KU LEUVEN

LOUVAIN
School of Management

Louvain School of Management

and

Katholiek Universiteit Leuven

Gendered effects on scientific novelty.

The link between gender, knowledge, team and the novelty of science in academia.

Research Master's Thesis submitted by
Amandine Deheneffe

with a view of getting the degrees

Master 120 in Toegepaste Economische Wetenschappen: Handelsingenieur

Master in Business Engineering, professional focus

Supervisor at KUL
Dennis Verhoeven

Supervisor at LSM
Julie Hermans

Academic Year 2017-2018

Acknowledgements

This thesis would have been impossible without the help of important persons. It is why I want to thank them for their contributions in the redaction of this work.

First of all, I would like to show my gratitude to Dennis Verhoeven, my promotor, for his advices as well as his expertise which highly improved this thesis. My thanks go to Thomas Schaper too, for the help he provided as co-promotor.

I also want to thank Yannik Schenk, who helped me and improved my understanding of different econometric considerations.

I am very thankful to my brother and my friend Théophile for their precious help all the long of the thesis. Special thanks to my great friend Margaux for the important mental support she provided during these 5 years of studies.

Finally, I would like to thank my family members for their attentive ears during the discussions about my thesis. Their support was crucial.

Table of Contents

1 General Introduction

In our actual society, novelty and creativity of research in academia are crucial in order to make society move forward. Novelty and creativity are related, and often used as synonyms in literature. In fact, creativity is the sum of two components: novelty and impact (Y. N. Lee, Walsh, & Wang, 2015). In this thesis, we will focus on the former procedures which deal with the novelty part of creativity. Precisely, throughout the argumentation, a combinatorial definition of novelty will be used (Schoenmakers & Duysters, 2010).

Novel researches have a higher impact than conventional research (Y. N. Lee et al., 2015; Uzzi, Mukherjee, Stringer, & Jones, 2013). Initiatives promoting novelty, creativeness and innovativeness are nowadays decisive, given the central place taken by knowledge-driven competition.

Nevertheless, some biases play against novelty. Funding agencies are increasingly risk-averse and novel search is intrinsically risky because the result of the research is hard to predict from the inputs (Arrow, 2008; Trapido, 2015). Most funding agencies have incentives schemes with very low tolerance for failure and short time-cycles for grants(Joshua & John, 2012). They tend to invest in less novel, less risky projects, based on existing knowledge rather than projects with new approaches (Azoulay, Zivin, & Manso, 2011). The use of bibliometric indicators makes them even more short-term oriented. As novel search is more likely to take time to be recognised, money is given to surer and short-term projects(Cattaneo, Meoli, & Signori, 2016; Collin, 2014; Garfield, 1980; Hicks, 2012; Monastersky, 2005).

Still, the biggest bias against novelty is due to the gender gap that still prevails in the scientific careers (Frietsch, Haller, Funken-Vrohlings, & Grupp, 2009; Mauleón, Daraio, & Bordons, 2014). The fact that women do not follow scientific tracks as much as men can be explained by both personal factors (specialization profile, education, trainings), more general factors such as the job market structure or the structural organization in academia (Probert, 2005; Schiebinger & Schraudner, 2011) and their divergent personality traits (Fellner & Maciejovsky, 2007; Niederle & Vesterlund, 2007; Van Vugt, De Cremer, & Janssen, 2007). Finally, systematic biases in favor of males in the assessment of the scientific performance for attribution of authorship rights, or in reviews are constantly feeding the gender gap. (Budden et al., 2008; Lissoni, Montobbio, & Zirulia, 2013; Moss-Racusin, Dovidio, Brescoll, Graham, & Handelsman, 2012; Webb, O'Hara, & Freckleton, 2008).

Beside the ethical considerations raised by the gap, the economic implications of the loss of knowledge, potential new ideas and creativity from women are also important (Hunt, Garant, Herman, & Munroe, 2013).

In this thesis, we will highlight to the policy makers, the various reasons why women are prone to enhance novelty and why closing the gender gap in academia is beneficial for the research, and thus to society. The general research question can be described as follow: "Do women presence in a co-authorship network enhance the novelty of the research in academia?"

The thesis is divided into five main sections. The first section contains the review of the literature composed of three parts. The first part explains the impact of the knowledge range and teams of co-authors on the novelty of the research. The second part highlights the particular reasons of why women improve novelty through higher knowledge sharing inside the teams. The last part describes the hypotheses in detail.

The second main section of the thesis contains the econometric considerations and is divided in three parts: the description of the data we will use, the methodology adopted and the ruling out of alternative explanations.

The third section is composed of the results of the statistical analysis and includes the descriptive statistics of our sample and the results for the different regressions.

Finally, the last section is the general conclusion of the thesis and consists of the discussion of the results, recommendations and limitations.

2 Literature review

The literature review contains three main parts. The first part analyzes the concept of combinatorial novelty and the role of knowledge and diversity through teams regarding novelty. The second part considers the different gendered characteristics that impact novelty and explains why women are likely to positively impact novelty. Finally, the last part defines the hypotheses that will be tested in the statistical section.

2.1 Knowledge, teams and combinatorial novelty

Combinatorial novelty is an explorative search process expressed by the combination of existing pieces of knowledge in original and brand-new ways (Y. N. Lee et al., 2015) This combinatorial view of novelty is shared by a wide stream of literature (Franzoni, 2010; Mednick, 1962; Nelson & Winter, 1982; Schoenmakers & Duysters, 2010; Schumpeter, 1934; Simonton, 2004; Wang, Veugelers, & Stephan, 2017; Weitzman, 1998; Youn, Strumsky, Bettencourt, & Lobo, 2015). In this vision, the knowledge expands by reassembling different pieces of knowledge that were not put together before and this statement has several consequences.

First of all, the limit to the expansion of knowledge in society does not lie in our ability to generate new ideas but more in our ability to recombine in useful manners and deal with all the pieces of existing knowledge (Weitzman, 1998).
Secondly, novelty is not seen as the outcome of an individual genius scientist but depends on the searcher environment and social interactions. In 1987, Amabile and Gryskiewicz (as cited in Franzoni, 2010) explained that "there is increasing consensus that creativity depends not only on talent, experience or cognitive capacities, but also on milieu, i.e., the combination of social, environmental and organizational factors that surround an individual and his working environment". Therefore, novelty is linked to social interactions and the world around the scientist.
Third, the associative perspective of novelty relies heavily on the existing knowledge stock, the broader it is, the more chances of useful novel combinations.

2.1.1 The importance of the knowledge base and sharing

Knowledge is the basis in order to make new combinations and trigger novelty. The transfer of knowledge can be done in different manners. On the one hand, an individual can extend its knowledge by studying on its own. Or, on the other hand, knowledge can be transmitted through interactions and collaborations with experts, scientists, other academics. In academic research, collaboration, in the sense of co-authorship is recurring, hence there are a lot of knowledge transfers and spillovers. (Fallah, 2004; Fershtman & Gandal, 2011).

We can make the link between knowledge and combinatorial novelty: the larger the knowledge's stock of an individual or a team, the more chances there are for new combinations. A great part of the knowledge is embodied in people and therefore the only way to access it is through scientific collaborations and teams. Those collaborations favor the exchange of ideas and the creation of new ideas (Mohnen, 2016). When people work in teams, beside the direct knowledge productivity of an individual, each member of a team has a peer effect on the others through the sharing of its own knowledge (Nonaka, Byosiere, Borucki, & Konno, 1994).

Consequently, an effective way to broaden the knowledge base in order to trigger combinatorial novelty is collaboration and working in teams: the more shared and disseminated (distant) the knowledge is, the more likely novel associations will arise.

2.1.2 Teams and diversity: catalyzers for novelty, to a certain extent

2.1.2.1 The burden of knowledge

Nowadays, the role of teams in the discovery of new knowledge is important (Packalen & Bhattacharya, 2015b). There is a shift from individual-based to team-based research . Wuchty, Jones, and Uzzi (2007), found evidences that team size nearly doubled over the last 45 years from 1.9 to attain 3.5 authors per paper. They also noticed that results of team's work surpass solo author work: team researches are more frequently cited and related to high-impact research (Wuchty et al., 2007).

The reason for this increasing use of teams is the phenomenon called "the burden of knowledge" (B. F. Jones, 2009). Novelty of scientific research pushes further the knowledge's frontier: every time a discovery is made, it broadens the existing base of knowledge. As a matter of fact, the increasingly wider knowledge stock compels researchers to specialize in restricted fields, it is not possible for them to learn deeply about various and remoted fields because there is currently just too much information. Yet, this specialization comes at the cost of a less individual, creative and innovative capacity and a greater tendency towards team's work with various specialists in order to solve collectively an issue that spans over different fields (Y. N. Lee et al., 2015; Wuchty et al., 2007).
The basis for our future study will thus be the team, or more precisely the network of co-authors around a focal scientist.

2.1.2.2 Types of diversity and their impacts on the team

The principal reason for the good results of team work is that it brings more diversity. Diversity is a catch-all term defined as “differences between individuals on any attribute that may lead to the perception that another person is different from self”(van Knippenberg, De Dreu, & Homan, 2004, p. 1008). Diversity can be of two types: demographic or technological.

Demographic diversity covers clear and striking characteristics of individuals, such as age, sex, ethnicity and is also called social category diversity (Williams & O'Reilly, 1998).

Technological diversity refers to characteristics that are anchored more deeply in people and aren’t visible at first sight such as education, functional-background, knowledge and skills. Technological diversity comprises three types of diversity:

- informational diversity, referring to the dissimilarities in knowledge and perspectives between individuals (Jehn, Northcraft, & Neale, 1999)
- functional diversity alluding to functional backgrounds differences between the members of the team (Pelled, Eisenhardt, & Xin, 1999)
- educational diversity (Dahlin et al., 2005)

Both types of diversity engender different mechanisms and effects impacting the novelty of the research conducted by the team.

Demographic diversity is related to the social categorization perspective. When diverse people are brought together to perform a task, it can lead to conflicts. People tend to categorize others in groups based on their differences and similarities. This forms in-groups and out-groups and leads to intergroup biases which are defined as “more favourable perceptions of, and attitudes and behaviour toward, in-group than out- group” (Brewer, 1979 cited in van Knippenberg et al., 2004). Intergroup biases “invite a closing of the mind to dissimilar others, reducing the willingness to share and discuss information and diverse perspectives, as well as a tendency to see diverse others as less trustworthy and knowledgeable sources of information, and thus lead members to pay less attention to diverse viewpoints even if they are shared”(Pieterse, van Knippenberg, & van Dierendonck, 2013, p. 783).These biases trigger conflicts, communication and sharing problems and finally impacts badly on team novelty.

The effect of technological diversity on the novelty of the team is not straightforward. Regarding informational diversity, referring to diversity in knowledge and perspectives, Jehn et al. (1999) showed that increasing it positively impacts group performance because the discovery of new insights by a team is related to the attendance of diverse perspectives.

Indeed, increasing informational diversity inside the team allows the cross-fertilization of ideas by broadening the knowledge stock available to make combinations of existing pieces of knowledge into novel outputs (Y. N. Lee et al., 2015, p. 686).
Moreover, from the individual point of view, working in team leads to great exposure to other beliefs, perspectives, experiences, careers and thus enhances the individual's creativity which in turn increases group creativity (Taylor, Greve, Alva, & Henrich, 2006).

A consequent current of literature has studied and agreed on the fact that team work increases novelty and creativity through higher diversity, this is called the information/decision-making perspective (Landry, Traore, & Godin, 1996; Taylor et al., 2006; Uzzi et al., 2013; van Knippenberg et al., 2004; Woolley, Chabris, Pentland, Hashmi, & Malone, 2010).

Unfortunately, beside this perspective, technological diversity also has drawbacks for novelty. Pelled et al. (1999) discovered that functional diversity increases appearance of task conflicts because the members of the team each rely on their mental structures and beliefs when they perform a task. The more diverse functional backgrounds there are, the harder it gets for the member to agree and to move forward on the tasks. Therefore, functional diversity has an inverted U-shape relationship with novelty meaning that in the beginning, increasing it enhances novelty but only up to a certain point. After this optimal point, functional diversity badly impacts novelty. Educational diversity also has an inverted U-shape relationship with the range and the depth of information use and hence novelty (Dahlin et al., 2005).

The Figure 1 below summarizes the previous paragraph to make the various relationships between diversity and novelty clearer.

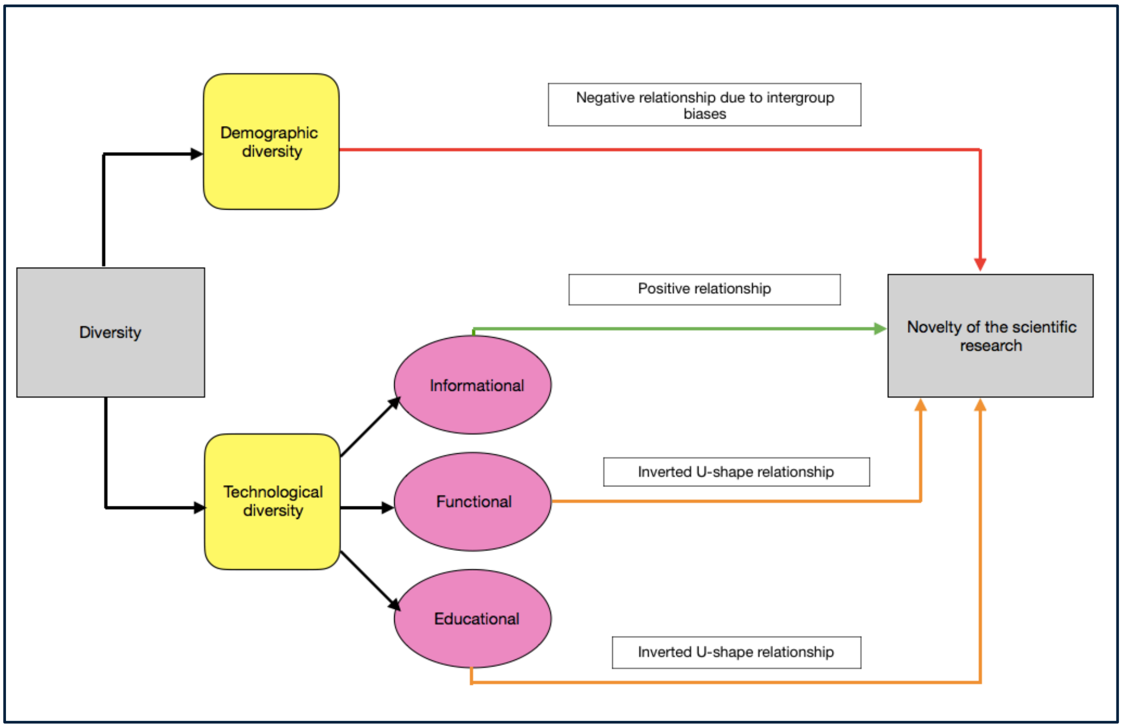

Figure 1: The relationships between diversity and novelty

To conclude, diversity, through the use of teams, is extremely important to boost the novelty of the scientific search. Yet, diversity is a double-edged sword because of the intergroup biases it can create (Pieterse et al., 2013) and the inverted U-shape relationships at stake. Indeed, diversity boosts novelty by having a larger pool of knowledge to make combinations but at some point, adding further diversity to a group will not have positive but well negative effects because of conflicts, misunderstandings, the time required to discuss the various point of views etc.(Chatman, Polzer, Barsade, & Neale, 1998; Y. N. Lee et al., 2015; van Knippenberg et al., 2004).

