treat racial and ethnic minorities*

equally Strongly Disagree/ Disagree/ Neither agree nor disagree/ Agree/ Strongly agree

18 *If you disagreed with any of the statements above please provide more information to explain your answer* Open-Ended Response

19 *Of the statements below choose the one that you agree with the most:* Women are under-represented in HPC but I do not think this is something that needs to be changed/ Women are under-represented in HPC and I think there should be more women/ There are equal proportions of men and women in HPC/ None of the above (please specify)

20 *What do you think could be done to increase the proportion of women in the HPC community?* Open-Ended Response

21 *Do you believe working in a mixed gender team* Has a positive effect on team work and productivity/ Has no effect on team work and productivity/ Has a negative effect on team work and productivity - Please explain your answer

22 *Do you believe that a gender balanced international HPC community* Has a positive effect on the community/ Has no effect on the community/ Has a negative effect on the community - Please explain your answer

23 *Do you believe you belong to a minority group?* Yes/ No - If yes, please specify the minority you belong to

24 *Have you heard of the term "Stereotype Threat"?* Yes/ No

25 *"Stereotype threat is when a person believes that their performance in a particular situation may confirm other people's negative stereotypes about them". Have you ever felt that you are confirming other people's negative stereotypes about you and your social group?* Yes/ No - If you have any additional comments please include them here [free text]

26 *Please indicate what sector you currently work in* Industry/Commercial/ Academia/ Non-profit/ Other (please specify)

27 *Please select from the following the option that most closely fits your current role* Research (academia, tenured)/ Research (academia, not-tenured)/ Research (postdoctoral)/ Research (industry)/ Research (student)/ User support/ Software developer/ Systems administrator/ Technician/ Sales/ Teaching/ Marketing/ HPC Center/Service Director/ Other (please specify)

28 *Please select your highest qualification* PhD/ Masters degree/ Undergraduate/First degree/ High School qualifications/diploma/ Other (please specify)

29 *Do you have children?* Yes/ No/ Prefer not to specify

30 *Do you believe that having children has impacted your career?* Yes/ No/ Prefer not to specify

31 *Do you believe that having children would impact your career?* Yes/ No/ Prefer not to specify

32 *Please select your gender* Female/ Male/ Other/ Prefer not to specify

33 *Please select your age* 16-20/ 20-29/ 30-39/ 40-49/ 50-59/ 60-69/ 70+/ Prefer not to specify

34 *Did you attend ISC 2015 (Frankfurt, Germany) and/or Supercomputing 2015 (Austin, Texas, USA)?* ISC 2015/ Supercomputing 2015 (SC15)/ Neither

35 *How many times have you attended Supercomputing?* Never/ SC15 was my first SC conference/ Twice (including SC15)/ 3-5/ 6+

36 *How many times have you attended ISC?* Never/ ISC15 was my first ISC conference/ Twice (including ISC15)/ 3-5/ 6+

37 *Please select all of the registrations you have for SC15 (please select all that apply)* Admission to exhibit hall only - ISC 2015/ Admission to exhibit hall only - SC15/ Technical program - ISC 2015/ Technical program - SC15/ Tutorial passport - ISC 2015/ Tutorial passport - SC15/ Workshop passport - ISC 2015/ Workshop passport - SC15/ Exhibitor - ISC 2015/ Exhibitor - SC15/ Press pass - ISC 2015/ Press pass - SC15/ Other (please specify)

The **STEM group** included the options:

Physical and Mathematical Sciences

Computer Science

Engineering and Technology

Biological Science

Medical and Health Science

The **non-STEM group** included the options:

Humanities and Arts

Administrative and Business Studies

Social and Political Science

Education

B.4 The Use of HPC by Non-STEM Disciplines

Here, we include the invitation to participation to the survey, and the survey questions.

Invitation:

Dear Sir/Madam,

My name is Athina Frantzana and I am a PhD student at the University of Edinburgh. I work for the Edinburgh Parallel Computing Centre (EPCC) and Women in High Performance Computing (WHPC), under the supervision of Dr. Toni Collis. My research is focused on the gender balance and differences in STEM, and more specifically in High Performance Computing (HPC).

My research so far has shown that the majority of the HPC users come from a STEM background. At the moment I am trying to find the reasons why this happens, with the hypothesis that HPC and programming are not popular for research in non-STEM disciplines, where the gender balance is better than STEM.

One of the methods I use to collect data for my research is surveys. I am currently conducting a very short survey with simple questions which might help me identifying the potential reasons of the under-representation of non-STEM users and of women in the HPC community: [link]

I would like to finish collecting data until the end of September 2017, so I have enough time to analyse it. If you agree to participate in my research, I would like to ask your assistance in circulating this email or the survey link to the academic staff and postgraduates students of your department/school.

This research is still a work in progress, thus details of the methods and research questions may still evolve. Parts of my research results might be published as scientific papers or articles. I can assure you that the research will be conducted in a professional manner and that any confidential information will be treated anonymously. If you have any questions or concerns, please feel free to contact me, or my supervisor, Dr. Toni Collis at acollis@epcc.ed.ac.uk.

Thank you in advance.

Best wishes, Athina Frantzana

Questions:

This short survey started by asking a first main question:

1 *Do you know what High Performance Computing (HPC) is and what is used for?*

Then, according to the participant's response (Yes/ No), the survey would lead the participant to two different second questions.

