Words Matter?
Gender Disparities in Speeches, Evaluation and Competitive Performance

ISBN: 978 90 361 0670 2

Cover design: Crasborn Graphic Designers bno, Valkenburg a.d. Geul

This book is No. **789** of the Tinbergen Institute Research Series, established through cooperation between Rozenberg Publishers and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

Words Matter?
Gender Disparities in Speeches, Evaluation and Competitive Performance

Woorden doen ertoe? Genderverschillen in toespraken, evaluatie en competitieve prestaties

Thesis

to obtain the degree of Doctor from the

Erasmus University Rotterdam

by command of the Rector Magnificus

Prof. dr. L. A. Bredenoord

and in accordance with the decision of the Doctorate Board.

The public defence shall be held on December 1, 2021, at 15:30 hours

by

HUYỀN THỊ THANH NGUYỄN

born in Hà Nội, Việt Nam

Erasmus University Rotterdam

Doctoral Committee:

Promotor:	Prof. dr. O. H. Swank
Other members:	Prof. dr. J. B. Hirschberg Prof. dr. M. P. García-Gómez Prof. dr. H. D. Webbink
Co-promotor:	Dr. J. Delfgaauw

Acknowledgments

It feels surreal to type down these words, reflecting on my entire Ph.D. adventure, at this moment. Back in the days battling the first-year MPhil courses at the Tinbergen Institute, and during the toughest moments in this pandemic year, reaching the Ph.D. defense altar feels like such an impossible height for me. And it definitely would have, without the incredible humans and institutions that embraced me throughout these years.

First and foremost, this dissertation and the researcher I am today would not have come into existence without my advisor, Josse Delfgaauw. Your unwavering trust in my research capability, your sympathetic guidance, and your unflinching support, from the MPhil thesis stage to all these Ph.D. years, nurture my growth as a researcher. The unmatched intellectual freedom you enabled for my academic growth, even when it means venturing into other disciplines than economics, means the world to me. Stories about your *hundreds* of meticulous comments for my paper drafts (by now they collectively are in the thousands...), your never-ending enthusiasm to constructively discuss and critique research ideas, and your day-to-day open-door policy are often met with wide-eyed disbelief, if not downright jealousy. I learn how to become a supportive, inspiring, and effective academic from your dedicated attitude towards the work of colleagues and department visitors, as well as your efficient handling of numerous administrative tasks. You are crucial not only for my academic growth but also for my growth as a person. I am indebted to your sympathetic ear in my times of need, to your pragmatic advice about how to overcome the toughest of struggles, and to your incredible kindness. The fact that I could complete this dissertation against all odds, is (one of) the testaments to how blessed I am to have you as my supervisor.

Secondly, I would like to thank my promotor, Otto Swank, and all members of my defense committee for their valuable time and effort in assessing my dissertation. Julia Hirschberg, thank you for your detailed comments to significantly improve my dissertation writing, as well as the countless ideas and practical tips on furthering my research agenda. Dinand Webbink and Pilar García-Gómez, thank you for your constructive feedback on the next steps for these chapters. Thomas Buser, I am honored to have you as part of my defense committee. Your expansive research work on gender and labor economics has been a major source of inspiration for me. On top of that, you have always been exceptionally generous in sharing your feedback and ideas on my work in earlier stages. Malia Mason, having my academic female role models, not just one, but two of them on my defense committee, makes this day exceptionally wonderful.