A remark needs to be done here: throughout the thesis, we argue that knowledge diversity or distance increases novelty, i.e. the less overlapping the knowledge is, the higher chances for novel recombinations. This relation is always positive and refers to informational diversity. The trick is, in order to have higher informational diversity and hence a larger stock of knowledge, scientists need to work in teams. Working in teams increases the diversity of education, specialization, backgrounds and has the side effect to lead to task conflicts and intergroup biases. So, knowledge distance, through informational diversity has a positive linear relationship with novelty while the rest of the team diversity has an inverted U-shape or negative relationship with novelty.

2.2 The gender's impact on novelty

In the previous sections, we described various phenomena impacting novelty of research. Nevertheless, the goal of this paper is to discuss the importance of women in the scientific world and how their presence in research teams can enhance novelty.

As explained above, the driver of combinatorial novelty is the stock of knowledge from which the re-combinations are made and an effective way to extend and share the knowledge is by working inside diverse teams.

Still, working with teams is not sufficient. On the one hand, the knowledge of each team member still needs to be shared and transferred to the others in order to effectively trigger novelty. That's where women intervene. On the other hand, the forming teams need to be composed of people with as distant knowledge as possible in order to effectively increase the knowledge range. Once again, women are likely to play an important role there.

In the next paragraphs, we will explain how women's characteristics enhance the transfer of knowledge and thus extend the knowledge base and enhance novelty. We need to note that the following mechanisms are observed on average and that of course, one individual can differ from another.

2.2.1 Willingness to compete, narcissism

Women and men have divergent personality traits in their willingness to compete, narcissism (Fellner & Maciejovsky, 2007; Niederle & Vesterlund, 2007; Van Vugt et al., 2007).

Regarding competition, men tend to compete more and choose more prestigious and competitive career's tracks than women (Apicella & Dreber, 2015; Buser, Niederle, & Oosterbeek, 2014; Gneezy, Leonard, & List, 2009; Niederle & Vesterlund, 2007). This greater willingness to compete is mainly driven by the overconfidence of men and their higher narcissism (Grijalva et al., 2015). Taking the women's perspective, their smaller competitive attitude can be explained by their lower confidence and higher risk-aversion which push them to opt out from competitive schemes (Fellner & Maciejovsky, 2007; van Veldhuizen, 2016). The strongest incentive of men for competition also has repercussion on how they collaborate. They are more prone to collaborate with a group if this group is in competition with other groups, while collaboration for women is not influenced by intergroup competition (Van Vugt et al., 2007).

Concerning narcissism, Grijalva et al. (2015), found out that the social roles and gender stereotypes can be classified into two main categories: agentic characteristics

(competitiveness, dominance, assertiveness and need for achievement) and communal characteristics (friendliness, tenderness, self-lessness).
Narcissist people, who are more likely to be men, show higher level of agency and are more focused “on getting ahead than getting along socially”(Foster & Brennan, 2011, p. 90). They view scientific research and publication as a contest and they are less likely to invest themselves in the group dynamics or to take actions to enhance peer productivity. Women, on average, are less narcissist and more communal-oriented, focusing on maintaining and strengthening social relationships.

Differences in women’s competitiveness and narcissism can positively impact the exchange and relations inside the team, by extent its knowledge sharing and thus the novelty of the search.

2.2.2 Gender and the big five personality traits

A stream of literature has also studied the gender disparities in the five basic personality traits: openness, extraversion, agreeableness, conscientiousness and neuroticism (Weisberg, Deyoung, & Hirsh, 2011).

Significant gender differences with women scoring higher in extraversion and agreeableness have been discovered (Forrester & Tashchian, 2010; Lehmann, Denissen, Allemand, & Penke, 2013).
Extraversion is divided in two subsequent facets: enthusiasm reflecting sociability, gregariousness and experience of positive emotions and assertiveness reflecting traits linked to agency and dominance. Women are more enthusiast and men are more assertive. Agreeableness corresponds to altruism and qualities such as empathy, kindness, a propensity for cooperation, maintain social harmony and consideration for other people’s concerns. Taken together, extraversion and agreeableness are personality traits positively linked to interpersonal interactions (Berings, De Feyter, Brebels, Van den Broeck, & Proost, 2013a, 2013b; Forrester & Tashchian, 2010; Lehmann et al., 2013).

Given the higher scores of women on agreeableness, they are more likely to cooperate and thus to share their knowledge. This taken together with their bigger enthusiasm make them more sociable and people-oriented: it is important for them to keep good relations inside their groups which can lead to better dynamics and impact group processes.

2.2.3 Women’s team spirit characteristics

Women are more likely to endorse low-promotability tasks, defined as work which is benefiting the organization and the collective but is not improving one’s evaluation and career development (Babcock, Recalde, & Vesterlund, 2017; Babcock, Recalde, Vesterlund, & Weingart, 2017). Low-promotability tasks involve activities such as enhancing the group cohesion, the welfare of others or the working of the organization (De Pater, 2011).

Beside low-promotability tasks, women also endorse more communal goals than men and like people and/or society-oriented work (Diekman, Brown, Johnston, & Clark, 2010).

Communal goals and communion are related to the desire of working with, helping others, by giving material or emotional support (Diekman & Steinberg, 2013). Communal motivation is linked to affiliation, intimacy and altruism (Boucher, Fuesting, Diekman, & Murphy, 2017). Agentic motivations are linked to power, achievement and to traits that promote the interests of the self while communion emphasizes promoting the interests of others. As said earlier, men follow more agentic motivations while women pursue communal motivations.

Being communal goal-oriented is beneficial for the person itself but also for people around as it helps to "smooth social functioning" (Diekman, Steinberg, Brown, Belanger, & Clark, 2017, p. 151).. Pursuing this type of goals helps one to fulfil various needs such as the need for connection, belonging or affiliation. The achievement of those needs leads in turn to constructive events like experiencing greater social support, being more positive emotionally speaking, less subject to stress, more empathetic and collaborative. Those positive events extend to the environment around the person and thus to the workplace and lead to better collaboration. Communal-oriented people have closer contact with colleagues at work, making them more motivated and likely to stick to one job (Boucher et al., 2017; Diekman et al., 2017).
As women tend to be more communal-oriented than men, their presence in teams can bring those beneficial events and improve collaboration.

2.2.4 Pro-social behaviour and knowledge sharing

The paper of Iorio, Labory, and Rentocchini (2017) sheds light on how knowledge sharing is impacted by the different motivations scientists have to effectively engage in those transfers. Particularly, the authors work on the pro-social motivation, which is defined as "the desire for academic scientists to advance the societal role of universities" (Iorio et al., 2017, p. 497). Scientists following this motivation are more altruistic and put the interests of the community first.
As women aspire to careers that can fulfill their social needs, involve interacting with people and that are socially meaningful (Watt et al., 2012), they are more likely to share this pro-social motivation. Indeed, men tend to give more importance to extrinsic rewards such as

payments and promotions, while women grant greater interest to social rewards such as affiliation to a group and good relations with co-workers (Berings et al., 2013a; Mottaz, 1986).
Furthermore, the authors argue that pro-social motivation can lead to more effective knowledge transfers and a higher willingness to enter in knowledge sharing relationships.

Finally, knowledge transfers are better when communication and interactions are more frequent. Trust is also an important factor in knowledge transfer activities (Levin & Cross, 2004). Actually, numerous interactions build social capital, common norms and values which make communications and comprehension simpler and hence, makes knowledge transfers more efficient. As women tend to also communicate and build stronger relationships, their presence in teams can help the transfer of knowledge which broadens the knowledge base of the team.

2.2.5 Communication and leadership

Woolley et al. (2010) found evidence of the existence of a collective intelligence factor defined as “the general ability of the group to perform a wide variety of tasks” (Woolley et al., 2010, p. 687). This collective intelligence factor is significantly and positively correlated with the number of women inside the team. This positive effect of women on collaboration is partly explained by their greater social sensitivity: their ability to perceive and interpret what other members are feeling or thinking during a social interaction.

The collective intelligence factor is also dependent on how people interact within the groups. Group with a more equal repartition of speaking time are more collectively intelligent than groups with one or two members dominating the conversation. Women positively impact the interpersonal communication in groups. They have a more interpersonal oriented style of communication, are warmer and smile more to others and are also collaborative and supportive during communication (Carli, 2010). The presence of women brings a more equal distribution of the conversational turn-taking which allows a better use of individual knowledge and skill resources (Woolley et al., 2010). In the book of Tannen (1990) (as cited in Fenwick & Neal, 2001) women are also identified as communicating through negotiation and empowerment, particularly by turning their opinions into questions and thus sharing information and inviting others to participate. Women try to build relations through communication whereas men use it more as a way to convey information, when useful, to disclose their status and expertise and consequently adopt a different tone and way of communicating.

Women and men also diverge in the way they lead a team with men being more autocratic and women more democratic. Male leaders tend to focus more on their subordinates’ mistakes and penalize poor performance while women are more likely to reward good performance, to ask for the opinions of others and support them (Carli, 2010). The study of Eagly and Johnson (1990) concluded the same results: women being more interpersonally oriented, with higher interests in maintaining good relations and having a democratic leadership style while men are more autocratic and directive.

The presence of women gives rise to a more participative environment than men. People also reported to feel more confident about efficiency of their work in the presence of women (Fenwick & Neal, 2001; C. Lee & Farh, 2004). Those differences in interpersonal communication and leadership explain the positive effect of female presence on group processes and collaborations (Bear & Woolley, 2011), which leads to better intra-group relationships and exchange of information broadening the knowledge base from which recombinations are made to reach novel findings.

2.3 Hypotheses

Figure 2 summarizes the relationships discovered throughout the review of literature. The starting point of our analysis is the team of authors and co-authors of scientific papers. Our econometric analysis is based on panel data and as it is difficult to find teams that stay constant over time, we have decided to fix the unit of study on the network of co-authors around a focal scientist, i.e. the authors with whom a focal scientist has co-written a paper. A deeper explanation of the network will follow in the section 3. In the rest of this thesis, the word "team" thus corresponds to the network of co-authors around a focal scientist.

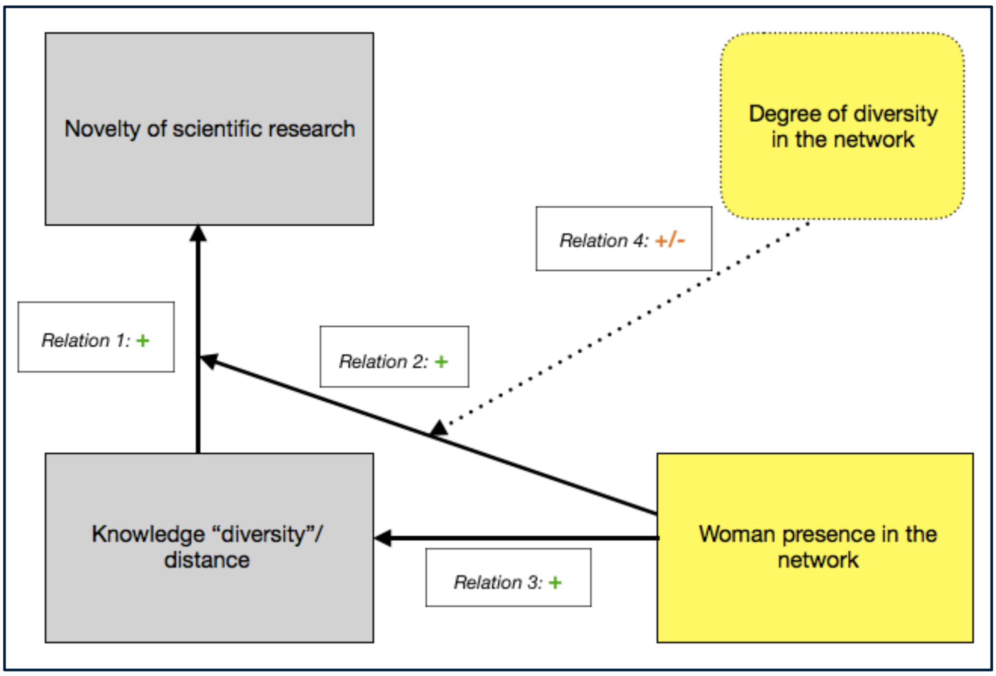

Figure 2: Relations between women, knowledge and novelty

2.3.1 Relation 1: Knowledge distance and novelty of scientific research

The first relationship is the one relating informational diversity, i.e. the width of the knowledge base, and the novelty of the scientific research: the larger the knowledge base, the better chances for novel recombinations. We will not demonstrate this relationship in our analysis as it is already well demonstrated in the current literature (Boudreau, Guinan, Lakhani, & Riedl, 2016; Fallah, 2004; Fershtman & Gandal, 2011; Hansen, Owan, & Pan, 2006; Ruiz-Jimenez, Fuentes-Fuentes, & Ruiz-Arroyo, 2016; Simonton, 2004; Taylor et al., 2006; Uzzi et al., 2013).

2.3.2 Relation 2: woman presence in the network as positive moderator of relation 1

Having women in the co-authorship network positively moderates the first relationship between knowledge diversity and novelty of scientific research. It means that for a given level of knowledge diversity, which impacts novelty, if there are women in the network, the impact will be higher. This relationship is related to various characteristics of women detailed earlier: women impact positively knowledge flows in teams because of skills (communication skills, social skills) and preferences (communal goal endorsement, competition, …).

In other words, women have higher peer effects on the member of their team than men. Peer effects are the productivity (here we concentrate more precisely on the **novelty**) enhancements in the group of peers (here defined as the network of **co-authors**) that are due to the presence and social interactions with the focal individual (here the **women** inside the teams). Peer effects are in fact, indirect productivity effects due to the presence of women in the teams. Therefore, we expect to find in our results a higher novelty in teams with female presence as they are more likely to engender bigger positive peer effects than males.

> ***Hypothesis 1****: Woman presence in the co-authorship network enhances the novelty of the co-authors' researches because of their higher peer effects.*

2.3.3 Relation 3: Woman presence and knowledge distance

The third relationship states that having woman presence in the network increases knowledge diversity.

This relationship is related to the formation of the teams. Indeed, the knowledge distance inside the network depends on the people who are present in it. The more scientists with distant knowledge are working together in the network, the more diverse the knowledge is and the higher the odds are for novel combinations. Therefore, we will measure the degree of knowledge distance between the co-authors around the focal scientist. We expect this degree to be higher when the focal scientist is a woman as we claim that they are better than men at creating relationship ties and communication among members of the team.

Relationships and sharing of ideas are more complicated when the knowledge of the members highly differ because higher informational diversity also means that they have higher various education, specialization which increase the attendance of conflicts. So, we expect individuals with more diverse knowledge, because they think in different ways, to have more difficulties to collaborate then people with similar knowledge. We thus expect that scientists with diverse knowledge are more likely to enter and collaborate inside teams with women because they ease the contact and communication.

> ***Hypothesis 2****: Knowledge diverse networks are more likely to be formed in the presence of women.*

2.3.4 Relation 4: impact of the gender diversity on the first hypothesis

The relationship 4 is related to the degree of other types of diversity than informational diversity inside the network (educational, functional and demographic diversity). Following the theory, we also expect those diversities to impact our results. In the context of this thesis we focus on the gender diversity, which is part of demographic diversity, inside the co-authorship network. The effect of gender diversity inside the network is tricky to predict because two opposite drivers are at play.