The respondents who stated that they knew what HPC is and what is used for, were asked the following question:

2 *Do you use HPC for your research?*

The respondents who answered *"No"* to Question 1, were provided with an explanation on what HPC is and what is used for (*"High Performance Computing (HPC) refers to any computational activity requiring more than a single computer to execute a task. It is used to solve advanced computational problems in various research fields."*; then they were asked the following question:

2 *Would you be interested in using HPC for your research?*

Later, both groups of respondents were asked the same following questions (and available response options):

3 *Does your school/department offer any computational/programming course/-training?* Yes/ No

4 *Which Operating System do you prefer?* Windows/ Linux/ Mac OSX

5 *Please choose the subject area of your work/studies* Social & Political science/ Arts & Humanities/ Business/ Education/ Other (Please specify)

6 *Please choose your current job position* Postgraduate student (Master/PhD)/ PDRA or Early-career stage researcher/ Lecturer or Reader/ Professor/ Other

7 *Please choose your gender* Female/ Male/ Prefer not to say/ Other

8 *Please choose your age group* < 25/ $25 - 35$/ $35 - 45$/ $45 - 55$/ $55 - 65$/ > 65

Appendix C

Qualitative Approaches

C.1 Interview Documentation

We included here: the interview invitation, which was sent by email to targeted and recommended individuals of the HPC community, the script of the interview consent, which was agreed verbally before the interview, and the questions that were aimed to be asked during the interview. The audio data of the verbal consent and the interview were recorded and archived together for each participant. We also included the written consent, which was sent to the interviewees who had chose to be interviewed by email, prior to the interview (C.1).

Invitation:

"Dear [name],

My name is Athina Frantzana and I am a PhD student at EPCC, the University of Edinburgh. My research topic is about the gender representation and balance in the HPC community, and my supervisor is Dr. Toni Collis.

Currently, I am conducting interviews with people that work for HPC or are HPC users. For this reason, I would like to invite you to be interviewed and discuss about your experiences in working in HPC, as well as your personal views on the gender balance matter.

Please, let me know if you are interested in contributing to our research, and the dates and times that work best for you for a Skype call interview in February.

Looking forward to your response and thank you in advance.

Best wishes, Athina Frantzana."

Interview, Questions and Verbal Consent script:

"Hello, I am Athina Frantzana. I am conducting interviews about the representation of women in the HPC community as part of my PhD research at EPCC at the University of Edinburgh. Im working under the direction of my supervisor, Dr. Toni Collis. If you have any concerns about this study or my behaviour, you can contact her and I will provide you with her contact details.

If you agree to participate in this study, I will ask you questions about your personal experiences in working in HPC or using HPC facilities, and your views on gender equality matters. This usually lasts about 45-60 minutes. Some of the questions may be personal or sensitive. You are free to choose not to answer any question. I will also stop taking notes if you prefer. You are also free to end the conversation at any time. If you decide to stop the conversation, you do not have to give me a reason why.

Our discussion will be audio recorded for my own research records. Anything you tell me will be private. I will not use your name or any details that might identify you when I write and publish my work. I will not tell anyone that we have had a conversation, and I will not reveal personal details. My notes and the audio data will be stored securely. I can provide you with the transcript of the full conversation on request.

You will not be paid for taking part in this study, but your contribution to this study is highly appreciated. I would be pleased to send you a short summary of the study results when I finish going over our results. Please let me know if you would like a summary and what would be the best way to get this to you.

Do you have any questions about me, my project, or this conversation before we begin? [Answer questions]

Do you agree to participate in this study knowing that you can withdraw at any point with no consequences to you? [If yes, begin the interview.] [If no, thank the participant for his/her time.]

Questions:

1 *What is your current job? Describe how you are related to HPC and the*

route of your career until now.

2 *Do you future career plans?*

3 *Why did you choose an HPC career/to use HPC facilities?*

4 *Did you have to attend any HPC training courses when you started using HPC? Do you have to attend any courses now/often?*

5 *If you couldnt work in HPC, what other career/job would you choose?*

6 *Were there any obstacles/challenges when you first enter the HPC field?*

7 *Are you facing any obstacles/challenges in the workplace at the moment?*

8 *Is there anything you would like to change in the HPC field/community?*

9 *What would/do you do to promote HPC?*

10 *Do you think that HPC facilities are only for use in STEM research? Is HPC known and available in other research areas?*

11 *Do you believe there is an under-representation of women in HPC? Why/why not?*

12 *Do you think that gender balanced teams produce higher performance and better results?*

13 *How important are female role models and female leaders?*

14 *How can parenthood impact the career of both men and women?*

15 *What is your opinion on mentoring? (being mentor being mentee)*

16 *Have you ever heard of the terms: Unconscious Bias, Stereotype Threat, Impostor Syndrome, In-group Out-group Bias? (What do you know? Do you identify yourself with any of these?)*

17 *Have you ever attended an E&D training? How did you find it? Do you think it is important?*

18 *Are there any social activities at your workplace? Do you join them? Are they important? (added 25/3)*

19 *What do you think it should be done to achieve equality and diversity in HPC?*

We are getting close to the end of our time and our interview and before I thank you for your time and your precious responses, I would like to ask you if there is anything you would like to add or if there is any question that you believe I should have asked you and I should add for future use (Question 20).