All these years would have been so dull without the supportive, approachable yet critical minds of ESE Economics Department. Otto, thank you for being the best head of department ever. The easy-to-mingle atmosphere in the seminars, the special wine tasting and department get-together events, are certainly the unique trademark of our department under your leadership. Vladimir, as much as I dread your intensive Micro I course, I hugely appreciate the opportunity to be your TA for Game Theory course, and learn from you a structured. Furthermore, your spot-on feedback on my paper presentation in its infancy stage definitely shapes it towards the right path. Dirk, Bas, Dinand and Sacha, thank you for your critical comments and warm encouragement for my work. Anne Gielen, thank you for your essential guidance on the academic paths, especially during pandemic times. Anne Boring, thank you for giving me necessary pointers to improve my paper in its infancy stages and organizing many meaningful network meeting opportunities for female academics at EUR. Anna Baiardi, Yao and Felix, thank you for giving me useful advice on my paper drafts, as well as the meaningful life conversations we had together. Ankimon and the wonderful secretariat ladies, thank you for always navigating the myriads of administrative and contractual arrangements for me. Judith and Caroline, thank you for all your timely help with any administrative and dissertation publication matters, as well as the wonderful moments during my MPhil years.

Next, I would like to thank the ESE Diversity & Inclusion Office, the Economics department and the Erasmus Trusfond for the generous financial support to collect the data and complete part of my research at Columbia University. I thank my research assistants: Fenna ten Haaf and Jan-Gunther Gosselke, for their concerted effort and adeptness in assisting me collecting and making sense of the data. Franka Boender, thank you for your initial assistance in making sense of the argumentation system. Jan-Gunther, thank you especially for your timely coding expertise, which helps me through many complex data hurdles. Furthermore, I am extremely grateful to the ESE economics department for enabling me to participate in many useful research conferences, summer schools and workshops across Europe. From economics, political science to data science and machine learning, the intriguing discussions and presentations at these venues have definitely opened my minds to the ever-growing excitement and importance of interdisciplinary research.

It would be an unforgivable amiss to forget the highlight of my Ph.D. time: the research visit at Columbia Business School in New York. I am incredibly grateful to Malia Mason, my host, my mentor, and my role model female academic, for making me feel as inspired and homely as possible, during my entire stay. Ashley, Sam, Pete and Patrick, thank you for being so welcoming and inclusive towards me. Julia and Sarah Ita, thank you so much for enthusiastically embracing my debate research agenda; and I feel so blessed about our collaboration thus far. Furthermore, it was my honor to be part of the awe-inspiring I-House NY community. The deep exchanges at the bar, the fun ice cream socials, the Swedish Fit Mondays, and the talks of inspiring figures make me feel instantly at home in New York. Singing *Empire State of Mind* together at the Fall Fiesta after-party will always have a special place in my heart. This three-and-a-half month period at I-House and Columbia University is undoubtedly one of the most eye-opening experiences in my life.

Back to the days on the H-building *promovendus* 8^{th} Floor, I am grateful to the joyful moments with my PhD friends. Nam, thank you for being the most cynical, yet indispensable office-mate, especially for our weekly restaurant exploration in Rotterdam. Yan, Benjamin, Jingni, Esmee and Mathijs, thank you for enliven the office with your optimistic and welcoming characters. All my MPhil friends, specifically, to name a few, Jenny, Dieter, Lingwei, Kostas and Johan, thank you for the numerous moments of fun get-together, of insightful conversations and of being the wonderful company you are, as always.

It is fair to say that, the vast majority of my enriching Ph.D life experiences is credited to the European debate community. Be it sparing with these critical minds over controversial topics, weighing up the merits and what-nots of their speeches, or simply just exasperating over bizarre motions with these "idea junkies" over drinks, debaters have always been the indispensable spices to enlighten the seemingly never-ending dark tunnels of research.

I thank my TU Delft Debate Club friends, Gerson, Bram, Tanya, Jorino and Albert, for embracing me into debating. Milla, Kat, Lucia, Sarah, and Gigi, I am in awe of your incredible wits and grit to navigate the game to the top, and even more inspired by your tireless effort to make this "mental" sport inclusive to everyone. Annabelle, my Corona buddy sister, I feel very lucky to share the ups and downs of this pandemic in Rotterdam with you. Milla and Migle, thank you for all the beautiful moments and soul-touching stories we share. To all debate friends whom I have been fortunate enough to cross paths with, this thesis is as much as what I hope to contribute to economics, as it is to our debate community.