On the one hand, as the gender diversity inside the network is growing, the positive incidence of women presence on their peers' novelty and their effect on team formation have higher odds to occur: a single woman in the network is likely to have smaller effects than a few women inside it. So, we expect to have a higher impact on the novelty of the co-authors as the gender diversity inside the network is growing.

On the other hand, as more women are inside the network, the demographic diversity is increasing and there are higher chances for the formation of intergroup biases. So, we expect that the gender diversity inside the network will only impact novelty positively up to a certain point, then intergroup biases will disrupt the relationship.

3 Econometric considerations

In the following pages, we will describe the data, the methodology and regressions used to assess our two hypotheses.

3.1 Data

This section highlights the process through which the data for the econometric analysis were collected as well as the different measurements used in order to test the hypotheses.

3.1.1 Co-authors' network and use of sudden deaths

Our statistical design is based on the use of a co-authors' network defined as the group of scientists that have collaborated, i.e. co-write papers, with a focal scientist during a certain period of time. The reason of this choice is simple: on the one hand, our analysis is based on panel data, in which the core component is the time and the fact that we observe the same entities at different points in time. On the other hand, scientists do not write their papers with the same colleagues all the time, they change their partners. As such, it is not possible to define our unit of observation for the panel data on the couple team-year as we can't observe the same teams of authors over time. To resolve this problem, we have chosen to replace the team by the co-authors' network of scientists. For example, we take the scientist named Mark van Vugt, i.e. a focal scientist, and we define the network around him as the co-authors who have worked with him within a certain period of time. Several authors have based their studies on this type of networks (Azoulay, Zivin, & Wang, 2010; Jaravel, Petkova, & Bell, 2015; Mohnen, 2016). The unit of observation for our analysis is thus the couple co-author-year.

However, it is complicated to extract causal relationships from a network of co-authors because its formation is endogenously determined as the decision to co-author is often strategic and there are numerous unobservable factors that impact a scientist's productivity and his peer's productivity (Mohnen, 2016). We thus use the sudden/premature deaths of focal scientists in the network as a technique bringing an exogenous shock in the network around him leading to randomization. Sudden deaths allow the uncovering of the causal relationship from the focal scientist on his peer's productivity, i.e. his co-authors, and thus certify that the effect we will measure is due to the deceased person (Azoulay, Fons-Rosen, & Graff Zivin, 2015; Azoulay et al., 2010; Fracassi & Tate, 2012; Jaravel et al., 2015; Oettl, 2012). Specifically, we define a death as sudden when it occurs before the scientist's retirement age, meaning that at the time of the death, the focal scientist who passes away, was still active in the scientific publication world and that his passing was sudden as he was still young enough.

We will measure the change in novelty and in knowledge distance of the co-authors, before and after the death of the focal scientist. If there are any differences, it can be attributed to the deceased scientist. Following the literature review we thus expect this difference to be higher when the focal scientist is a woman versus a man.

3.1.2 Description of the treatment and matched control samples

Our econometric analysis is based on a difference-in-difference strategy, we thus need two groups: the treatment sample and the matched control sample.

To collect the data, we start from the articles indexed in PubMed – including most journals in the medical sciences – we compile data on individual scientists using the disambiguation provided by the Torvik Research Group (Torvik, 2015; Torvik & Agrawal, 2016; Torvik & Smalheiser, 2009). To identify sudden deaths, we run a procedure containing several stages. First, using a web scraping algorithm, we extract the first 10 BING search results for each scientist in the database using the scientist’s name and affiliation as a search key. Second, we define a search key to identify potentially deceased scientist and search for entries that might indicate death (using key words such as ‘deceased’, ‘died’, ‘obituary’). Third, for each hit, we manually look up the cause of death and extract sudden/premature deaths. We end up with a sample of scientists that passed away suddenly.

3.1.2.1 Unit of observation

Starting from the sample of sudden deceased scientists, we create the co-authorship network around each of them. Our unit of observation is thus the couple co-author-year. A co-author is defined as someone within the co-authorship network of the focal scientist, i.e. the scientist from which we consider the peer effect is flowing, at time t-10 until t-5, where t is the year of treatment (year of the focal scientist’s death) for the treatment and control group. We follow these co-authors from t-5 up until t+5, and we end up with a panel data-set of these co-authors for the five years before and after the treatment (i.e. the decease of the focal scientist).

3.1.2.2 Treatment and control sample

In the treatment sample are the co-authors (as defined above) of the focal treatment scientist.

A focal treatment scientist is a scientist who:

- dies within the time period 2000-2005

- has published at least 10 articles in total
- has published at least 5 articles in the 5 years before his/her last article
- has stopped publishing between 2000 and 2008
- does not have an affiliation with a non-English speaking country during his/her final 4 years of publishing

For each focal treatment scientist, we consider the population of scientists that has not die as a potential focal control scientist – and assign the treatment year to them. The control group will be restricted by following criteria:

- similar number of publications before the treatment year
- similar career age
- same gender
- similar network size at t-5
- similar gender diversity in their co-authorship network at t-5.

After these procedures, we end up with a set of treatment focal scientists, i.e. set of scientists who passed away a certain year, and a set of control focal scientists who have not passed away but to which we attribute a treatment year. The set of control focal scientists has the same likelihood to be treated – as if they and their co-authorship network are randomized into treatment.

3.1.3 Measure of novelty

In order to test our first hypothesis, we need to measure the novelty of the scientific research of the co-authors.

We use the novelty measures developed by Mishra and Torvik (2016) which are based on the Medical Subject Headings of articles indexed in MEDLINE. MeSH are keywords, not assigned by the authors but by specialised expert librarians from the National Institute of Health, which reflect the content of an article and categorize it. A MeSH corresponds to a concept, each article is related to several MeSH and thus to several concepts. The authors developed measures of novelty based on the age of a concept or a pair of concepts, represented by the age of its MeSH. The age of a concept is measured as the years since the first time it appears in the MeSH bulk. Therefore, the higher the novelty measure is (i.e. the older the concept is), the less novel the article is. The authors have established two types of novelty measure for a paper:

- Time novelty: based on the number of years since a concept's first appearance in the bulk of MeSH

- Volume novelty: based on the number of articles since a concept's first appearance in the bulk of MeSH

This two types or novelty measures are also available for individual concept and pairs of concepts. Pairs of concept's novelty is more related to the combinatorial view of novelty: two concepts cannot be novel but their recombinations in a pair is well novel. Individual concept novelty refers to the appearance of a concept alone.

To attribute time novelty score to articles, we look at the minimum concept age in years among all the article's concepts. We do the same for pairs of concepts to have the time novelty pairs score. Regarding volume novelty, we look at the youngest concept in term of papers, i.e. the concept that have been used in the less papers, and we also do the same for pairs of concepts.
In our database, we calculate those scores for all the articles of a co-author for each year and divide it by the numbers of articles by years to have measures of his average novelty each year. We thus end up, for each year and for each co-author, with 4 measures of his novelty: time novelty, time novelty for pairs, volume novelty, volume novelty for pairs.

3.1.4 Measure of knowledge distance

In order to measure the knowledge distance, we have taken the number of distinct MeSH in a co-author network and divide by the number of co-authors inside this network. This measure gives the knowledge diversity inside the network around a co-author in a certain year.

3.1.5 Description of the variables

Our panel-data set, with as unit of observation the couple co-author-year, will be composed of the following variables:

- By co-author – year (index ***jt***): novelty measures, network knowledge diversity measure, number of co-authors, career age, network gender diversity
- By co-author (index ***j***): gender, gender focal scientist, number of articles written with focal scientist before his death (intensity), time since last article with focal scientist (recency), if co-author and focal scientist were located at the same university (proximity)

A complete description of all the variables is available in the Annex 1.

3.2 Methodology

In this part, we highlight the regressions used to test our two hypotheses.

3.2.1 Test of H1: Woman presence in the co-authorship network enhances the novelty of the co-authors' researches because of their higher peer effects.

For this relation, the argument to prove is the fact that women have higher peer effects on their co-authors than men through higher knowledge flows because of skills (communication skills, social skills) and preferences (communal goal endorsement, competition etc.). When knowledge flows more, likelihood of novelty increases because of higher potential for recombination.

This predicts a positive difference between the peer effect on novelty of women versus men. The measures of novelty are inverted, the greater the measure, the less novel the co-author is. If the theory is true, at least we would expect to find a positive value of β_1 in our first regression, meaning a decrease in the co-authors' novelty after the death of a female focal scientist.

$$(1)\quad novelty_{jt} = \beta_0 + \beta_1\, AFTERDEATHjt * FEMALEFOCALj + \beta_2\, AFTERDEATH_{jt} + \beta_3 careerage_{jt} + \beta_4 genderdiversity_{jt} + \beta_5 nrofcoauthors_{jt} + \alpha_t + \delta_j + \epsilon_{jt}$$

with:

- $novelty_{jt}$: the dependent variable of interest corresponding to the novelty of the output of co-author *j* in year *t*
- $AFTERDEATH_{jt}$: an interaction effect represented by a dummy variable, switching to 1 after focal scientist of co-author *j* passed away in year *t* for the treated sample.
- $FEMALEFOCALj$: dummy variable equals to 1 when the focal scientist of co-author j is a woman and equals to 0 when it's a man
- β_1: coefficient that captures the difference in the causal effect of the focal scientist's death on co-author j, i.e. the change in novelty of co-author j, when the focal scientist is a woman versus man.
- β_2: coefficient that captures the causal effect of the focal scientist's death on co-author j, i.e. the change in novelty of co-author j, when the focal scientist is a man
- β_1+β_2: captures the causal effect of the focal scientist's death co-author j, i.e. the change in novelty of co-author j, when the focal scientist is a woman
- $careerage_{jt}$: controls for the career age of co-author j in year t
- $genderdiversity_{jt}$: controls for the gender ratio in the network of co-author j in year t

- $nrofcoauthors_{jt}$: controls for the number of co-authors in the network of co-author j in year t
- α_t: corresponds to time fixed effects for the 5 years before and after the death of the focal scientist
- δ_j: corresponds to the individual fixed effects such as education, motivation, ability etc
- ϵ_{jt}: corresponds to the error term

3.2.1.1 Moderating effect of the ties strength

As stated earlier, the strength of the ties between the co-authors and the focal scientist can impact our results (Carrington, Scott, & Wasserman, 2009). Indeed, our argument for hypothesis 1 is likely to be more pronounced for collaborators that have work ‘closely’ together, where ‘closely’ refers to the intensity, recency and proximity (the co-author and the focal scientist are affiliated with the same university) of collaboration.

Co-authors that had intense collaboration relationships, in the sense that they have collaborated at least a certain number of times with the focal scientist, will be more affected by his passing and we expect the treatment effect to be stronger when the number of collaborations increases. Indeed, an intense collaboration increases trust among members which improves the sharing of knowledge (G. Jones, 2017; Levin & Cross, 2004). The members also know better the way of working of the other and the exchange of ideas is more efficient.

Regarding recency, the later the collaboration occurs before the scientist passes away, the stronger treatment effect we expect.

Finally, for proximity, we expect the treatment effect to be stronger for co-authors collocated with the focal scientist. Indeed, the positive effects of women appear more during direct social interactions and exchanges and those types of communications are more likely to occur if the authors are collocated inside the same university. If the co-authors are not collocated, there are higher chances that they communicate via email and with this sort of interpersonal connection, the peer effects of women are less probable.

For each of these dimensions (intensity, recency and proximity of collaboration), we run an interaction between the main variable and the measures for tie strengths. Expected is that β_1 from regression 1 increases with tie strength – which suggests that the argument that females make knowledge flows more easily is supported, because knowledge flows more easily between strong ties. The regressions are the following:

$$(2)\ novelty_{jt} = \beta_0 + \beta_1\, AFTERDEATHjt * FEMALEFOCALj * RECENCYj + \beta_2\, AFTERDEATHjt * FEMALEFOCALj + \beta_3\, AFTERDEATH_{jt} + \beta_4 careerage_{jt} + \beta_5\, genderdiversity_{jt} + \beta_6 nrofcoauthors_{jt} + \alpha_t + \delta_j + \epsilon_{jt}$$

$$(3)\ novelty_{jt} = \beta_0 + \beta_1\, AFTERDEATHjt * FEMALEFOCALj * INTENSITYj + \beta_2\, AFTERDEATHjt * FEMALEFOCALj + \beta_3\, AFTERDEATH_{jt} + \beta_4 careerage_{jt} + \beta_5\, genderdiversity_{jt} + \beta_6 nrofcoauthors_{jt} + \alpha_t + \delta_j + \epsilon_{jt}$$

$$(4)\ novelty_{jt} = \beta_0 + \beta_1\, AFTERDEATHjt * FEMALEFOCALj * PROXIMITYj + \beta_2\, AFTERDEATHjt * FEMALEFOCALj + \beta_3\, AFTERDEATH_{jt} + \beta_4 careerage_{jt} + \beta_5\, genderdiversity_{jt} + \beta_6 nrofcoauthors_{jt} + \alpha_t + \delta_j + \epsilon_{jt}$$

with:

- $RECENCYj$: dummy variable that switches to 1 for co-authors who have published with focal scientist during the 2 years before his death
- $INTENSITYj$: dummy variable that switches to 1 for co-authors who have published 2 times or more with focal scientist before his death
- $PROXIMITYj$: dummy variable that switches to 1 for co-authors who are from the same university as the focal scientist
- β_1: coefficient that captures the difference in causal effect of the focal scientist's death on co-author j, i.e. the change in novelty of co-author j, when the focal scientist is a woman versus man and when the collaboration was either recent, close or in proximity.

3.2.2 Test of H2: woman presence positively impacts knowledge diversity inside the network

Basically, there are two main mechanisms in our review of literature corresponding to our two hypotheses. On the one hand, women might play an important role in forming ties, this mechanism corresponds to our second hypothesis, i.e. women increase the attendance of knowledge diverse networks. On the other hand, once ties are formed, women help to increase co-ordination, knowledge transfers and novelty, this is the main mechanism, corresponding to our first hypothesis explained above.

In order to disentangle these, we check whether the main effect is mainly explained by differing networks after the decease – so knowledge diversity of the network of co-authors decreases after the focal scientist's death. If knowledge diversity, measured for each year, decreases more when females die and if controlling for knowledge diversity decreases the main effect substantially, i.e. decrease the value of β_1 in the first regression, then the ties formation mechanism is, or partly is, responsible for the main effect of women on novelty. The extent to which the main effect stays once controlling for knowledge diversity, might give an indication for the presence of the main mechanism – that is, females increase knowledge flows within co-author teams.

3.2.2.1 Women and knowledge diversity

To measure it, we need to first calculate the change in knowledge diversity around each co-author j before and after the death of the focal scientist. On the contrary of novelty measures, knowledge diversity measure is not inverted. We expect a higher drop in knowledge diversity of co-author j when the focal scientist is a woman versus a man.

$$(5)\ knowledgediversity_{jt} = \widehat{\beta_0} + \widehat{\beta_1}\ AFTERDEATHjt * FEMALEFOCALj + \widehat{\beta_2}\ AFTERDEATH_{jt} + \widehat{\beta_3} careerage_{jt} + \widehat{\beta_4} genderdiversity_{jt} + \alpha_t + \delta_j + +\epsilon_{jt}$$

With all the variables corresponding to the same description as in the main regression but regarding knowledge diversity instead of novelty.