Thank you for supporting our study. Have a nice day."

C.2 Focus Group Discusssions Documentation

Here we included a sample of the invitations sent to individuals of the HPC community, who might have been interested in participating to the focus group discussions of our study (C.2). We also included the written consent that the focus group discussion participants read and signed before the discussion (C.3).

Interview for HPC community – by Athina Frantzana

Consent:

I am Athina Frantzana. I am conducting interviews about the representation of women in the HPC community as part of my PhD research at EPCC at the University of Edinburgh. I'm working under the direction of my supervisor, Dr. Toni Collis. If you have any concerns about this study or my behaviour, you can contact her and I will provide you with her contact details.

If you agree to participate in this study, I will ask you questions about your personal experiences in working in HPC or using HPC facilities, and your views on gender equality matters.

Some of the questions may be personal or sensitive. You are free to choose not to answer any question and you do not have to give me a reason why.

Anything you tell me will be private. I will not use your name or any details that might identify you when I write and publish my work. I will not tell anyone that we have had a conversation, and I will not reveal personal details. My notes will be stored securely.

You will not be paid for taking part in this study, but your contribution to this study is highly appreciated.

I would be pleased to send you a short summary of the study results when I finish going over my results. Please let me know if you would like a summary and what would be the best way to get this to you.

If you agree with all the above and you would like to participate in this study, please sign here:

____

Thank you.

Figure C.1 *The written consent which was sent to the by-email interviewees. It was signed and sent back by the interviewees before they received the interview questions.*

Are you an HPC user or do you work in the HPC community? If so, then I would like to invite you to take part in a focus group (small discussion group) about gender balance in the HPC community, which will be held in Leeds on Wednesday the 5th of April 2017, alongside the Women in HPC (WHPC) event. Please note everyone is welcome to join this discussion, irrespective of gender, and we would like to encourage participation in these discussions from men as well as women.

The focus group should last no longer than one hour. Your views will be used to help us with our research on understanding better women's representation and needs in the HPC community.

If you are interested in participating in this focus group, please email a.frantzana@sms.ed.ac.uk. The deadline for replying is the 19th of March.

I very much hope you will be able to take part and look forward to hearing from you.

Best wishes,

Athina Frantzana

Figure C.2 *A sample of the invitations to focus group discussion sent to the HPC community.*

Focus Group Discussion for HPC Community –
by Athina Frantzana, Ph.D. Candidate, University of Edinburgh

Consent Form

Purpose:

The purpose of the study is to examine the female representation in the High Performance Computing (HPC) community. Specifically, we want to understand if there is gender balance in HPC and if not why, and suggest solutions.

Procedures:

If you participate in this study, you will be in a group of approximately 4-6 people. There will be a facilitator who will ask questions and facilitate the discussion, and one note-taker to write down the ideas expressed within the group. This focus group will be recorded (video or Dictaphone), so it can be transcribed after the focus group is held. If you volunteer to participate in this focus group, you will be asked some questions relating to your experience with working in HPC and your views on gender matters. These questions will help us to better understand the female representation in HPC.

Your participation is completely voluntary. You may withdraw from this study at any time without penalty.

Benefits and Risks:

No risk greater than those experienced in ordinary conversation are anticipated. Everyone will be asked to respect the privacy of the other group members. All participants will be asked not to disclose anything said within the context of the discussion, but it is important to understand that other people in the group with you may not keep all information private and confidential.

Confidentiality:

The transcript will be anonymized before the researchers perform any analysis. No individual participant will be identified or linked to the results. The results of this study may be presented in the doctoral thesis of the researcher; however, your identity or any identifying information will not be disclosed. All information obtained in this study will be kept strictly confidential. All materials will be stored in a securely, safely and in accordance with Data Collection Act within the University of Edinburgh, and access to files will be restricted to the researcher and her supervisor.

Consent:

By signing this consent form, you are indicating that you fully understand the above information and agree to participate in this focus group.

Participant's signature: ____

Printed name: ____

Date: ____

Figure C.3 *The written consent which was read and signed by the focus group discussion participants prior to the discusssion.*

Bibliography

[1] A. Frantzana A.B.K. Collis C.E. MacPhee. Gender differences in research software: A uk survey. *Journal of Open Research Software*, 2016.

[2] UNESCO Institute for Statistics. Women in science, fact sheet no 51. June 2018.

[3] UNESCO Institute for Statistics. Women in science, quarterly thematic publication. March 2015.

[4] J. S. Hyde M. C. Linn. Gender similarities in mathematics and science. *Science*, 321:599–600, 2006.

[5] I. J. Deary et al. Skull size and intelligence ad king robert bruce's iq. *Intelligence*, 35:519–525, 2007.

[6] R. J. Haier R. E. Jung R. A. Yeo K. Head M. T. Alkire. The neuroanatomy of general intelligence: sex matters. *NeuroImage*, 25:320–327, 2005.

[7] J. Valla S. J. Ceci. Can sex differences in science be tied to the long reach of prenatal hormones? brain organization theory, digit ratio (2d/4d), and sex differences in preferences and cognition. *Prospect Psychol Sci.*, 6:134–136, 2011.

[8] M. Hines et al. Prenatal androgen exposure alters girls' responses to information indicating gender-appropriate behaviour. *Philosophical Transactions B*, 371, 2015.