Undeniably, what keeps me going forward against all odds, is the boundless emotional support from my family. Mẹ, Tí, and Sunny, thank you for believing in my wildest of decisions, for better or worse, under any circumstances. I owe my ever-expanding life experiences and wisdom to your wholehearted trust and all-embracing love. Thank you to my extended family - chú Thành, cô Trang, Cindy, Nhật and Charly. Without your above-and-beyond generosity with your time, space, and practical assistance, I would have failed to keep my wits about me, especially in this past tumultuous year. Thank you for taking me in through the darkest hours, for making my pandemic relocation trips a reality, and for your loving care, ever since I came to Europe.

In these rare moments of gratitude, I would like to leave the last words to myself. As common wisdom goes, the hardest lesson one needs to learn is to love oneself, especially in academia. Hence, to my past, current and future self: Thank you for failing, wailing and trying over and again, for never letting go of yourself even in the darkest hours, and for taking the time and space to care for yourself. Keep on being open to what makes you glow, inspired and fulfilled in life.

Huyền T.T. Nguyễn, 09/2021

"Không có việc gì khó
Chỉ sợ lòng không bền
Đào núi và lấp biển
Quyết chí ắt làm nên."

Bác Hồ

CONTENTS

1 Introduction

Humans speak to make ourselves understood, explain to get our ideas across, and debate to settle conflicts between one another. From convincing a friend to dine with you at your favorite restaurant to arguing for a case in front of the court, we engage with and persuade others in all aspects of our everyday lives. Mastering the art of persuasion is arguably the key to success, especially in highly competitive jobs with multi-dimensional tasks and complex organizational settings.

In the higher rungs of high-flying careers in business, academia, the law, or politics, one commonly observed fact is that women are persistently under-represented [Goldin et al., 2017; Blau and Kahn, 2017; Eckel et al., 2020]. In fact, recent research on gender disparities in willingness to negotiate salaries [Bohnet and Bowles, 2008; Leibbrandt and List, 2015], promote oneself [Exley and Kessler, 2019], perform a real effort task [Alan et al., 2020] or speak publicly [Buser and Yuan, 2020] has shed light on an important behavioral aspect: persuasion styles in high-stake contexts.

Notwithstanding the importance to understand how differences in persuasion styles across genders matter to gender representation and outcome gap, there exists very limited systematic evidence on gender disparities in speech patterns and evaluations. In most real-world settings, determining whether gender differences in outcomes are driven by differences in behavior or gender-specific evaluation patterns (i.e. discrimination) is inherently difficult due to two reasons. First, large-scale text data sets in a competitive setting where argumentation strength is unconfounded by *ad hominem* strategies or backdoor agreements are extremely scarce. In existing large-scale, textual communication data e.g. political debates [Gentzkow et al., 2019], central bank communication strategies [Hansen et al., 2018] or judicial court opinion polarization [Ash et al., 2017], in addition to an inordinate gender imbalance in the actors at hand, a transparent and rigorous evaluation procedure is absent. Consequently, linking specific speech patterns to the evaluation of persuasiveness across gender is infeasible. Second, persuasion, or communication in general, is context-dependent i.e. whether what one says is *more* persuasive, or deemed so, compared to others in the room, depends on who one faces and on who evaluates their speeches. Undeniably speaking, we listen through our own brain filters, biases of our lenses, and varying perceptions of the individuals across the tables. In order to comprehensively understand how and what matters for persuasion across genders, we need a well-defined, high-stake competitive setting with clear-cut and rigorous rules to value exclusively argumentation merits across contestants.