3.2.2.2 Disambiguation of the two hypotheses

Once we have the results for regression number 5 and if they show well a higher drop in knowledge diversity of the co-authors when the focal scientist is a woman (negative $\widehat{\beta_1}$), then we need to control for it in the main regression in order to see if the ties formation mechanism is responsible for the main mechanism, and if yes, to which extent.

$$(6)\ novelty_{jt} = \beta_0 + \beta_1\ AFTERDEATHjt * FEMALEFOCALj + \beta_2\ AFTERDEATH_{jt} + \beta_3 age_{jt} + B_4\ genderdiversity_{jt} + \beta_5 nrofcoauthors_{jt} + \beta_6\ knowledgediversity_{jt} + \alpha_t + \delta_j + \epsilon_{jt}$$

We expect $\beta_6\ in\ regression\ 6$ to be smaller than 0 meaning that $knowledgediversity_{jt}$ positively impacts $novelty_{jt}$. We also expect $\widehat{\beta_1}\ in\ regression\ 5$ to be smaller than 0 meaning that the death of a female focal scientist negatively impacts $knowledgediversity_{jt}$.

If those two expectations are true, then we also expect $\beta_1\ in\ regression\ 6$ to be smaller than $\beta_1\ in\ regression\ 1$ because part of the main effect (women increasing novelty) is explained by the tie formation theory and not only by the higher coordination and knowledge flow theory.

3.3 Ruling out alternative explanations through design

3.3.1 Women select themselves into more novel projects

It could be possible that women choose more novel projects than men. This hypothesis relies on the fact that women are more likely to endorse communal goals and to be motivated by moving the society and/or the research forward than men do (Diekman et al., 2010; Iorio et al., 2017). Novel researches have greater odds to have a big impact and really change the field of scientific research. The argument can be further supported by the fact that men, as they are more competitive, tend to focus on incremental searches that have higher chances to lead to lots of publications in scientific journals (Apicella & Dreber, 2015; Buser et al., 2014; Gneezy et al., 2009). So, you can either count your citations and publications, what men tend to do more, or you can engage in researches that have the potential to really move the field forward, what women are more likely to do.

On the other hand, the hypothesis can be weakened by the fact that women are more risk-averse than men are (Fellner & Maciejovsky, 2007; van Veldhuizen, 2016). As novel projects are riskier and takes more time to be recognised (Arrow, 2008), women maybe choose less novel projects than men.

In any case, our design allows us to rule out this alternative explanation through the use of sudden deaths that bring an exogenous shock in the network around the focal scientist leading to randomization and thus certifies that the effect we will measure is due to the person itself.

3.3.2 Males with interests in novel researches and/or specific attributes chose more mix-gendered teams

A second potential explanation could be the fact that men that are more likely to engage in novel projects are also more likely to choose gender-mixed teams and to work with women. If this is the case, the higher novelty measured will not be due to the presence of women and their peer effects but to the fact that men that are working on novel projects also work with more women. Or, it could also be possible that men that are more communicative, have higher ability, better education or are more motivated choose to work with women too. As we concentrate on combinatorial novelty, if men with better communicative skills are more likely to work with women in teams, the higher novelty of the research will be due to the presence of those men and will not be the consequence of the women' peer effects.

However, our design also rules out this explanation. The combined use of individual fixed effects and of sudden deaths leads to randomization of the treatment and controls for differences in individual capacities and thus excludes this potential alternate explanation.

3.3.3 Women with higher ability and/or better communicative skills choose more novel projects.

Some women may be more likely to work on novel projects than other women. Women that are more motivated, have better education, more motivation, higher ability or communicative skills may be more likely to work on novel projects. As such, the effect we will measure will not be due to the fact that women, in general, have higher peer effects.

The randomization through the use of sudden deaths and the utilization of individual fixed effects controls for this alternative explanation.

4 Results

The result section describes the outcomes of the statistical analysis using the data defined previously. This section is composed of three parts. The first comprises the data's descriptive statistics. The second part contains the results for our regressions in order to draw conclusions about our two hypotheses. The last part contains the discussion of the results: what are the big trends that we can extract from the data.

4.1 Descriptive statistics

We divide our descriptive statistic into three groups: one for dataset, one for time-variant measures and one for time-invariant measures.

4.1.1 Dataset

The final dataset is composed of 1,222,1998 observations corresponding to 167,250 co-authors arranged in networks around 6,124 focal scientists. In total, there are 355 female and 5,769 male focal scientists. Regarding the treatment and control sample, we have a treatment sample composed of 57,380 co-authors and a control matched sample composed of 1,091,387 co-authors. Finally, the mean age for the premature death of a focal scientist is 56 years old. We only kept the premature/sudden deaths in our dataset, i.e. deceases before the retirement age fixed at 65 years old.

4.1.2 Time-variant measures

Time-variant measures correspond to the different novelty measures, the knowledge diversity measure, the gender diversity inside the network, the number of co-authors and the career age.

The average career age of the co-authors is 20 years and the average number of co-authors is 20. Regarding gender diversity inside the co-authors' networks, the mean is 0.3071, implying that on average women account for one third of the scientists. The novelty scores using pairs of concepts are lower (lower score means higher novelty) compared to the scores with individual concept. This matches with the view that novelty is more a combinatorial process, discoveries of totally new concepts and thus new MeSH are rare. The complete descriptive statistics for those variables are available in table 1 in Appendix 2. For descriptive statistics, separated by treatment and control group and before/after the treatment, see tables 2 to 5 in Annex 2.

We conducted t-tests and found significant differences for all time-variant variables, except for the pair volume novelty measure, between the treatment and the control sample (see table 7 Appendix 2). First, we took a closer look to differences and saw that even though the confidence intervals did not overlap they were very close and that the difference in means were also small. We then plotted those variables by group (control or treatment sample) on graphs for the five years before the treatment, i.e. the death of the focal scientist, to observe time-trends (see Appendix 3 for the graphs). We found parallel time trends for all of them expect for the pair volume novelty. As the time-trends are parallel, it means that the control and treatment samples evolve in the same manner and that the use of fixed effects will control for the differences between them. Regarding the measure of volume novelty using pairs, it is the only one that does not have parallel time-trends. Insights on the reason why it is so will later be developed in this section. The non-significance of the tests may also be related to the fact that the control and treatment sample are not the same size.

4.1.3 Time-invariant measures

Recency, proximity and intensity are used to determine the extent to which the ties between the focal scientists and his co-authors are strong. Nevertheless, those measures do not vary over time. On average, the last collaboration with the focal scientist takes place 5 years before his passing and the co-authors write on average 4 papers with the focal scientist before his death (complete statistics in table 7 Appendix 2). Running t-tests, we found significant differences between the treatment and control sample except for the measure of the recent intensity, i.e. number of collaborations within the 10 years before the passing. Still, the differences in means between the control and treatment samples are small, so are the standard errors, meaning the small differences should not impact the results (see table 8 Appendix 2).

4.2 Regressions results

In this part, we will describe the results of our various regressions and confirm or invalidate our two hypotheses.

4.2.1 Results for the first hypothesis: women enhance the novelty of their peer's researches

The results of the first regression described just below with the 4 measures of novelty as dependent variables are summarized in the tables 9 to 12 in Appendix 4.

$$novelty_{jt} = \beta_0 + \beta_1 \, AFTERDEATHjt * FEMALEFOCALj + \beta_2 \, AFTERDEATH_{jt} + \beta_3 careerage_{jt} + \beta_4 genderdiversity_{jt} + \beta_5 nrofcoauthors_{jt} + \alpha_t + \delta_j + + \epsilon_{jt}$$

4.2.1.1 *With average time novelty as dependent variables (for individual and pairs of concepts)*

The regression using the average time novelty as dependent variable gives significant results (see table 9 Appendix 4). The coefficient β_1 equals 0.785 (p-value < 0.01). The coefficient β_2 equals -0.191 but with a p-value of 0.11. The coefficient of β_1 captures the difference in the novelty of the co-author when a female passes away versus a man and its positive value confirms our hypothesis. It is important to remember that the measure of novelty we are using is "inverted": the smaller the value of the measure, the greater the novelty. What we expected was a greater drop in the novelty when a female focal scientist passes away compared to when a man passes away. As β_1 is positive and significant it means that the disappearance of a female focal scientist entails a drop in the novelty of her co-authors as we predicted: when a woman passes compared to when a man dies, there is a difference in the novelty drop of 0.783 (the novelty measure increases, meaning that the novelty of the co-authors decreases).

The control variable career age's coefficient equals 0.937 (p-value< 0.01), implying that the longer the co-author is working, the less novel output he produces. The number of co-authors (coefficient equals -0.014) has a small but significant effect on the novelty (p-value< 0.01). The more co-authors there are in the network, the more novel the output will be. This is consistent with the argument developed in the literature review that the larger the knowledge base is, the greater chances for novel output. Finally, the gender diversity inside the network also has a coefficient of -1.259 (p-value < 0.01). It means that the higher the gender diversity inside the network, the more novel the research will be.

The measure of average time novelty with pairs of concepts gives approximately the same results as with individual concept. The coefficient of β_1 equals 0.267 (p-value< 0.05), so the passing of female focal scientist stills entails a drop in novelty compared to when a male dies. β_2, the effect of a male scientist dying, equals -0.023 but is not significant. Career age decreases the novelty of the co-authors while the number of co-authors and the gender diversity both enhance novelty (negative scores) (p-values< 0.01).

4.2.1.2 *With volume novelty scores as dependent variables*

The results using volume novelty as dependent variable are less straightforward. The average volume novelty score for individual concept gives the same trends, in the coefficients of β_1 and β_2 but both are non-significant. The career age, number of co-authors and gender diversity inside the network all have the same effects as with time novelty scores (p-values< 0.01). The results with average volume novelty for pairs of concepts are not following the same trends for β_1 and β_2 but those results are not-significant. Career age also has a non-significant coefficient. Number of co-authors (p-value< 0.01) and gender diversity (p-value< 0.05) both have negative coefficients meaning they still enhance the co-authors' novelty.

The non-significant results using volume novelty as dependent variables (both for individual and pairs of concepts) can be linked to the fact that those measures are less stable than the time novelty measures. Indeed, time novelty looks at the younger MeSH term of a paper, with the age of a MeSH term defined as the number of years since its first appearance in the corpus of MeSH terms. Therefore, once the MeSH term is born, there are no trends inside the measure, each year passing just add another year to the age of the MeSH term. In other words, time novelty is a linear measure. But for the volume novelty measures, the age depends on the number of papers using the MeSH, the more papers are using the term, the older the MeSH gets. The problem is that there are trends in the use of novel concepts, for example if the term becomes trendy at one point in time. Mishra and Torvik (2016) have identified 4 phases in the publication profile of a new MeSH: burn-in phase, accelerating growth phase, decelerating growth phase and constant growth phase. The measures using volume novelty scores are thus less stable and for this reason we prefer the time novelty measure.

4.2.2 Results for the moderating effect of ties strength

We have decided to perform the analyses for our preferred measures (individual and pair time novelty) for the reasons explained above. We ran the following regressions for proximity, recency and intensity measures.

$$\begin{aligned} novelty_{jt} = {} & \beta_0 + \beta_1\, AFTERDEATHjt * FEMALEFOCALj * RECENCYj \\ & + \beta_2\, AFTERDEATHjt * FEMALEFOCALj + \beta_3\, AFTERDEATH_{jt} \\ & + \beta_4 careerage_{jt} + \beta_5\, genderdiversity_{jt} + \beta_6 nrofcoauthors_{jt} + \alpha_t \\ & + \delta_j + \epsilon_{jt} \end{aligned}$$

4.2.2.1 Recency

Before running the regressions, we need to create a dummy variable switching to one when the recency, i.e. the number of years since the last article the co-author has written with the focal scientist, is below a certain number of years. We consider the collaboration to be recent when it happens maximum 4 years before the focal scientist passes away. The results for both individual and pair time novelty scores are summarized in tables 13 and 14 in Appendix 5.

For individual time novelty score, the recent collaboration with women enhances even more the peer effect of women as we predicted. β_1 equals 1.104 (p-value< 0.05) and β_2 equals 0.341 (p-value< 0.05), the peer effect of women is thus higher when the collaborations are recent as the drop in novelty when a female focal scientist passes away versus a man is greater when the collaboration happened maximum 4 years before the decease, i.e. β_1 in this regression is bigger than β_2 . For pair of concepts time novelty score, the trends are similar but the coefficient is not significant.

4.2.2.2 Proximity

The proximity measure is a dummy variable switching to one if at least one of the co-authored publications with the focal scientist was at the same university. As it was not possible to find the universities, we use as a proxy the cities where the co-authors and the focal scientists were located. The results for individual time novelty are not conclusive. β_1 equals -0.888 (p-value< 0.1), i.e. the proximity of the collaboration does not enhance the peer effect of women. The results are the same using time novelty for pairs of concepts, β_1 equals -0.39 (p-value< 0.1).

Those results have to be taken carefully since measuring the proximity was complicated and we were only able to use a proxy. Furthermore, the dummy variable is switching to one if the co-author and the focal scientist are collocated for one publication, but we expect the effect of proximity to appear when the people are working together for a certain time (complete results in tables 15 and 16 in Appendix 5).

4.2.2.3 Intensity

Intensity is split in three measures:

- Intensity: total number of collaborations with the focal scientist before his death
- Old intensity: number of collaborations with the focal scientist for the years before t-10, with t= the year of the death
- Recent intensity: number of collaborations with the focal scientist for the 10 years before his death, i.e. after t-10

We thus create 3 dummies that are switching to one for each of the intensity measure when the co-author and focal scientist have collaborated more than 4 times. None of the coefficients for the 3 measures had a significant p-value and therefore we can't conclude about the effect of the collaboration intensity (results in table 17 to 22 in Appendix 5).

4.2.3 Results for the second hypothesis: Knowledge diverse networks are more likely to be formed in the presence of women

We first ran the following regression to see if women impact network knowledge diversity.

$$knowledgediversity_{jt} = \widehat{\beta_0} + \widehat{\beta_1}\, AFTERDEATHjt * FEMALEFOCALj + \widehat{\beta_2}\, AFTERDEATH_{jt} + \widehat{\beta_3} careerage_{jt} + \widehat{\beta_4} genderdiversity_{jt} + \alpha_t + \delta_j + +\epsilon_{jt}$$

We expect the measure of knowledge diversity to drop further when a female focal scientist passes away versus when a man dies because we think women are better at creating ties between knowledge diverse persons. It is important to note that on the contrary of novelty scores, a higher knowledge diversity is better, while a higher novelty score means a lower novelty of the co-author. We thus expect $\widehat{\beta_1}$ to be negative, as the death of a female focal scientist decreases the attendance of knowledge diverse networks. $\widehat{\beta_1}$ equals -1.692, well negative as we predicted, but is non-significant. We can thus not conclude that women impact knowledge diversity.

The coefficient for gender diversity equals -103.177 and is significant (p-value< 0.01), meaning it decreases the knowledge diversity, i.e. the more gender diverse the network is, the less knowledge diversity. The career age also has a negative coefficient of –9.195 (p-value< 0.01), the longer the co-author has been working, the less knowledge diverse his networks is (complete results table 23 in Appendix 6).

Even though we can't conclude if women really engender more knowledge diverse networks, we are still interested in the effect of knowledge diversity on novelty. We run the following equation on the four measures of novelty (results in tables 24 till 27 in Appendix 6):

$$novelty_{jt} = \beta_0 + \beta_1\, AFTERDEATHjt * FEMALEFOCALj + \beta_2\, AFTERDEATH_{jt} + \beta_3 careerage_{jt} + \beta_4\, genderdiversity_{jt} + \beta_5 nrofcoauthors_{jt} + \beta_6\, knowledgediversity_{jt} + \alpha_t + \delta_j + \epsilon_{jt}$$

The knowledge diversity slightly enhances the time novelty for both individual and pair of concepts (p-value< 0.01). Regarding volume novelty, for individual concept, the knowledge diversity slightly decreases the novelty of the research (positive coefficient). For pairs of concepts with volume novelty, the coefficients are non-significant.