[9] S. J. Ceci W. M. Williams. *Why aren't more women in science?: Top researchers debate the evidence.* American Psychological Association, first edition, 2007.

[10] S. J. Ceci W. M. Williams S. M. Barnett. Women's underrepresentation in science: Sociocultural and biological considerations. *Psychological Bulletin*, 135:218–261, 2009.

[11] C. Fine. *Delusions of Gender - The real science behind sex differences.* Icon Books Ltd, UK, second edition, 2011.

[12] A. E. Preston. *Leaving Science: Occupational exit from scientific careers.* New York: Russell Sage Foundation, first edition, 2004.

[13] M. M. Ferreira. Gender differences in graduate students' perspectives on the culture of science. *Journal of Women and Minorities in Science and Engineering*, 9, 2003.

[14] K. Bradley M. Charles. *A Matter of Degrees: Female Underrepresentation in Computer Science programs cross-nationally. In J. McGrath Cohoon and W. Aspray (Editors), Women and Information Technology: Research on underrepresentation.* MIT Press, 2006.

[15] J.Eccles B. Barber D. Jozfowicz. *Linking gender to educational, occupational and recreational choices: Applying the Eccles et al. related model of achievement related choices. In W. Swan, J. Langlois and L. Gilbert (Editors), Sexism and stereotypes in modern society: The gender science of Janet Taylor Spence.* Washington DC: American Psychological Association Books, 1999.

[16] A. M. Masnick S. S. Valenti B. D. Cox C. J. Osman. A multidimensional scaling analysis of students' attitudes about science careers. *International Journal of Science Education*, 2:653–667, 2009.

[17] R Su J. Rounds P. I. Armstrong. Men and things, women and people: A meta-analysis of sex differences in interests. *Psychological Bulletin*, 135:859–884, 2009.

[18] Y. Yang J. M.Barth. Gender differences in stem undergraduates' vocational interests: Peoplething orientation and goal affordances. *Journal of Vocational Behavior*, 91:65–75, 2015.

[19] A. B. Diekman E. R.Brown A. M. Johnson E. K. Clark. Seeking congruity between goals and roles: A new look at why women opt out of science, technology, engineering, and mathematics careers. *Psychological Science*, 21:1051–1057, 2010.

[20] S-J. Leslie A.Cimpian M. Meyer E. Freeland. Expectations of brilliance underlie gender distributions across academic disciplines. *Science*, 347:262–265, 2015.

[21] J. L. Smith K. L. Lewis L. Hawthorne S. D. Hodges. When trying hard isnt natural - womens belonging with and motivation for male-dominated stem fields as a function of effort expenditure concerns. *Personality and Social Psychology Bulletin*, 39:131–143, 2013.

[22] A. Kiefer M.Shih. Gender differences in persistence and attributions in stereotype relevant contexts. *Sex Roles*, 54:859–868, 2006.

[23] C. Hakim. Women, careers, and work-life preferences. *British Journal of Guidance and Councelling*, 34:279–294, 2006.

[24] D. W. Leslie. The reshaping of america's academic workforce. *Research Dialogue - TIAA-CREF institute*, 2007.

[25] M. A. Mason M. Goulden. Marriage and baby blues: Redefining gender equity in the academy. *The ANNALS of the American Academy of Political and Social Science*, 596:86–103, 2004.

[26] J. A. Jacobs S. E. Winslow. Overworked faculty: Job stresses and family demands. *The ANNALS of the American Academy of Political and Social Science*, 596:104–129, 2004.

[27] D. K. Ginther S. Kahn. Does science promote women? evidence from academia 19732001. *Unpublished Manuscript*, 2006.

[28] D. F. Halpern et al. The science of sex differences in science and mathematics. *Psychol Sci Public Interest*, 8:1–51, 2007.

[29] D. E. Shalala et al. *Beyond Bias and Barriers: Fulfilling the Potential of Women in Academic Science and Engineering.* National Academies Press, 2006.

[30] C. Herman S. Lewis. Entitled to a sustainable career? motherhood in science, engineering, and technology. *Journal of Social Issues*, 68:767–789, 2013.

[31] M. J. Hodges M. J. Budig. Who gets the daddy bonus? organizational hegemonic masculinity and the impact of fatherhood on earnings. *Gender and Society*, 24:717–745, 2010.

[32] J. L. Glass S. Sassler Y. Levitte K. M. Michelmore. What's so special about stem? a comparison of women's retention in stem and professional occupations. *Soc Forces*, 92:723–756, 2013.

[33] J. Hunt. Why do women leave science and engineering? *National Bureau of Economic Research Working Paper Series*, 2010.

[34] J. L. Smith C. Samsone P. H. White. The stereotyped task engagement process: The role of interest and achievement motivation. *Journal of Educational Psychology*, 99:99–114, 2007.

[35] H. Lytton D. M. Romney. Parents' differential socialization of boys and girls: A meta-analysis. *Psychological Bulletin*, 109, 1991.

[36] C. Wolfgang L. Stannard I. Jones. Block play performance among preschoolers as a predictor of later school achievement in mathematics. *Journal of Research in Childhood Education*, 15:173–180, 2001.

[37] D. B. Downey A. S. Vogt Yuan. Sex differences in school performance during high school: Puzzling patterns and possible explanations. *The Sociological Quarterly*, 46:299–321, 2005.