This dissertation contributes novel insights on the role of gender in persuasion tactics, competitive performance and evaluation patterns, in a unique setting of international university debate tournaments. These tournaments and their participants provide an attractive setting to systematically answer these questions. First, these competitions take place annually, at the European or worldwide scale, in a multi-round tournament setting following the widely used British Parliamentary Debate format across a variety of controversial topics. Second, participants are intrinsically motivated students representing various academic institutions worldwide, whose persuasion motives are similar to those of lawyers, politicians, and academics. In fact, many famous politicians, lawyers, and judges trained their persuasion skills in competitive debating, thus making this setup externally relevant to real-life competitive contexts. For each debate round, participants are randomly assigned debate topics, speaking positions (i.e. for or against the topic), opponents, and judge panels. Every participant gives a 7-minute speech to convince a panel of trained judges, who are incentivized to evaluate speeches fairly given past achievements and peer performance feedback. Such incentive architecture mirrors real-life committee decisions, where career concerns, authority play, and social pressures matter. Importantly, comparative argument strength is the yardstick to success in these tournaments. In other words, *ad hominem* argumentation strategies, which is the common confound with argumentation merits in political debates, are outlawed in debate tournament speeches. In the introduction section of each chapter, the reader will find the uniquely relevant debate tournament advantages to study the respective research questions.

By combining state-of-the-art, persuasion-relevant natural language processing with econometric techniques, this dissertation is a collection of three empirical essays investigating persuasive communication performance and evaluations *across genders* in a relevant competitive context i.e. high-profile international debate tournaments. Chapter 2: The (Great) Persuasion Divide? Gender Disparities in Debate Speeches and Evaluations draws on recent advances in dictionary-based persuasion methods [Pennebaker et al., 2015] to extract spoken verbal tactics across genders in 1517 speech transcripts of the highest-profile inter-varsity debate tournaments to understand: (1) whether men and women persuade differently; and (2) how their persuasion patterns matter for competitive evaluations among committees. I find significant variation in speech patterns across genders. Female speakers use a more personal and disclosing speaking style, with more hedging phrases and non-fluencies in their speeches. In their answers to questions from opponents, women negate less, while having longer and more vague answers. On average, women receive lower evaluation scores than men. Across debates, having a less analytical speaking style and more positive sentiment is associated with higher scores for speeches by women, but not by men. Within debates, except for non-fluencies, there is no robust evidence of gender-specific evaluation standards. Noteworthily, within debates, even though evaluation patterns are similar for male and female speakers across judges and the judge panel gender compositions, committees with more female judges are significantly harsher towards female speakers. Overall, these insights suggest that the gender score gap arises because speeches of female speakers contain more score-reducing and fewer score-enhancing features, rather than discrimination.

Since evaluators play a critical role in determining persuasiveness among contestants, Chapter 3: Gender Composition of Committees and Performance Evaluation: Evidence from Debate Tournaments explores the causal impact of the gender composition of 4896 committees on 39 168 competitive speech performance scores across European and World Universities Debate Championships. Here I find that committees with a female chair judge give lower scores to both male and female speakers, particularly in higher-ranked debates. The gender of other committee members does not affect evaluations. While accomplished male chair judges are more generous in scoring, they are notably less so towards female speakers. These results demonstrate that gender quotas on evaluation committees do not necessarily eliminate the glass ceiling for women in high-stake, repeated competition contexts.

Last but not least, given that the gender of opponents has been hypothesized to impact the competitive performance of real-world contestants, Chapter 4: Choking upon Facing (Fe)male Opponents? Evidence from Debate Tournaments examines whether the gender of debate opponents causally affect the competitive performance of contestants, by exploiting the random assignment of 3153 participants to multiple rounds of debate matches. On average, I find that the performance of neither men nor women is affected by the gender composition of opponents. In higher-ranked debates, female speakers perform comparatively worse in rooms with more female opponents. These findings indicate that more inflow of women into competitions for high-profile careers does not necessarily reduce the thickness of the glass ceiling.