5 General Conclusion

We have started this thesis by noticing an important gender gap still prevailing in academia. We also observe that novelty of the research is nowadays crucial but that important biases against novelty still exist through the funding schemes and bibliometric indicators such as the impact factor. Throughout the review of literature, we have learned about various characteristics of women: communication and leadership skills, communal-goal endorsement, pro-social behavior, personality traits, … These various characteristics are likely to affect the novelty of the research and we think that women are likely to enhance their co-authors' novelty through higher peer effects compared to men. We have tested our hypothesis through a difference-in-difference design.

5.1 Discussion of results

This part highlights the bigger trends that we have uncovered from our dataset. Our first hypothesis that women enhance the novelty of their co-authors through higher peer effects is confirmed for the measure of time novelty for both individual and pairs concepts. Women thus well enhance the novelty of the people they are working with. For both volume novelty measures, the coefficients were not significant. It is likely to be due to the fact that the volume novelty measures are less stable measures than the time novelty ones.

The effects of the ties strengths are not conclusive, except for the recency of the collaboration. The positive effect of women on their peers' novelty is thus well higher when they have collaborated recently, which is not surprising as the death of the focal scientist is more likely to have higher repercussions on the co-authors when they are in closer relationships.

Our second hypothesis, i.e. women enhancing the attendance of knowledge diverse networks, is not confirmed because the coefficient of the women's impact on the knowledge diversity is not significant. Though, it is important to discuss the effect of knowledge diversity on novelty. For the individual and pairs of concepts time novelty scores, the knowledge diversity enhances the novelty of the co-authors. It is consistent with the argument developed in our review of literature that the broader the knowledge base, the higher chances for novel recombinations and thus novel outputs. For volume novelty measures, the results were not conclusive or not significant.

Throughout the different regressions we have also uncovered interesting trends related to the following control variables: the career age of the co-authors, the number of the co-authors within the network and the gender diversity within the network.

We have observed that the co-author's career age badly impacts both its novelty and its knowledge diversity. It means that the longer a co-author has been working, the less novel his output is and the less knowledge diverse his network is. Packalen and Bhattacharya (2015a) argue that younger scientists are more likely to build on new ideas even though on the on the other hand, the best research teams also need an experimented scientist. There is an optimal combination of young scientists with some experimented ones to achieve higher level of novelty. Older scientists may have less incentives to engage in novel researches as they know better that their chances of success are very small. They also have less time to invest in very intensive researches because they are asked various other types of activities (committee work, mentorship, advising etc.). The lower network knowledge diversity of older scientists can be explained by the fact that they have specialized throughout their career and collaborate with more similar scholars.

The co-authors' novelty is slightly impacted by the number of co-authors inside their network. The larger their networks are, i.e. the higher the number of co-authors inside it, the more novel their publications are. The coefficients are small but significant across the different regressions.

Finally, gender diversity also impacts both novelty and knowledge diversity. The novelty is positively affected by gender diversity. Therefore, increasing the share of women in scientific research enhances novelty as we argue in the review of literature.
However, the gender diversity negatively affects the knowledge diversity inside the network. It highlights the complicated nature of diversity discussed in the literature. Indeed, gender diversity positively impacts novelty but negatively impacts knowledge diversity. So, increasing the share of women enhances the novelty of the research due to their different skills and characteristics. At the same time, increasing it also boosts demographic diversity which rises intergroup biases and can have negative effects such as decreasing the knowledge diversity. The optimal level of gender diversity is thus difficult to determine but given the fact that women do well enhance novelty of their peers, which is the most important, we think that increasing their presence in academia is beneficial.

5.2 Recommendations

Based on our findings we have suggestions for universities and science policy makers. Our study suggests universities to investigate further in order to understand the reasons of the gender gap still existing in academia and to take actions to close, or at least reduce it. We advise universities to engage and promote the entry of women in academia and their research teams as they boost novelty of their peers and thus of the research. We also advise both universities and funding agencies to finance research teams with young scientists as the earlier in their career stage, the more likely they are to produce novel outputs. Taking a broader perspective, we also suggest to funding agencies to invest in novel projects even though there are riskier because they are more likely to really impact the research and to lead to great innovations. We also advise universities to conduct team building activities as the novelty arises from the exchange of ideas and knowledge flows. It is not sufficient to have knowledge diverse teams, there is still the need to exchange those ideas and knowledge.

5.3 Limitations and suggestions

First of all, our research is limited to the biomedical sector as we have extracted our data from MEDLINE. It would be very interesting to extend the research to other scientific sectors in order to see if the women peer's effect on novelty is consistent across the fields.

Secondly, some of the measures we use as variables can be further improved. Having a better measure of proximity for the study of the effect of the ties strengths could give interesting results. Another measure for the knowledge diversity could also lead to more accurate evaluation of the women effect on it. A more precise measure that we were not able to perform on our data, is developed in Appendix 7. Further studies implementing this measure could uncover more precisely relations between knowledge diversity and women and the impact of gender diversity on it.

Thirdly, futures researches should develop in more details the matching between the control and the treatment group.

Finally, our study is limited to the research in academia, it could be interesting to extend it to innovation teams in companies by measuring the novelty of the patents developed by those teams and see if female presence also enhances it. Indeed, the various skills and characteristics of women improving novelty of the research are likely to also positively affect innovation.

Appendices

<u>Appendix 1: Description of the variables contained in the database</u>

- **rank**: identifier for a scientist. This means that the person was a collaborator of a deceases/control scientist at t-10 until t-5, where t is the year of treatment. Note that also the control group is a assigned a year of treatment, because that is the year in which we deem the focal scientist to be comparable to one of our dead scientists.
- **year**: a (calendar) year in which the scientist published at least one paper
- treated: dummy for being treated (having collaborating focal scientist who died)
- **year_treat**: the year in which the focal scientist died (or based on which it was selected to be a control
- **focrank**: identifier of the focal scientist
- **focgenni:** gender of focal scientist
- **focfemale**: dummy variable switching to 1 if focal scientist is female, missing means that we don't know
- **matched_to_focrank**: tells you, for the control sample, the rank of the deceased focal scientist with whom the focal scientist was matched. for the treated sample, matched_to_focrank = focrank
- **age_dead:** the age at which focal scientist died, if we know, for the treated sample
- **timenovelty_avg**: average over the year the scientist published in of what is called the timenovelty of its publications. It tells you for a publication how old the youngest MeSH-term is in years
- **timenovelty_sum**: sum instead of average of the above
- **volumenovelty_avg:** average over the year the scientist published in of what is called the volumenovelty of its publications. It tells you for a publication how old the youngest MeSH-term is in terms of how many times it is used before with 1 meaning it is the first time it is used.
- **volumenovelty_sum**: sum instead of the average above.
- **pair[...]:** The four above measures are also given when looking at combinations of MeSH-terms instead of MeSH-terms themselves.
- **careerage:** the time in years since the scientist's first publication
- **nr_coauthors**: number of co-authors in this year in the network of co-author j
- **genderdiversity**: number of female coauthors divided by total co-authors
- **recency:** number of years passed from the treatment year since the last co-authored publication with the focal scientist

- **proximity**: whether at least one of the co-authored publications with the focal was at the same university (this is a proxy, as we use the city they worked in)
- **intensity:** total number of collaborations before treatment year with the focal scientist
- **intensity_recent**: the above for only the last 10 years before treatment
- **intensity_old**: the above for the years before t-10, with t= year of treatment
- **networkdiversity:** the number of distinct MeSH-terms on publications of any co-author
- **knowledge_div:** standardized version of networkdiversity so networkdiversity/nr_coauthors
- **after:** dummy variable switching to 1 for the years after the treatment
- **after_X_treated:** dummy variable switching to 1 for co-author j for the years after the treatment and if co-author j is from the treatment sample (corresponds to afterdeath in the regressions)
- **after_X_treated_X_female:** dummy variable switching to one for the years after the treatment for the treatment sample if the focal scientist is a woman

Appendix 2: Tables for descriptive statistics

Table 1: Summary statistics time-variant measures

Variable	Observations	Mean	Std. Dev.	Min	Max
Time novelty	1,113,094	21.67218	10.27683	0	100
Pair time novelty	1,112,551	2.712507	4.079196	0	87
Volume novelty	1,113,094	3584.678	6562.699	1	1261170
Pair volume novelty	1,112,551	22.69404	569.278	1	144157
Knowledge diversity	1,110,648	242.9288	111.2958	0	2430
Gender diversity	1,105,101	.3070886	.2078829	0	1
Career age	1,122,248	20.70772	11.27968	0	102
Number of co-authors	1,122,248	20.83874	29.7796	0	1375

Table 2: Summary statistics for time-variant measure for the control sample before the treatment (treated=0 and after=0)

Variable	Obs	Mean	Std. Dev.	Min	Max
Time novelty	650,884	20.48728	9.982396	0	100
Pair time novelty	650,465	2.435215	3.892156	0	87
Volume novelty	650,884	3140.694	5865.133	1	1261170
Pair volume novelty	650,465	20.72035	510.0308	1	124707
Knowledge diversity	649,920	255.5535	113.735	0	2034
Gender diversity	646,400	.2904136	.2068165	0	1
Career age	656,339	18.51804	11.18077	0	97
Number of co-authors	656,339	18.73092	24.90698	0	861

Table 3: Summary statistics for time-variant measure for the control sample after the treatment (treated=0 and after=1)

Variable	Obs	Mean	Std. Dev.	Min	Max
Time novelty	431,656	23.50627	10.43917	0	73
Pair time novelty	431,539	3.142906	4.323775	0	63
Volume novelty	431,656	4258.616	7449.712	1	1031724
Pair volume novelty	431,539	25.5445	651.6655	1	144157
Knowledge diversity	430,282	224.7714	104.8456	0	2430
Gender diversity	428,440	.3318068	.2062931	0	1
Career age	435,048	24.09239	10.5984	6	102
Number of co-authors	435,048	24.08955	35.70511	0	1375

Table 4: Summary statistics for time-variant measure for the treatment sample before the treatment (treated=1 and after=0)

Variable	Obs	Mean	Std. Dev.	Min	Max
Time novelty	18,579	19.8449	10.15907	0	60
Pair time novelty	18,572	2.249696	3.695243	0	55
Volume novelty	18,579	3056.844	5635.715	1	362043.7
Pair volume novelty	18,572	21.73595	374.6412	1	20959.8
Knowledge diversity	18,558	240.7666	112.5344	0	1212
Gender diversity	18,434	.2999916	.2182767	0	1
Career age	18,783	17.27014	10.78069	0	83
Number of co-authors	18,783	17.56056	22.67692	0	795

Table 5: Summary statistics for time-variant measure for the treatment sample after the treatment (treated=1 and after=1)

Variable	Obs	Mean	Std. Dev.	Min	Max
Time novelty	11,975	22.79777	10.40892	0	63
Pair time novelty	11,975	2.982198	4.115938	0	57
Volume novelty	11,975	4242.627	7473.165	2	290395.7
Pair volume novelty	11,975	28.66712	676.0653	1	59040
Knowledge diversity	11,888	213.3074	104.3012	0	1484
Gender diversity	11,827	.3340877	.2149614	0	1
Career age	12,078	23.12908	10.22508	6	77
Number of co-authors	12,078	23.38558	36.39573	0	729

Table 6: T-tests for time-variant measures

T-test for average time novelty

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
0	650,894	20.48755	.012373	9.982292	20.46329	20.5118
1	18,575	19.84944	.0745599	10.16178	19.70329	19.99558
combined	669,469	20.46984	.0122069	9.987858	20.44592	20.49377
diff		.6381072	.0743181		.4924462	.7837683

```
    diff = mean(0) - mean(1)                                      t =   8.5862
Ho: diff = 0                                     degrees of freedom =   669467

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

T-test for average volume novelty

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
0	650,894	3140.868	7.274444	5868.876	3126.61	3155.126
1	18,575	3057.506	41.34749	5635.254	2976.461	3138.551
combined	669,469	3138.555	7.165062	5862.531	3124.512	3152.598
diff		83.36246	43.62445		-2.140053	168.865

diff = mean(0) - mean(1)   t = 1.9109
Ho: diff = 0   degrees of freedom = 669467

Ha: diff < 0	Ha: diff != 0	Ha: diff > 0
Pr(T < t) = 0.9720	Pr(|T| > |t|) = 0.0560	Pr(T > t) = 0.0280

T-test for average time novelty (pairs)

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
0	650,475	2.43507	.0048256	3.891964	2.425612	2.444528
1	18,568	2.248675	.0271176	3.695164	2.195522	2.301828
combined	669,043	2.429897	.0047518	3.886755	2.420583	2.43921
diff		.1863946	.028927		.1296987	.2430906

diff = mean(0) - mean(1)   t = 6.4436
Ho: diff = 0   degrees of freedom = 669041

Ha: diff < 0	Ha: diff != 0	Ha: diff > 0
Pr(T < t) = 1.0000	Pr(|T| > |t|) = 0.0000	Pr(T > t) = 0.0000

T-test for average volume novelty (pairs)

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
0	650,475	20.66924	.6316888	509.4699	19.43115	21.90733
1	18,568	21.6925	2.749584	374.6706	16.30306	27.08193
combined	669,043	20.69764	.6188799	506.2131	19.48466	21.91062
diff		-1.023254	3.767583		-8.407595	6.361086

diff = mean(0) - mean(1)   t = -0.2716
Ho: diff = 0   degrees of freedom = 669041

Ha: diff < 0	Ha: diff != 0	Ha: diff > 0
Pr(T < t) = 0.3930	Pr(|T| > |t|) = 0.7859	Pr(T > t) = 0.6070

T-test for career age

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
0	656,343	18.51834	.0138007	11.18062	18.49129	18.54539
1	18,779	17.27462	.0786601	10.7793	17.12043	17.4288
combined	675,122	18.48374	.0135963	11.17152	18.45709	18.51039
diff		1.243722	.0826664		1.081698	1.405745

```
    diff = mean(0) - mean(1)                                      t =  15.0451
Ho: diff = 0                                     degrees of freedom =   675120

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

T-test for number of co-authors

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
0	656,343	18.73408	.0307511	24.91304	18.67381	18.79435
1	18,779	17.60594	.1667028	22.84436	17.27919	17.9327
combined	675,122	18.7027	.0302541	24.8585	18.6434	18.76199
diff		1.128134	.1839724		.7675539	1.488714

```
    diff = mean(0) - mean(1)                                      t =   6.1321
Ho: diff = 0                                     degrees of freedom =   675120

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

T-test for gender diversity

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
0	646,402	.2904045	.0002572	.2068208	.2899003	.2909086
1	18,429	.299987	.0016077	.218255	.2968357	.3031383
combined	664,831	.2906701	.0002541	.2071521	.2901721	.291168
diff		-.0095826	.0015475		-.0126156	-.0065495

```
    diff = mean(0) - mean(1)                                      t =  -6.1923
Ho: diff = 0                                     degrees of freedom =   664829

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000
```

T-test for knowledge diversity