[38] M. G. Jones J. Wheatley. Gender differences in teacherstudent interactions in science classrooms. *Journal of Research in Science Teaching*, 27:861–874, 1990.

[39] J. R. Becker. Differential treatment of females and males in mathematics classes. *Journal of Research in Mathematics Education*, 12:40–53, 1981.

[40] V. Lavy. On the origins of gender human capital gaps: short and long term consequences of teachers' stereotypical biases. *National Bureau of Economic Research Working Paper Series*, 2015.

[41] K. A. Frank et al. The social dynamics of mathematics coursetaking in high school. *AJS*, 113:1645–1696, 2008.

[42] C. Leaper T. Farkas C. S. Brown. Adolescent girls' experiences and gender-related beliefs in relation to their motivation in math/science and english. *Journal of Youth and Adolescence*, 41:268–282, 2012.

[43] U. Kessels B. Hannover. When being a girl matters less: Accessibility of genderrelated selfknowledge in singlesex and coeducational classes and its impact on students' physicsrelated selfconcept of ability. *British Journal of Educational Psychology*, 78:273–289, 2008.

[44] J. Steele. Children's gender stereotypes about math: The role of stereotype stratification. *Journal of Applied Social Psychology*, 33:2587–2606, 2003.

[45] A. H. Herzig. Becoming mathematicians: Women and students of color choosing and leaving doctoral mathematics. *Review of Educational Research*, 74:171–214, 2004.

[46] D. C. Holland M. A. Eisenhart. *Educated in Romance: Women, Achievement, and College Culture*. University of Chicago Press, 1991.

[47] C. M. Steele J. Aronson. Stereotype threat and the intellectual test performance of african americans. *Journal of Personality and Social Psychology*, 69:797–811, 1995.

[48] E. S. Shaffer D. M. Marx R. Prislin. Mind the gap: Framing of women's success and representation in stem affects women's math performance under threat. *Sex Roles*, 68:454–463, 2003.

[49] A. C. Krendl J. A. Richeson W. M. Kelley T. F. Heatherton. The negative consequences of threat - a functional magnetic resonance imaging investigation of the neural mechanisms underlying women's underperformance in math. *Psychological Science*, 19:168–175, 2008.

[50] M. Cadinu A. Maass A. Rosabianca J. Kiesner. Why do women underperform under stereotype threat? evidence for the role of negative thinking. *American Psychological Society*, 16:572–578, 2005.

[51] P. R. Clance S. Imes. The imposter phenomenon in high achieving women: Dynamics and therapeutic intervention. *Psychotherapy Theory, Research and Practice*, 15, 1978.

[52] J. C. Royse Roskowski. *Imposter Phenomenon and Counceling Self-Efficacy: The Impact of Imposter Feelings. PhD Thesis.* Ball State University, Indiana, 2010.

[53] S. Kumar C. M. Jagacinski. Imposters have goals too: The imposter phenomenon and its relationship to achievement goal theory. *Personality and Individual Differences*, 40, 2006.

[54] C. A. Moss-Racusin J. F. Dovidio V. L. Brescoll M. J. Graham J. Handelsman. Science faculty's subtle gender biases favor male students. *PNAS*, 109:16474–16479, 2012.

[55] E. Reuben P. Sapienza L. Zingales. How stereotypes impair women's careers in science. *PNAS*, 111:4403–4408, 2014.

[56] K. De Welde S. L. Laursen. The glass obstacle course: Informal and formal barriers for women phd students in stem fields. *International Journal of Gender, Science and Technology*, 3:572–595, 2011.

[57] L. Bornmann R.Mutz H-D. Daniel. Gender differences in grant peer review: A meta-analysis. *Journal of Informetrics*, 1:226–238, 2007.

[58] F. Trix C. Psenka. Exploring the color of glass: Letters of recommendation for female and male medical faculty. *Discourse and Society*, 14:191–200, 2003.

[59] E. Judson et al. Examination of implicit gender biases among engineering faculty when assigning leadership, research, and service roles. *American Society for Engineering Education*, 2017.

[60] M. F. Fox. Women, science, and academia : Graduate education and careers. *Gender and Society*, 15:654–666, 2001.

[61] L.A. Isbell T.P. Young A.H. Harcourt. Stag parties linger: Continued gender bias in a female-rich scientific discipline. *PLoS ONE*, 7, 2012.

[62] J. Schroeder et al. Fewer invited talks by women in evolutionary biology symposia. *Journal of Evolutionary Biology*, pages 2063–2069, 2014.

[63] R.F. Kalejta A.C.Palmenberg. Gender parity trends for invited speakers at four prominent virology conference series. *Journal of Virology, American Society of Microbiology*, 91, 2017.

[64] V. Lariviere C.Ni Y. Gingras B. Cronin CR. Sugimoto. Bibliometrics: Global gender disparities in science. *Nature*, 504:211–213, 2013.

[65] J.D. West J.Jacquet M.M. King S.J. Correll C.T. Bergstrom. The role of gender in scholarly authorship. *Physics and Society*, arXiv:1211.1759, 2012.

[66] D. Conley J. Stadmark. Gender matters: A call to commission more women writers. *Nature*, 488, 2012.