All in all, this dissertation serves to expand our understanding of how and to what extent oral persuasion patterns matter to performance evaluation, in a uniquely relevant competitive context. Findings from these chapters have three important implications. First, assuming that these results carry over to workplace settings, gender differences in outcomes of negotiations and job interviews would be attributable to differences in persuasion tactics, rather than how negotiation is evaluated. Since the lexical features investigated in this high-stake, competitive, male-dominated context correlate with confidence and charisma, the finding that female speakers have more features correlated with lower confidence and performance scores speaks to the exhibited gender gap in self-promotion, leadership tendency, and workplace authority. Second, the null finding of increasing female members or having a female chair in a committee raises doubts about the *direct* effectiveness of gender quota law, in and of its own, on smashing the glass ceiling for women to the top. Given the ever-growing implementation of such a law across the world, it is important to keep in mind other crucial institutional setups and mechanisms in truly creating an equitable competitive environment. Finally, since female speakers perform comparatively worse in debates with more female contestants, it is crucial that policymakers consider alternative setups of competitions into high-profile careers if their goal is to taper the barriers at the top.

"Persuasion is achieved by the speaker’s personal character when the speech is so spoken as to make us think him credible. We believe good men more fully and more readily than others: this is true generally whatever the question is, and absolutely true where exact certainty is impossible and opinions are divided."

Aristotle

2 The (Great) Persuasion Divide? Gender Disparities in Debate Speeches and Evaluations

" When I’m sometimes asked when will there be enough women on the Supreme Court and I say, ’When there are nine,’ people are shocked. But there’d been nine men, and nobody’s ever raised a question about that."

Ruth Bader Ginsburg

3 Gender Composition of Committees and Performance Evaluation: Evidence from Debate Tournaments

"Know thy self, know thy enemy. A thousand battles, a thousand victories."

Sun Tzu

4 Choking upon Facing (Fe)male Opponents? Evidence from Debate Tournaments

4.1 Introduction

It is a well-established fact that the higher up the rungs of career ladders, the fewer women are present, especially in competitive occupations [Goldin et al., 2017; Blau and Kahn, 2017; Eckel et al., 2020]. A significant body of research have linked women's distaste for competition [Niederle and Vesterlund, 2007; Buser et al., 2014; Niederle and Vesterlund, 2011; Villeval, 2012; Datta Gupta et al., 2013] and under-performance in competitive settings [Niederle, 2017; Gneezy et al., 2003; Gneezy and Rustichini, 2004; Antonovics et al., 2009; Shurchkov, 2012] to this persistent gap. One prominent hypothesis revolves around the idea that the gender of opponents affects competitive performance [Shurchkov and Eckel, 2018]. Yet, evidence on this hypothesis is mixed, and especially scarce when it comes to real-life, dynamic competitions on multi-dimensional and complex tasks.

On the one hand, women have been shown to perform worse, especially when facing men, in the seminal lab evidence of [Gneezy et al., 2003] and [Niederle and Vesterlund, 2007], and across high-stake field studies [van Dolder et al., 2020; Säve-Söderbergh and Sjögren Lindquist, 2017]. On the other hand, numerous lab and field studies demonstrate that women compete similarly or better against men, from one-on-one lab experiments [Moely et al., 1979; Conti et al., 2001; Mago and Razzolini, 2019], real-effort team competition [Ivanova-Stenzel and Kübler, 2011] to high-stake field settings [Antonovics et al., 2009; Jetter and Walker, 2018; Iriberri and Rey-Biel, 2019]. Noteworthily, most of the relevant studies assess either one-on-one competitive settings or the overall gender composition in a static competition environment. Since competitive success often requires repeated interaction in larger groups, over multiple rounds of applications and assessments in labor markets, it raises the question: To what extent can the observed gendered performance patterns given the gender composition of opponents be generalized to other real-life contests?