```
Two-sample t test with equal variances

   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 | 649,923     255.566    .1410783    113.7343    255.2894    255.8425
       1 |  18,553    240.8061    .8256716    112.4642    239.1877    242.4245
---------+--------------------------------------------------------------------
combined | 668,476    255.1563    .1390954     113.725    254.8837    255.4289
---------+--------------------------------------------------------------------
    diff |            14.75986    .8465688                13.10061    16.41911
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =  17.4349
Ho: diff = 0                                     degrees of freedom =   668474

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

Table 7: Summary statistics for time-invariant variables

Variable	Obs	Mean	Std. Dev.	Min	Max
Recency	167,101	5.035769	4.035177	-13	10
Proximity	167,101	.3924872	.4883057	0	1
Intensity	167,101	4.057289	7.258504	1	229
Intensity (old)	167,101	.915955	3.672429	0	167
Intensity (recent)	167,101	2.732958	3.973243	1	108

Table 8: T-tests for time-invariant measures

T-test for recency

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
0	1091393	4.653741	.0042139	4.402202	4.645482	4.662
1	30,855	5.183439	.0201642	3.541967	5.143916	5.222961
combined	1122248	4.668304	.0041361	4.381664	4.660197	4.676411
diff		-.529698	.0252898		-.5792651	-.4801308

diff = mean(0) - mean(1) t = -20.9451

Ho: diff = 0 degrees of freedom = 1.1e+06

Ha: diff < 0 Pr(T < t) = 0.0000

Ha: diff != 0 Pr(|T| > |t|) = 0.0000

Ha: diff > 0 Pr(T > t) = 1.0000

T-test for proximity

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
0	1091393	.4417446	.0004753	.4965949	.440813	.4426763
1	30,855	.4738616	.0028426	.4993244	.4682899	.4794333
combined	1122248	.4426277	.0004689	.4966977	.4417087	.4435466
diff		-.032117	.0028672		-.0377366	-.0264973

diff = mean(0) - mean(1) t = -11.2015

Ho: diff = 0 degrees of freedom = 1.1e+06

Ha: diff < 0 Pr(T < t) = 0.0000

Ha: diff != 0 Pr(|T| > |t|) = 0.0000

Ha: diff > 0 Pr(T > t) = 1.0000

T-test for intensity

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0	1091393	4.70599	.008009	8.366932	4.690293	4.721688
1	30,855	4.069908	.0435977	7.658187	3.984454	4.155361
combined	1122248	4.688502	.0078811	8.348895	4.673055	4.703949
diff		.6360828	.0481933		.5416256	.73054

```
    diff = mean(0) - mean(1)                                      t =  13.1986
Ho: diff = 0                                     degrees of freedom =  1.1e+06

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

T-test for recent intensity

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0	1091393	3.084995	.0044185	4.615949	3.076335	3.093655
1	30,855	3.020613	.0266829	4.687004	2.968313	3.072912
combined	1122248	3.083225	.0043592	4.617927	3.074681	3.091769
diff		.0643825	.0266586		.0121326	.1166324

```
    diff = mean(0) - mean(1)                                      t =   2.4151
Ho: diff = 0                                     degrees of freedom =  1.1e+06

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9921         Pr(|T| > |t|) = 0.0157          Pr(T > t) = 0.0079
```

T-test for old intensity

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0	1091393	1.032727	.0037697	3.938145	1.025339	1.040115
1	30,855	.8291687	.0228079	4.006342	.7844643	.8738731
combined	1122248	1.02713	.0037194	3.940174	1.019841	1.03442
diff		.2035583	.0227453		.1589783	.2481383

```
    diff = mean(0) - mean(1)                                      t =   8.9495
Ho: diff = 0                                     degrees of freedom =  1.1e+06

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

Appendix 3: Figures for summary statistics

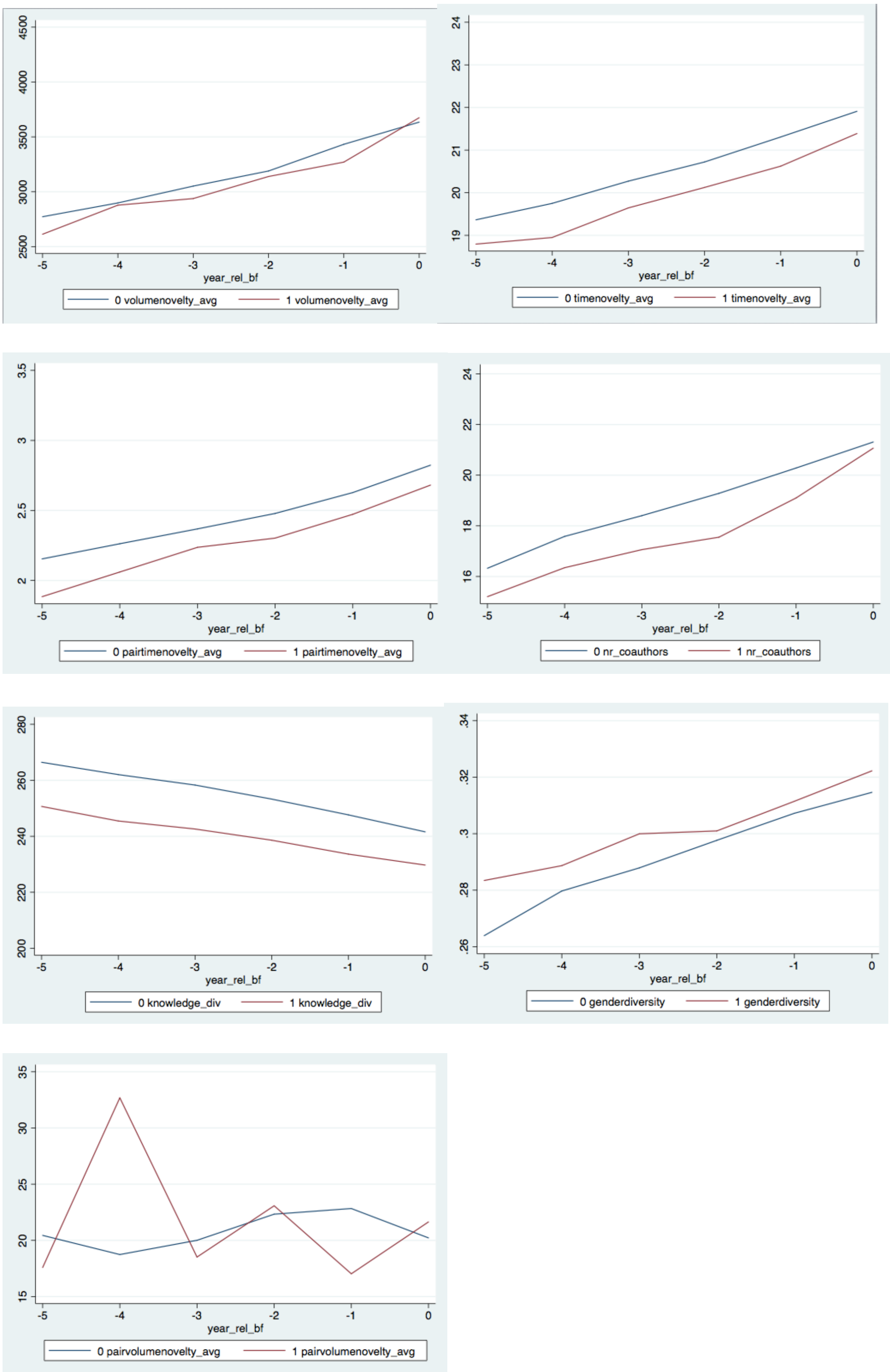

Figure 3: Graphics for time-variant variables

Appendix 4: Results for the first hypothesis

Table 9: Results of regression 1 using time novelty measure

VARIABLES	Average time novelty measure
after_X_treated_X_female	0.785***
	(0.279)
after_X_treated	-0.191
	(0.120)
Career age	0.937***
	(0.190)
Number of coauthors	-0.014***
	(0.001)
Gender diversity	-1.259***
	(0.064)
Year relative dummy1	3.841**
	(1.904)
Year relative dummy2	3.367**
	(1.713)
Year relative dummy3	2.970*
	(1.523)
Year relative dummy4	2.506*
	(1.333)
Year relative dummy5	2.140*
	(1.143)
Year relative dummy6	1.855*
	(0.953)
Year relative dummy7	1.555**
	(0.763)
Year relative dummy8	1.181**
	(0.573)
Year relative dummy9	0.844**
	(0.383)
Year relative dummy 10	0.407**
	(0.195)
Omitted year relative dummy11	-
Constant	0.810
	(4.978)

Observations 1,091,612
Number of id 164,652
R-squared 0.051
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 10: Results of regression 1 using pair time novelty measure

VARIABLES	Average time novelty using pairs
after_X_treated_X_female	0.267**
	(0.116)
after_X_treated	-0.023
	(0.052)
Career age	0.286***
	(0.091)
Number of coauthors	-0.005***
	(0.000)
Gender diversity	-0.390***
	(0.031)
Year relative dummy1	1.692*
	(0.911)
Year relative dummy2	1.522*
	(0.820)
Year relative dummy3	1.340*
	(0.729)
Year relative dummy4	1.163*
	(0.638)
Year relative dummy5	1.023*
	(0.547)
Year relative dummy6	0.947**
	(0.456)
Year relative dummy7	0.792**
	(0.365)
Year relative dummy8	0.639**
	(0.274)
Year relative dummy9	0.516***
	(0.183)
Year relative dummy 10	0.227**
	(0.094)
Omitted year relative dummy11	-
Constant	-4.016*
	(2.382)

Observations 1,091,307
Number of id 164,579
R-squared 0.013
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 11: Results of regression 1 using average volume novelty

VARIABLES	Average volume novelty
after_X_treated_X_female	-166.599
	(153.610)
after_X_treated	77.365
	(92.126)
Career age	387.674***
	(92.334)
Number of coauthors	-2.147***
	(0.317)
Gender diversity	-420.643***
	(44.629)
Year relative dummy1	1,939.941**
	(923.771)
Year relative dummy2	1,657.479**
	(831.456)
Year relative dummy3	1,419.453*
	(739.309)
Year relative dummy4	1,161.682*
	(647.061)
Year relative dummy5	989.431*
	(554.994)
Year relative dummy6	840.037*
	(462.755)
Year relative dummy7	621.725*
	(370.865)
Year relative dummy8	478.897*
	(279.317)
Year relative dummy9	336.668*
	(188.271)
Year relative dummy 10	194.481*
	(100.932)
Omitted year relative dummy11	-
Constant	-5,261.414**
	(2,414.127)

Observations 1,091,612
Number of id 164,652
R-squared 0.011
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 12: Results of regression 1 using average volume novelty (pairs)

VARIABLES	Average volume novelty using pairs
after_X_treated_X_female	15.315
	(21.584)
after_X_treated	0.820
	(7.800)
Career age	3.874
	(2.884)
Number of coauthors	-0.062***
	(0.022)
Gender diversity	-6.773*
	(3.726)
Year relative dummy1	34.732
	(29.020)
Year relative dummy2	29.013
	(26.113)
Year relative dummy3	26.670
	(23.279)
Year relative dummy4	23.778
	(20.459)
Year relative dummy5	20.931
	(17.664)
Year relative dummy6	14.331
	(14.713)
Year relative dummy7	12.353
	(11.946)
Year relative dummy8	11.325
	(9.266)
Year relative dummy9	9.845
	(6.913)
Year relative dummy 10	5.029
	(4.563)
Omitted year relative dummy11	-
Constant	-73.731
	(75.362)

Observations 1,091,307
Number of id 164,579
R-squared 0.000
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix 5: Results for the moderating effect of ties strength

Table 13: Recency for time novelty

VARIABLES	Average time novelty
after_X_treated_X_female	0.341
	(0.344)
after_X_treated	-0.191
	(0.120)
After_X_treated_X_female_X_recent	1.104**
	(0.513)
Career age	0.937***
	(0.190)
Number of coauthors	-0.014***
	(0.001)
Gender diversity	-1.259***
	(0.064)
Year relative dummy1	3.842**
	(1.904)
Year relative dummy2	3.367**
	(1.713)
Year relative dummy3	2.970*
	(1.523)
Year relative dummy4	2.506*
	(1.333)
Year relative dummy5	2.140*
	(1.143)
Year relative dummy6	1.855*
	(0.953)
Year relative dummy7	1.555**
	(0.763)
Year relative dummy8	1.181**
	(0.573)
Year relative dummy9	0.844**
	(0.383)
Year relative dummy 10	0.407**
	(0.195)
Omitted year relative dummy11	-
Constant	0.810
	(4.978)

Observations 1,091,612
Number of id 164,652
R-squared 0.051
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 14: Recency for time novelty (pairs)

VARIABLES	Average time novelty using pairs
after_X_treated_X_female	0.136
	(0.137)
after_X_treated	-0.023
	(0.052)
After_x_treated_X_female_recent	0.325
	(0.217)
Career age	0.286***
	(0.091)
Number of coauthors	-0.005***
	(0.000)
Gender diversity	-0.390***
	(0.031)
Year relative dummy1	1.692*
	(0.911)
Year relative dummy2	1.522*
	(0.820)
Year relative dummy3	1.340*
	(0.729)
Year relative dummy4	1.163*
	(0.638)
Year relative dummy5	1.023*
	(0.547)
Year relative dummy6	0.947**
	(0.456)
Year relative dummy7	0.792**
	(0.365)
Year relative dummy8	0.639**
	(0.274)
Year relative dummy9	0.516***
	(0.183)
Year relative dummy 10	0.227**
	(0.094)
Omitted year relative dummy11	-
Constant	-4.017*
	(2.382)

Observations 1,091,307
Number of id 164,579
R-squared 0.013
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 15: Proximity for time novelty

VARIABLES	Average time novelty
after_X_treated_X_female	1.277***
	(0.385)
after_X_treated	-0.191
	(0.120)
After_X_treated_X_female_X_proximity	-0.888*
	(0.503)
Career age	0.937***
	(0.190)
Number of coauthors	-0.014***
	(0.001)
Gender diversity	-1.259***
	(0.064)
Year relative dummy1	3.841**
	(1.904)
Year relative dummy2	3.367**
	(1.713)
Year relative dummy3	2.970*
	(1.523)
Year relative dummy4	2.506*
	(1.333)
Year relative dummy5	2.140*
	(1.143)
Year relative dummy6	1.855*
	(0.953)
Year relative dummy7	1.555**
	(0.763)
Year relative dummy8	1.181**
	(0.573)
Year relative dummy9	0.844**
	(0.383)
Year relative dummy 10	0.407**
	(0.195)
Omitted year relative dummy11	-
Constant	0.810
	(4.978)

Observations 1,091,612
Number of id 164,652
R-squared 0.051
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 16: Proximity for time novelty (pairs)

VARIABLES	Average time novelty (pairs)
after_X_treated_X_female	0.484***
	(0.170)
after_X_treated	-0.023
	(0.052)
After_X_treated_X_female_X_proximity	-0.390*
	(0.211)
Career age	0.286***
	(0.091)
Number of coauthors	-0.005***
	(0.000)
Gender diversity	-0.390***
	(0.031)
Year relative dummy1	1.692*
	(0.911)
Year relative dummy2	1.522*
	(0.820)
Year relative dummy3	1.340*
	(0.729)
Year relative dummy4	1.163*
	(0.638)
Year relative dummy5	1.023*
	(0.547)
Year relative dummy6	0.947**
	(0.456)
Year relative dummy7	0.792**
	(0.365)
Year relative dummy8	0.639**
	(0.274)
Year relative dummy9	0.516***
	(0.183)
Year relative dummy 10	0.227**
	(0.094)
Omitted year relative dummy11	-
Constant	-4.016*
	(2.382)