[67] A. L. Hopkins J. W. Jawitz C. McCarty A. Goldman N. B. Basu. Disparities in publication patterns by gender, race and ethnicity based on a survey of a random sample of authors. *Scientometrics*, 96:515–534, 2013.

[68] L. A. Rhoton. Distancing as a gendered barrier: Understanding women scientists' gender practices. *Gender and Society*, 25:696–716, 2011.

[69] S. J. Ceci. Women in academic science: Experimental findings from hiring studies. *Educational Psychologist*, 53:22–41, 2017.

[70] A. G. Greenwald M. R. Banaji. Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102, 1995.

[71] R. E. Steinpress K.A. Anders D. Ritzke. The impact of gender on the review of curricula vitae of job applicants and tenure candidates: A national empirical study. *Sex Roles*, 41, 1999.

[72] M. E. Heilman T. G. Okimoto. Why are women penalized for success at male tasks?: The implied communality deficit. *Journal of Applied Psychology*, 92, 2007.

[73] A. E. Budden et al. Double-blind review favours increased representation of female authors. *Trends in Ecology and Evolution*, 23:4–6, 2007.

[74] A.W. Wooley C.F. Chabris A.Pentland N. Hashmi T.W.Malone. Evidence for a collective intelligence factor in the performance of human groups. *Science Magazine*, 330:686–688, 2010.

[75] A. Joshi. By whom and when is women's expertise recognised? the interactive effects of gender and education in science and engineering teams. *Administrative Science Quarterly*, 59:201–239, 2014.

[76] S. Hoogendoorn H. Oosterbeek M.van Praag. The impact of gender diversity on the performance of business teams: evidence from a field experiment. *Management Science*, 59:iv–vii, 2013.

[77] M.W. Nielsen et al. Opinion: Gender diversity leads to better science. *PNAS*, 114:1740–1742, 2017.

[78] A. Joshi H. Roh. The role of context in work team diversity research: A meta-analytic review. *Academy of Mangement Journal*, 52:599–627, 2009.

[79] J. Allmendinger R. J. Hackman. The more, the better? a four-nation study of the inclusion of women in symphony orchestras. *Social Forces*, 74:423–460, 1995.

[80] B.Vasilescu et al. Gender and tenure diversity in github teams. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 3789–3798, Seoul, Republic of Korea, 2015. ACM, New York, USA.

[81] C.Ashcraft B. McLain E.Eger. *Women in Tech: The Facts*. National Center for Women & Technology (NCWIT), 2016.

[82] M.A. Whitecraft W. M. Williams. *Why aren't more women in Computer Science? In A. Oram and G. Wilson (Editors), Making Software*. O'Reilly Media, Inc., first edition, 2011.

[83] L. Schiebinger M. Schraudner. Interdisciplinary approaches to achieving gendered innovations in science, medicine, and engineering. *Interdisciplinary Science Reviews*, 36:154–167, 2011.

[84] Sir A. Morse KCB. *Report: Delivering STEM skills for the economy*. National Audit Office Press Office, 2018.

[85] N. Broughton. *In The Balance - The STEM human capital crunch*. The Social Market Foundation, first edition, 2013.

[86] R. Marschke S. Laursen J. McCarl Nielsen P. Dunn-Rankin. Demographic inertia revisited: An immodest proposal to achieve equitable gender representation among faculty in higher education. *Journal of Higher Education*, 78:1–26, 2007.

[87] C. S. Morrissey M. L. Schmidt. Fixing the system, not the women: An innovative approach to faculty advancement. *Journal of Women's Health*, 17, 2008.

[88] M. Levine N. Serio B. Radaram S. Chaudhuri W. Talbert. Addressing the stem gender gap by designing and implementing an educational outreach chemistry camp for middle school girls. *Journal of Chemical Education*, 92:1639–1644, 2015.

[89] H. S. Mosatche S. Matloff-Nieves L. Kekelis E. K. Lawner. Effective stem programs for adolescent girls - three approaches and many lessons learned. *Afterschool Matter*, pages 17–25, 2013.

[90] G. Naizer M. J. Hawthorne T. B. Henley. Narrowing the gender gap: Enduring changes in middle school students' attitude toward math, science and technology. *Journal of STEM Education*, 15:29–34, 2014.

[91] E. Ruggs M. Hebl. *Diversity, Inclusion, and Cultural Awareness for Classroom and Outreach Education. In B. Bogue and E. Cady (Editors), Apply Research to Practice (ARP) Resources*. Assessing Women and Men in Engineering, 2012.

[92] S. Cheryan A. Master A. N. Meltzoff. Cultural stereotypes as gatekeepers: increasing girls' interest in computer science and engineering by diversifying stereotypes. *Frontiers in Psychology*, 6, 2015.

[93] J. C. Blickenstaff. Women and science careers: leaky pipeline or gender filter? *Gender and Education*, 17:369–386, 2005.

[94] S. Cheryan V. C. Plaut C. Handron L. Hudson. The stereotypical computer scientist: Gendered media representations as a barrier to inclusion for women. *Sex Roles*, 69:58–71, 2013.

[95] A. Griffith. Persistence of women and minorities in stem field majors: Is it the school that matters? *Economics of Education Review*, 29:911–922, 2010.

[96] E. S. Weisgram R. S. Bigler. Girls and science careers: The role of altruistic values and attitudes about scientific tasks. *Journal Of Applied Developmental Psychology*, 27:326–348, 2006.

[97] T. Breda J. Grenet M. Monnet C. Van Effenterre. Can female role models reduce the gender gap in science? evidence from classroom interventions in french high schools. *Paris School of Economics Workin Papers*, 2018–06, 2018.

[98] A. Kalev. Cracking the glass cages? restructuring and ascriptive inequality at work. *American Journal of Sociology*, 114:1591–1643, 2009.

[99] S. Stack. Gender, children and research productivity. *Research in Higher Education*, 45:891–920, 2004.

[100] A. Casadevall J.Handelsman. The presence of female conveners correlates with a higher proportion of female speakers at scientific symposia. *mBio*, 5, 2014.

[101] E. Fehr H. Bernhard B. Rockenbach. Egalitarianism in young children. *Nature*, 454:1079–1083, 2008.

[102] M. Van Vugt D. De Cremer D. P. Janssen. Gender differences in cooperation and competition - the male-warrior hypothesis. *Psychological Science*, 18:19–23, 2007.

[103] L. A. Rudman S. A. Goodwin. Gender differences in automatic in-group bias: Why do women like women more than men like men? *Journal Of Personality and Social Psychology*, 87:494–509, 2004.

[104] E. Apospori I. Nikandrou L. Panayotopoulou. Mentoring and women's career advancement in greece. *Human Resource Development International*, 9:509–527, 2007.

[105] T. D. Allen L. T. Eby M. L. Poteet E. Lentz L. Lima. Career benefits associated with mentoring for proteges: A meta-analysis. *Journal of Applied Psychology*, 89:127–136, 2004.

[106] M. Mollica L. Nemeth. Outcomes and characteristics of faculty/student mentorship in phd programs. *American Journal of Educational Research*, 2:703–708, 2014.

[107] D. Bilimoria S. Joy X. Liang. Breaking barriers and creating inclusiveness: Lessons of organizational transformation to advance women faculty in academic science and engineering. *Human Resource Management*, 47:423–441, 2008.

[108] J. Reid E. Smith N. Iamsuk J. Miller. Balancing the equation: Mentoring first-year female stem students at a regional university. *International Journal of Innovation in Science and Mathematics Education*, 24:18–30, 2016.

[109] A. Kahveci S. A. Southerland P. J. Gilmer. Retaining undergraduate women in science, mathematics, and engineering. *Journal of College Science Teaching*, 36:34–38, 2006.

[110] A. Roberts. Mentoring revisited: A phenomenological reading of the literature. *Mentoring & Tutoring: Partnership in Learning*, 8:145–170, 2000.

[111] S. T. Gorman M. C. Durmowicz E. M. Roskes S. P. Slattery. Women in the academy: Female leadership in stem education and the evolution of a mentoring web. *Forum on Public Policy: A Journal of the Oxford Round Table*, 2010.

[112] A .J. Ko L. Hwa K. Davis J. C. Yip. Informal mentoring of adolescents about computing: Relationships, roles, qualities, and impact. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, pages 598–603, Baltimore, USA, 2018. ACM, New York, USA.

[113] P. A.Stout J. Staiger N. A. Jennings. Affective stories: Understanding the lack of progress of women faculty. *NWSA*, 19:124–144, 2007.

[114] S. J. Correll. Sws 2016 feminist lecture: Reducing gender biases in modern workplaces: A small wins approach to organizational change. *Gender & Society*, 31:725–750, 2017.

[115] M. Carnes et al. Effect of an intervention to break the gender bias habit for faculty at one institution: A cluster randomized, controlled trial. *Acad Med*, 90, 2015.

[116] A. Kalev F. Dobbin E. Kelly. Best practices or best guesses? assessing the efficacy of corporate affirmative action and diversity policies. *American Sociological Review*, 71:589–617, 2006.

[117] F. Dobbin A. Kalev. Spotlight on building a diverse organisation - why diversity programs fail and what works better. *Harvard Business Review*, 2016.

[118] D. Atewologun T. Cornish F. Tresh. *Research Report 113 - Unconscious bias training: An assessment of the evidence for effectiveness*. Equality and Human Rights Commission, 2018.

[119] C. A. Moss-Racusin C. Sanzari N. Caluori H. Rabasco. Gender bias produces gender gaps in stem engagement. *Sex Roles*, pages 1–20, 2016.

[120] K. B. Whittington. Mothers of invention? gender, motherhood, and new dimensions of productivity in the science profession. *Work and Occupations*, 38:417–456, 2011.

[121] M. Mercer. Shared parental leave take-up levels, different approaches, and the keys to successful implementation. *Institutefor Employment Studies*, 2016.

[122] H. Brikett S. Forbes. Shared parental leave: Why is take up so low and what can be done? (policy brief). *Birmingham Business School, WIRC - Inclusion Matters*, 2018.

[123] C. Schimpf M. M. Santiago J. Hoegh D. Banerjee A. Pawley. Stem faculty and parental leave: Understanding an institution's policy within a national policy context through structuration theory. *International Journal of Gender, Science and Technology*, 5:102–125, 2013.

[124] E. Drew A. L. Humbert. "men have careers, women have babies": unequal parental care among irish entrepreneurs. *Community, Work & Family*, 15:49–67, 2010.

[125] L. Haas C. P. Hwang. The impact of taking parental leave on fathers' participation in childcare and relationships with children: Lessons from sweden. *Community, Work & Family*, 11:85–104, 2008.