Observations 1,091,307
Number of id 164,579
R-squared 0.013
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 17: Intensity for time novelty

VARIABLES	Average time novelty
after_X_treated_X_female	0.717**
	(0.344)
after_X_treated	-0.191
	(0.120)
After_X_treated_X_female_X_intensity	0.203
	(0.509)
Career age	0.937***
	(0.190)
Number of coauthors	-0.014***
	(0.001)
Gender diversity	-1.259***
	(0.064)
Year relative dummy1	3.841**
	(1.904)
Year relative dummy2	3.367**
	(1.713)
Year relative dummy3	2.970*
	(1.523)
Year relative dummy4	2.506*
	(1.333)
Year relative dummy5	2.140*
	(1.143)
Year relative dummy6	1.855*
	(0.953)
Year relative dummy7	1.555**
	(0.763)
Year relative dummy8	1.181**
	(0.573)
Year relative dummy9	0.844**
	(0.383)
Year relative dummy 10	0.407**
	(0.195)
Omitted year relative dummy11	-
Constant	0.810
	(4.978)

Observations 1,091,612
Number of id 164,652
R-squared 0.051
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 18: Intensity for time novelty (pairs)

VARIABLES	Average time novelty
after_X_treated_X_female	0.186
	(0.134)
after_X_treated	-0.023
	(0.052)
After_X_treated_X_female_X_intensity	0.244
	(0.227)
Career age	0.286***
	(0.091)
Number of coauthors	-0.005***
	(0.000)
Gender diversity	-0.390***
	(0.031)
Year relative dummy1	1.692*
	(0.911)
Year relative dummy2	1.522*
	(0.820)
Year relative dummy3	1.340*
	(0.729)
Year relative dummy4	1.163*
	(0.638)
Year relative dummy5	1.023*
	(0.547)
Year relative dummy6	0.947**
	(0.456)
Year relative dummy7	0.792**
	(0.365)
Year relative dummy8	0.639**
	(0.274)
Year relative dummy9	0.516***
	(0.183)
Year relative dummy 10	0.227**
	(0.094)
Omitted year relative dummy11	-
Constant	-4.017*
	(2.382)

Observations 1,091,307
Number of id 164,579
R-squared 0.013
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 19: Old intensity for time novelty

VARIABLES	Average time novelty
after_X_treated_X_female	0.809***
	(0.289)
after_X_treated	-0.191
	(0.120)
After_X_treated_X_female_X_intensityold	-0.387
	(0.878)
Career age	0.937***
	(0.190)
Number of coauthors	-0.014***
	(0.001)
Gender diversity	-1.259***
	(0.064)
Year relative dummy1	3.841**
	(1.904)
Year relative dummy2	3.367**
	(1.713)
Year relative dummy3	2.970*
	(1.523)
Year relative dummy4	2.506*
	(1.333)
Year relative dummy5	2.140*
	(1.143)
Year relative dummy6	1.855*
	(0.953)
Year relative dummy7	1.555**
	(0.763)
Year relative dummy8	1.181**
	(0.573)
Year relative dummy9	0.844**
	(0.383)
Year relative dummy 10	0.407**
	(0.195)
Omitted year relative dummy11	-
Constant	0.810
	(4.978)

Observations 1,091,612
Number of id 164,652
R-squared 0.051
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 20: Old intensity for time novelty (pairs)

VARIABLES	Average time novelty (pairs)
after_X_treated_X_female	0.262**
	(0.121)
after_X_treated	-0.023
	(0.052)
After_X_treated_X_female_X_intensityold	0.090
	(0.304)
Career age	0.286***
	(0.091)
Number of coauthors	-0.005***
	(0.000)
Gender diversity	-0.390***
	(0.031)
Year relative dummy1	1.692*
	(0.911)
Year relative dummy2	1.522*
	(0.820)
Year relative dummy3	1.340*
	(0.729)
Year relative dummy4	1.163*
	(0.638)
Year relative dummy5	1.023*
	(0.547)
Year relative dummy6	0.947**
	(0.456)
Year relative dummy7	0.792**
	(0.365)
Year relative dummy8	0.639**
	(0.274)
Year relative dummy9	0.516***
	(0.183)
Year relative dummy 10	0.227**
	(0.094)
Omitted year relative dummy11	-
Constant	-4.016*
	(2.382)

Observations 1,091,307
Number of id 164,579
R-squared 0.013
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 21: Recent intensity for time novelty

VARIABLES	Average time novelty
after_X_treated_X_female	0.822***
	(0.313)
after_X_treated	-0.191
	(0.120)
After_X_treated_X_female_X_intensityrec	-0.167
	(0.585)
Career age	0.937***
	(0.190)
Number of coauthors	-0.014***
	(0.001)
Gender diversity	-1.259***
	(0.064)
Year relative dummy1	3.841**
	(1.904)
Year relative dummy2	3.367**
	(1.713)
Year relative dummy3	2.970*
	(1.523)
Year relative dummy4	2.506*
	(1.333)
Year relative dummy5	2.140*
	(1.143)
Year relative dummy6	1.855*
	(0.953)
Year relative dummy7	1.555**
	(0.763)
Year relative dummy8	1.181**
	(0.573)
Year relative dummy9	0.844**
	(0.383)
Year relative dummy 10	0.407**
	(0.195)
Omitted year relative dummy11	-
Constant	0.810
	(4.978)

Observations 1,091,612
Number of id 164,652
R-squared 0.051
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 22: Recent intensity for time novelty (pairs)

VARIABLES	Average time novelty (pairs)
after_X_treated_X_female	0.256**
	(0.126)
after_X_treated	-0.023
	(0.052)
After_X_treated_X_female_X_intensityrec	0.049
	(0.264)
Career age	0.286***
	(0.091)
Number of coauthors	-0.005***
	(0.000)
Gender diversity	-0.390***
	(0.031)
Year relative dummy1	1.692*
	(0.911)
Year relative dummy2	1.522*
	(0.820)
Year relative dummy3	1.340*
	(0.729)
Year relative dummy4	1.163*
	(0.638)
Year relative dummy5	1.023*
	(0.547)
Year relative dummy6	0.947**
	(0.456)
Year relative dummy7	0.792**
	(0.365)
Year relative dummy8	0.639**
	(0.274)
Year relative dummy9	0.516***
	(0.183)
Year relative dummy 10	0.227**
	(0.094)
Omitted year relative dummy11	-
Constant	-4.016*
	(2.382)

Observations 1,091,307
Number of id 164,579
R-squared 0.013
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix 6: Results for the second hypothesis

Table 23: Women impact on knowledge diversity

VARIABLES	Knowledge diversity
after_X_treated	1.705
	(1.369)
after_X_treated_X_female	-1.678
	(3.169)
Career age	-9.195***
	(2.657)
Gender diversity	-103.177***
	(0.726)
Year relative dummy1	-43.747*
	(26.563)
Year relative dummy2	-39.325*
	(23.902)
Year relative dummy3	-33.096
	(21.250)
Year relative dummy4	-27.626
	(18.592)
Year relative dummy5	-22.859
	(15.940)
Year relative dummy6	-18.234
	(13.284)
Year relative dummy7	-15.156
	(10.629)
Year relative dummy8	-11.374
	(7.976)
Year relative dummy9	-8.528
	(5.324)
Year relative dummy 10	-4.219
	(2.684)
Omitted year relative dummy11	-
Constant	487.699***
	(69.367)

Observations 1,100,117
Number of id 165,559
R-squared 0.088
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 24: Effect of knowledge diversity on time novelty

VARIABLES	Average time novelty
after_X_treated	-0.183
	(0.120)
after_X_treated_X_female	0.775***
	(0.279)
Career age	0.897***
	(0.188)
Gender diversity	-1.738***
	(0.065)
Number of coauthors	-0.015***
	(0.001)
Knowledge diversity	-0.005***
	(0.000)
Year relative dummy1	3.659*
	(1.878)
Year relative dummy2	3.203*
	(1.690)
Year relative dummy3	2.833*
	(1.503)
Year relative dummy4	2.392*
	(1.315)
Year relative dummy5	2.046*
	(1.128)
Year relative dummy6	1.781*
	(0.940)
Year relative dummy7	1.493**
	(0.752)
Year relative dummy8	1.135**
	(0.565)
Year relative dummy9	0.810**
	(0.378)
Year relative dummy 10	0.390**
	(0.193)
Omitted year relative dummy11	-
Constant	3.015
	(4.911)

Observations 1,091,612
Number of id 164,652
R-squared 0.053
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 25: Effect of knowledge diversity on volume novelty

VARIABLES	Average volume novelty
after_X_treated	76.417
	(92.119)
after_X_treated_X_female	-165.482
	(153.517)
Career age	392.336***
	(92.559)
Gender diversity	-364.899***
	(46.055)
Number of coauthors	-2.067***
	(0.317)
Knowledge diversity	0.534***
	(0.104)
Year relative dummy1	1,961.167**
	(925.996)
Year relative dummy2	1,676.555**
	(833.465)
Year relative dummy3	1,435.418*
	(741.100)
Year relative dummy4	1,174.920*
	(648.620)
Year relative dummy5	1,000.329*
	(556.326)
Year relative dummy6	848.631*
	(463.871)
Year relative dummy7	628.891*
	(371.759)
Year relative dummy8	484.194*
	(279.978)
Year relative dummy9	340.620*
	(188.701)
Year relative dummy 10	196.441*
	(101.136)
Omitted year relative dummy11	-
Constant	-5,518.177**
	(2,420.254)

Observations 1,091,612
Number of id 164,652
R-squared 0.011
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 26: Effect of knowledge diversity on time novelty (pairs)

VARIABLES	Average time novelty (pairs)
after_X_treated	-0.022
	(0.052)
after_X_treated_X_female	0.266**
	(0.116)
Career age	0.279***
	(0.091)
Gender diversity	-0.472***
	(0.031)
Number of coauthors	-0.005***
	(0.000)
Knowledge diversity	-0.001***
	(0.000)
Year relative dummy1	1.660*
	(0.909)
Year relative dummy2	1.494*
	(0.818)
Year relative dummy3	1.316*
	(0.728)
Year relative dummy4	1.144*
	(0.637)
Year relative dummy5	1.006*
	(0.546)
Year relative dummy6	0.934**
	(0.455)
Year relative dummy7	0.781**
	(0.364)
Year relative dummy8	0.632**
	(0.274)
Year relative dummy9	0.510***
	(0.183)
Year relative dummy 10	0.224**
	(0.093)
Omitted year relative dummy11	-
Constant	-3.640
	(2.377)

Observations 1,091,307
Number of id 164,579
R-squared 0.014
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 27: Effect of knowledge diversity on volume novelty (pairs)

VARIABLES	Average volume novelty (pairs)
after_X_treated	0.814
	(7.800)
after_X_treated_X_female	15.322
	(21.585)
Career age	3.906
	(2.884)
Gender diversity	-6.393*
	(3.655)
Number of coauthors	-0.061***
	(0.022)
Knowledge diversity	0.004
	(0.010)
Year relative dummy1	34.879
	(29.012)
Year relative dummy2	29.145
	(26.112)
Year relative dummy3	26.781
	(23.275)
Year relative dummy4	23.870
	(20.456)
Year relative dummy5	21.008
	(17.659)
Year relative dummy6	14.391
	(14.711)
Year relative dummy7	12.403
	(11.947)
Year relative dummy8	11.362
	(9.265)
Year relative dummy9	9.873
	(6.911)
Year relative dummy 10	5.042
	(4.564)
Omitted year relative dummy11	-
Constant	-75.487
	(75.479)

Observations 1,091,307
Number of id 164,579
R-squared 0.000
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix 7: Other measure of knowledge diversity

For the analysis of our second hypothesis, i.e. women impact the team's ties formation and networks with diverse knowledge are more likely to occur in their presence, we have developed a measure to evaluate the knowledge distance between co-authors. The measure is also based on Medical Subject Headings. We start by collecting all the articles a co-author has written during the 5 years before the focal scientist passed away in order to construct a set of MeSH for each of the co-author. The second step is the comparison of this set of MeSH terms with the set of MeSH of another co-author in the network. The more common MeSH two co-authors have in common, the less distant their knowledge is. For example, if co-author A is related to 15 MeSH and co-author B to 20 MeSH and that in those total 35 MeSH, 10 are common, then the distance between co-author A and co-author B is 1- (10/35) = 0.714. If two co-authors are from two totally different fields and have 0 MeSH in common, then the knowledge distance is equal to one. The last step is to aggregate the measure of knowledge distance for each co-author. So, if co-author A has 10 co-authors, we sum up the measure of knowledge distance from A to each of its ten co-authors and then divide the results by ten to obtain the average knowledge distance of co-author A with his partners.

List of figures

List of tables

Sources

Books

Angrist, Joshua D, & Pischke, Jörn-Steffen. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton: Princeton university.

Campbell, W. K., & Miller, J. D. (2011). *The handbook of narcissism and narcissistic personality disorder: Theoretical approaches, empirical findings, and treatments*. John Wiley & Sons.

Carrington, P., Scott, John, & Wasserman, Stanley. (2007). *Models and methods in social network analysis* (Reprint ed., Structural analysis in the social sciences 27): Cambridge : Cambridge university press.

Nelson, R. R., & Winter, S. G. (1982). *An evolutionary theory of economic change*. Cambridge (Mass.) : Belknap press of Harvard university press.

Poutanen, S., & Kovalainen, A. (2017). *Gender and Innovation in the New Economy*. New York: Palgrave Macmillan US.

Schumpeter, J. A. (1934). *The theory of economic development : an inquiry into profits, capital, credit, interest, and the business cycle*: Cambridge : Harvard university.

Simonton, D. K. (2004). *Creativity in science: Chance, logic, genius, and zeitgeist*: Cambridge University Press.

Tannen, D. (1990). *You just don’t understand: Women and men in conversation*. New York: Ballantine Books.

Articles

Apicella, C. L., & Dreber, A. (2015). Sex Differences in Competitiveness: Hunter-Gatherer Women and Girls Compete Less in Gender-Neutral and Male-Centric Tasks. *Adaptive Human Behavior and Physiology, 1*(3), 247-269.

Arrow, K. J. (2008). Economic Welfare and the Allocation of Resources for Invention. *Revista Brasileira De Inovacao, 7*(2), 268-285.

Azoulay, P., Fons-Rosen, C., & Graff Zivin, J. S. (2015). Does science advance one funeral at a time? *NBER Working Paper Series*.

Azoulay, P., Zivin, J. S. G., & Manso, G. (2011). Incentives and creativity: evidence from the academic life sciences. *Rand Journal of Economics, 42*(3), 527-554.

Azoulay, P., Zivin, J. S. G., & Wang, J. L. (2010). Superstar extinction. *Quarterly Journal of Economics, 125*(2), 549-589.

Babcock, L., Recalde, M. P., & Vesterlund, L. (2017). Gender Differences in the Allocation of Low-Promotability Tasks: The Role of Backlash. *American Economic Review, 107*(5), 131-135.

Babcock, L., Recalde, M. P., Vesterlund, L., & Weingart, L. (2017). Gender Differences in Accepting and Receiving Requests for Tasks with Low Promotability. *American Economic Review, 107*(3), 714-747.

Bear, J. B., & Woolley, A. W. (2011). The Role of Gender in Team Collaboration and Performance. *Interdisciplinary Science Reviews, 36*(2), 146-153.