[126] S. Mattauch K. Lohmann F. Hannig D. Lohmann J. Teich. The gender gap in computer science a bibliometric analysis,. *Friedrich-Alexander-Universitt Erlangen-Nrnberg, Dept. of Computer Science, Technical Reports, CS-2018-02,*, 2018.

[127] D. Ford J. Smith P. J. Guo C. Parnin. Paradise unplugged: identifying barriers for female participation on stack overflow. In *Proceedings of the 2016 24th ACM SIGSOFT International Symposium on Foundations of Software Engineering*, pages 846–857, Seattle, USA, 2016. ACM, New York, USA.

[128] J. Terrell et al. Gender differences and bias in open source: pull request acceptance of women versus men. *PeerJ CompSci*, 3, 2017.

[129] M. Burnett et al. Gender differences and programming environments: Across programming populations. In *Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*. ACM, New York, USA, 2010.

[130] A. Brett et al. *Research Software Engineers: State of the Nation Report 2017*. Research Software Engineer Network (RSEN), 2017.

[131] K.S. Bonham M. I. Stefan. Women are underrepresented in computational biology: An analysis of the scholarly literature in biology, computer science and computational biology. *PLOS Computational Biology*, 13, 2017.

[132] D. de Vaus. *Social Research Today: Surveys in Social Research*. Routledge, fifth edition, 2002.

[133] C. Davis. *SPSS step by step: Essentials for Social and Political Science*. The Policy Press, first edition, 2013.

[134] K. M. DeWalt Billie R. DeWalt. *Participant Observation: A guide for fieldworkers*. AltaMira Press, 2002.

[135] H. Russel Bernard. *Research Methods in Anthropology: Qualitative and Quantitative Approaches*. AltaMira Press, fifth edition, 2011.

[136] M.V. Angrosino K.A. Mays dePerez. *Rethinking Observation: From Method to Context. In N. K. Denzin and Y. S. Lincoln (Editors), Handbook of Qualitative Research*. SAGE Publications, Inc., second edition, 2000.

[137] J.R.A. Davenport M. Fouesneau E. Grand A. Hagen K. Poppenhaeger L.L. Watkins. Studying gender in conference talks: Data from the 23rd meeting of the american astronomical society. *Physics and Society*, arXiv:1403.3091, 2014.

[138] A. Hinsley WJ.Sutherland A. Johnston. Men ask more questions than women at a scientific conference. *PLoS ONE*, 10, 2017.

[139] L. Borghans M. Romans J. Sauermann. What makes a good conference? analysing the preferences of labor economists. *Discussion paper series// Forschungsinstitut zur Zukunft der Arbeit*, 4870, 2010.

[140] M. A. Mason M.Goulden. Do babies matter (part ii)? closing the baby gap. *Academe, American Association of University Professors*, 90:10–15, 2004.

[141] C. A. Moss-Racusin L.A.Rudman. Disruptions in women's self-promotion: the backlash avoidance model. *Psychology of Women Quarterly*, pages 186–202, 2010.

[142] W.F. Laurance S.G. Laurance D.C. Useche. Gender differences in science: no support for the "homer simpson effect" among tropical researchers. *Trends in Ecology and Evolution*, 26:262–263, 2011.

[143] H.L. Dugdale M. Hinsch J. Schroeder. Biased sampling: no "homer simpson effect" among high achievers. *Trends in Ecology and Evolution*, 26:622–623, 2011.

[144] R. Sutton Koeser. Trusting others to "do the math". *Interdisciplinary Science Reviews*, 40, 2015.

[145] S. Hettrick. It's impossible to conduct research without software, say 7 out of 10 uk researchers. *Software Sustainability Institute Blog, www.software.ac.uk/blog*, 2014.

[146] Y. Hayashi et al. The missing link in engineering education: The arts and humanitites. *Workshop Proceedings of the 25th International Conference on Computers in Education, New Zeland: Asia-Pacific Society fro Computers in Education*, page 418 424, 2017.

[147] L.M. Laird. A new pathway: A software engineering master's program for liberal arts graduates. *American Society for Engineering Education*, 2017.

[148] S. Doyle N. Senske. Democratizing access and identifying inequalities: Gender, technology, architecture. *Architecture Conference Proceedings and Presentations*, pages 56 – 62, 2017.

[149] E. Vzquez-Cano E. Lopez Meneses E. Garca-Garzn. Differences in basic digital competences between male and female university students of social sciences in spain. *International Journal of Educational Technology in Higher Education*, 14, 2017.

[150] I. Seidman. *Interviewing as Qualitative Research: A guide for researchers in Education and the Social Sciences.* Teachers College Press, third edition, 2006.

[151] B. Diccico-Bloom B.F. Crabtree. The qualitative research interview. *Medical Education*, pages 314–321, 2006.

[152] R. Edwards J. Holland. *What is Qualitative Interviewing?* Bloomsbury Publishing Plc, first edition, 2013.

[153] D.W. Stewart P.N. Shamdasani D.W. Rook. *Focus Groups: Theory and Practice.* Sage Publications, Inc., second edition, 2007.

[154] R.A. Krueger M.A. Casey. *Focus Groups: A Practical Guide for Applied Research.* Sage Publications, Inc., fifth edition, 2014.