Berings, D., De Feyter, T., Brebels, L., Van den Broeck, A., & Proost, K. (2013b). A threefold differentiation in academic motivation to disentangle the complex relationship between gender, personality and academic performance. *Journal of Education Research*, *7*(2), 111-134.

Boucher, K. L., Fuesting, M. A., Diekman, A. B., & Murphy, M. C. (2017). Can I Work with and Help Others in This Field ? How Communal Goals Influense Interest and Participations in STEM Fields. *Frontiers in Psychology, 8*, 12.

Boudreau, K. J., Guinan, E. C., Lakhani, K. R., & Riedl, C. (2016). Looking Across and Looking Beyond the Knowledge Frontier: Intellectual Distance, Novelty, and Resource Allocation in Science. *Management Science, 62*(10), 2765-2783.

Budden, A. E., Tregenza, T., Aarssen, L. W., Koricheva, J., Leimu, R., & Lortie, C. J. (2008). Double-blind review favours increased representation of female authors. *Trends Ecol Evol, 23*(1), 4-6.

Buser, T., Niederle, M., & Oosterbeek, H. (2014). Gender, competitiveness and career choices. *Quarterly Journal of Economics, 129*(3), 1409-1447.

Cattaneo, M., Meoli, M., & Signori, A. (2016). Performance-based funding and university research productivity: the moderating effect of university legitimacy. *Journal of Technology Transfer, 41*(1), 85-104.

Chatman, J., Polzer, J., Barsade, S., & Neale, M. (1998). Being different yet feeling similar: The influence of demographic composition and organizational culture on work processes and outcomes. *Administrative Science Quarterly, 43*(4), 749-780.

Collin, F. (2014). Kuhn's The Structure of Scientific Revolutions Revisited. *Science & Education, 23*(5), 1217-1222.

Dahlin, K. B., Weingart, L. R., Hinds, P. J., Kristina, B. D., Laurie, R. W., & Pamela, J. H. (2005). Team Diversity and Information Use. *The Academy of Management Journal, 48*(6), 1107-1123.

De Pater, I. E. (2011). Gender differences in job challenge: a matter of task allocation. *Strategic Direction, 27*(1), 6-9.

Diekman, A. B., Brown, E. R., Johnston, A. M., & Clark, E. K. (2010). Seeking congruity between goals and roles: a new look at why women opt out of science, technology, engineering, and mathematics careers. *Psychological Science, 21*(8), 1051-1057.

Diekman, A. B., & Steinberg, M. (2013). Navigating Social Roles in Pursuit of Important Goals: A Communal Goal Congruity Account of STEM Pursuits. *Social and Personality Psychology Compass, 7*(7), 487-501.

Diekman, A. B., Steinberg, M., Brown, E. R., Belanger, A. L., & Clark, E. K. (2017). A Goal Congruity Model of Role Entry, Engagement, and Exit: Understanding Communal Goal Processes in STEM Gender Gaps. *Personality and Social Psychology Review, 21*(2), 142-175.

Eagly, A. H., & Johnson, B. T. (1990). Gender and leadership-style: a meta-analysis. *Psychological Bulletin, 108*(2), 233-256.

Eckles, D., Kizilcec, R. F., & Bakshy, E. (2016). Estimating peer effects in networks with peer encouragement designs. *Proceedings of the National Academy of Sciences of the United States of America, 113*(27), 7316-22.

Fallah, M. H., & Ibrahim, S. (2004). Knowledge spillover and innovation in technological clusters. *Proceedings, IAMOT 2004 Conference, Washington, DC*, 1-16.

Fellner, G., & Maciejovsky, B. (2007). Risk attitude and market behavior: Evidence from experimental asset markets. *Journal of Economic Psychology, 28*(3), 338-350.

Fenwick, G. D., & Neal, D. J. (2001). Effect of Gender Composition on Group Performance. *Gender, Work & Organization, 8*(2), 205-225.

Fershtman, C., & Gandal, N. (2011). Direct and indirect knowledge spillovers: the "social network" of open-source projects. *Rand Journal of Economics, 42*(1), 70-91.

Forrester, W. R., & Taschchian, A. (2010). Effects of Personality on Attitudes toward Academic Group Work. *American Journal of Business Education, 3*(3), 39-46.

Fracassi, C., & Tate, G. (2012). External Networking and Internal Firm Governance. *Journal of Finance, 67*(1), 153-194.

Franzoni, C. (2010). Exploratory attitude and creativity in science. *Science for Innovation: Exploiting and strengthening the linkage, Research Institute of Economy, Trade and Industry*.

Frietsch, R., Haller, I., Funken-Vrohlings, M., & Grupp, H. (2009). Gender- specific patterns in patenting and publishing. *Research Policy, 38*(4), 590-599.

Garfield, E. (1980). Premature discovery or delayed recognition-Why? *Current Contents*(21), 5-10

Gneezy, U., Leonard, K. L., & List, J. A. (2009). Gender differences in competition: evidence from a matrilineal and a patriarchal society. *Econometrica, 77*(5), 1637-1664.

Grijalva, E., Newman, D. A., Tay, L., Donnellan, M. B., Harms, P. D., Robins, R. W., & Yan, T. (2015). Gender Differences in Narcissism: A Meta-Analytic Review. *Psychological Bulletin, 141*(2), 261-310.

Hansen, Z., Owan, H., & Pan, J. (2006). The Impact of Group Diversity on Performance and Knowledge Spillover -- An Experiment in a College Classroom. *NBER Working Paper Series*, 12251.

Hicks, D. (2012). Performance-based university research funding systems. *Research Policy, 41*(2), 251-261.

Hunt, J., Garant, J. P., Herman, H., & Munroe, D. J. (2013). Why are women underrepresented amongst patentees? *Research Policy, 42*(4), 831-843.

Iorio, R., Labory, S., & Rentocchini, F. (2017). The importance of pro-social behaviour for the breadth and depth of knowledge transfer activities: An analysis of Italian academic scientists. *Research Policy, 46*(2), 497-509.

Jaravel, X., Petkova, N., & Bell, A. (2015). Team-Specific Capital and Innovation. *American Economic Review, 108*(4-5), 1034-1073.

Jehn, K. A., Northcraft, G. B., & Neale, M. A. (1999). Why differences make a difference: A field study of diversity, conflict, and performance in workgroups. *Administrative Science Quarterly, 44*(4), 741-763.

Jones, B. F. (2009). The Burden of Knowledge and the " Death of the Renaissance Man": Is Innovation Getting Harder? *Review of Economic Studies, 76*(1), 283-317.

Jones, G. (2017). Knowledge sharing and technological innovation: The effectiveness of trust, training, and good communication. *Cogent Business & Management*, *4*(1).

Joshua, M. N., & John, P. A. I. (2012). Research grants: Conform and be funded. *Nature, 492*(7427), 34-6.

Landry, R., Traore, N., & Godin, B. (1996). An econometric analysis of the effect of collaboration on academic research productivity. *Higher Education, 32*(3), 283-301.

Lee, C., & Farh, J. L. (2004). Joint effects of group efficacy and gender diversity on group cohesion and performance. *Applied Psychology, 53*(1), 136-154.

Lee, Y. N., Walsh, J. P., & Wang, J. (2015). Creativity in scientific teams: Unpacking novelty and impact. *Research Policy, 44*(3), 684-697.

Lehmann, R., Denissen, J. J. A., Allemand, M., & Penke, L. (2013). Age and Gender Differences in Motivational Manifestations of the Big Five from Age 16 to 60. *Developmental Psychology, 49*(2), 365-383.

Levin, D. Z., & Cross, R. (2004). The strength of weak ties you can trust: The mediating role of trust in effective knowledge transfer. *Management Science, 50*(11), 1477-1490.

Lissoni, F., Montobbio, F., & Zirulia, L. (2013). Inventorship and authorship as attribution rights: An enquiry into the economics of scientific credit. *Journal of Economic Behavior & Organization, 95*, 49-69.

Mauleón, E., Daraio, C., & Bordons, M. (2014). Exploring gender differences in patenting in Spain. *Research Evaluation, 23*(1), 62-78.

Mednick, S. (1962). The associative basis of the creative process. *Psychological Review, 69*(3), 220-232.

Mishra, S., & Torvik, V. I. (2016). Quantifying conceptual novelty in the biomedical literature. *D-Lib Magazine, 22(9-10)*.

Mohnen, M. (2016). Stars and brokers: peer effects among medical scientists. *Job market paper*.

Monastersky, R. (2005). The Number That's Devouring Science.(grading system for scholarly journals). *The Chronicle of Higher Education, 52*(8), 12-17.

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences of the United States of America, 109*(41), 16474-16479.

Mottaz, C. (1986). Gender Differences in Work Satisfaction, Work-Related Rewards and Values, and the Determinants of Work Satisfaction. *Human Relations, 39*(4), 359-377.

Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics, 122*(3), 1067-1101.

Nonaka, I., Byosiere, P., Borucki, C. C., & Konno, N. (1994). Organizational knowledge creation theory: A first comprehensive test. *International Business Review, 3*(4), 337-351.

Oettl, A. (2012). Reconceptualizing Stars: Scientist Helpfulness and Peer Performance. *Management Science, 58*(6), 1122-1140.

Packalen, M., & Bhattacharya, J. (2015a). Age and the Trying Out of New Ideas. *NBER Working Paper Series*, 20920.

Packalen, M., & Bhattacharya, J. (2015b). New Ideas in Invention. *NBER Working Paper Series*, 20922.

Pelled, L. H., Eisenhardt, K. M., & Xin, K. R. (1999). Exploring the Black Box: An Analysis of Work Group Diversity, Conflict and Performance. *Administrative Science Quarterly, 44*(1), 1-28.

Pieterse, A. N., van Knippenberg, D., & van Dierendonck, D. (2013). Cultural diversity and team performance: the role of team member goal orientation. *Academy of Management Journal, 56*(3), 782-804.

Probert, B. (2005). 'I just couldn't fit it in': Gender and unequal outcomes in academic careers. *Gender Work and Organization, 12*(1), 50-72.

Ruiz-Jimenez, J. M., Fuentes-Fuentes, M. D., & Ruiz-Arroyo, M. (2016). Knowledge Combination Capability and Innovation: The Effects of Gender Diversity on Top Management Teams in Technology-Based Firms. *Journal of Business Ethics, 135*(3), 503-515.

Schiebinger, L., & Schraudner, M. (2011). Interdisciplinary Approaches to Achieving Gendered Innovations in Science, Medicine, and Engineering. *Interdisciplinary Science Reviews, 36*(2), 154-167.

Schoenmakers, W., & Duysters, G. (2010). The technological origins of radical inventions. *Research Policy, 39*(8), 1051-1059.

Taylor, A., Greve, H. R., Alva, T., & Henrich, R. G. (2006). Superman or the Fantastic Four? Knowledge Combination and Experience in Innovative Teams. *The Academy of Management Journal, 49*(4), 723-740.

Torvik, V. I. (2015). MapAffil: A Bibliographic Tool for Mapping Author Affiliation Strings to Cities and Their Geocodes Worldwide. *Dlib Magazine, 21(11-12)*.

Torvik, V. I., & Agrawal, S. (2016). Ethnea – an instance-based ethnicity classifier based on geo-coded author names in a large-scale bibliographic database. *International Symposium on Science March 22-23, 2016*.

Torvik, V. I., & Smalheiser, N. (2009). Author name disambiguation in MEDLINE. *ACM Transactions on Knowledge Discovery from Data (TKDD), 3*(3), 1-29.

Trapido, D. (2015). How novelty in knowledge earns recognition: The role of consistent identities. *Research Policy, 44*(8), 1488-1500.

Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical Combinations and Scientific Impact. *Science, 342*(6157), 468-472.

van Knippenberg, D., De Dreu, C. K. W., & Homan, A. C. (2004). Work group diversity and group performance: An integrative model and research agenda. *Journal of Applied Psychology, 89*(6), 1008-1022.

van Veldhuizen, R. (2016). Gender differences in tournament choices: Riskpreferences, overconfidence or competitiveness? *WZB Discussion Paper, No. SP II 2016-207*.

Van Vugt, M., De Cremer, D., & Janssen, D. P. (2007). Gender differences in cooperation and competition - The male-warrior hypothesis. *Psychological Science, 18*(1), 19-23.

Wang, J., Veugelers, R., & Stephan, P. (2017). Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy, 46*(8), 1416-1436.

Watt, H. M. G., Shapka, J. D., Morris, Z. A., Durik, A. M., Keating, D. P., & Eccles, J. S. (2012). Gendered Motivational Processes Affecting High School Mathematics Participation, Educational Aspirations, and Career Plans: A Comparison of Samples From Australia, Canada, and the United States. *Developmental Psychology, 48*(6), 1594-1611.

Webb, T. J., O'Hara, B., & Freckleton, R. P. (2008). Does double-blind review benefit female authors? *Trends Ecol Evol, 23*(7), 351-353; author reply 353-354.

Weisberg, Y. J., Deyoung, C. G., & Hirsh, J. B. (2011). Gender Differences in Personality across the Ten Aspects of the Big Five. *Frontiers in Psychology, 2,* 178.

Weitzman, M. L. (1998). Recombinant growth. *Quarterly Journal of Economics, 113*(2), 331-360.

Williams, K. Y., & O'Reilly, C. A. (1998). Demography and diversity in organizations: A review of 40 years of research. *Research in Organizational Behavior, Vol 20, 1998: an Annual Series of Analytical Essays and Critical Reviews, 20*, 77-140.

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a Collective Intelligence Factor in the Performance of Human Groups. *Science, 330*(6004), 686-688.

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science, 316*(5827), 1036-1039.

Youn, H., Strumsky, D., Bettencourt, L. M. A., & Lobo, J. (2015). Invention as a combinatorial process: evidence from US patents. *Journal of the Royal Society, Interface, 12*(106).

Articles in books

Berings, D, De Feyter, T, Brebels, L, Van den Broeck, A, & Proost, K. (2013). Gender, personality and academic performance: Self-determination theory and regulatory focus theory as levers to open the black box between gender, personality, academic effort and performance. In *Handbook of Academic Performance: Predictors, Learning Strategies and Influences of Gender*. Nova Science ; New York. pp 1-33.

Carli, L. (2010). Gender and group behavior. In J. C. Chrisler & D. R. McCreary (Eds.), *Handbook of Gender in Psychology*. New York: Springer. pp. 337-358.

Foster, J. D., & Brennan, J. C. (2011). Narcissism, the agency model, and approach-avoidance motivation. In W. K. Campbell & J. D. Miller (Eds.), *The handbook of narcissism and narcissistic personality disorder: Theoretical approaches, empirical findings, and treatments*. Hoboken, NJ, US: John Wiley & Sons Inc. pp 89-100.

The novelty of the research is crucial in order to make the society move forward. We argue that the closure of the gender gap still prevailing in academia, by bringing more women into scientific research, can improve novelty. Indeed, women have different skills and personality traits likely to positively shape the knowledge transfers inside the research teams and thus to trigger novelty. Our hypothesis is that women have higher peer effects on their co-author than men. Using articles indexed in MEDLINE, we calculate various measures of novelty and test our hypothesis through a difference-in-difference strategy. We extract the impact of a woman on her peers by measuring the change in her coauthors' novelty before and after she disappears from the team. Our results confirm the hypothesis for some of the novelty measures: female presence in research teams does well enhance co-authors' novelty and thus novelty of the research.

Place des Doyens, 1 bte L2.01.01, 1348 Louvain-la-Neuve, Belgique **www.uclouvain.be/lsm**

UCL
Université catholique de Louvain