This chapter exploits the random assignment of 3153 contestants to multiple rounds of international university debate tournaments[1] to causally investigate how the gender composition of opponents affects speech performance. Competitive debating is a complex, multi-dimensional skill[2] that is externally relevant to careers in law, politics, business, and academia, where oral persuasion skill is instrumental to success [Buser and Yuan, 2020]. Three institutional features make these competitions an attractive setting to investigate whether the gender composition of opponents affects one's competitive performance. First, with 3153 contestants giving 39 168 speeches across nine debate rounds in each tournament, I can include individual fixed effects to control for unobserved factors, such as the innate ability of contestants. Second, in fixed teams of two, for every round, participants are exogenously matched to compete against three other teams (i.e. six other competitors).[3] This system creates a randomly allocated set of opponents in terms of gender across debates. Finally, except for Round 1 where team matching is completely randomized, in each N^{th} round, every debate consists of contestants with similar $(N-1)$ speech performance records. This *power-matching mechanism* mirrors the labor market contests, where repeated relative performance evaluations are used to assign jobs and promote employees. It also enables me to study whether the impact on speech performance given the gender composition of opponents is consistent across debate room levels, especially with heightened competitive pressure in high-ranked debates[4] as the preliminary rounds progress.

I find that, on average, the performance of *neither* male nor female debaters is affected by the gender composition of opponents. Overall, while an additional female opponent is

[1] Four annual World Universities Debating Championships (WUDC), and four annual European Universities Debating Championships (EUDC), from 2015 to 2018.

[2] Debating is considered one of the most effective activities to train four major language skills [Green III and Klug, 1990; Li et al., 2019] and leadership skill [Chikeleze et al., 2018].

[3] In general, opponents, judges, topics, and speaking position to argue for or against a policy-relevant topic are exogenously assigned. See Appendix 4.9 for details on the Debate Format.

[4] i.e. For N^{th} round, higher-ranked debates are those where teams with speech performance record equal to or higher than the *median* cumulative performance record in $(N-1)$ rounds.

associated with a reduction of 2.0 percentage point standard deviation in speech score this unconditional score gap vanishes upon controlling for speaker fixed effect. This result is confirmed in the non-parametric specification of the number of female opponents. In other words, within individuals, the speech performance of neither men nor women responds to the number of female opponents they face in a debate.

In higher-ranked debates, women perform comparatively worse when facing more *female* opponents, whereas the performance of male speakers is unaffected by the gender composition of opponents. Controlling for speaker fixed effects, I estimate that for female speakers, an additional female opponent yields a 2.1 percentage point standard deviation reduction in score. No significant finding regarding the gender of opponents is detected in lower-ranked debates. An alternative analysis with a non-parametric specification of opponents' gender among these rooms shows negative and significant results for female speakers given *any* number of female speakers they face. For male speakers in higher-ranked debates, their speech performances are only comparatively worse in rooms with four female opponents or more. Furthermore, this gender score gap concerning female opponents is observed in women in female-only teams, and not those in mixed-gender teams.

This chapter contributes a causal finding on the interplay between one's gender and the gender composition of opponents in high-profile debate tournaments. This task and tournament setup particularly complements the current literature, which usually involves tasks without oral persuasion elements. A copious body of literature in one-on-one settings has shown that women perform comparatively worse when they face men, for instance, in the seminal work of [Gneezy et al., 2003], [Niederle and Vesterlund, 2007]. In [Delfgaauw et al., 2013], they show that sales competition among employees increases sales growth, but only in stores where the majority have the same gender. Meanwhile, [Datta Gupta et al., 2013] found that men choose to compete for less against other men than against women. In the field, [van Dolder et al., 2020] use data from the Dutch *Jeopardy!* shows to demonstrate female contestants perform worse when facing men, especially when taking into account the competitiveness of others. Conversely, men become more competitive in anticipation of decreasing competitiveness of their female contestants. In [Säve-Söderbergh and Sjögren Lindquist, 2017], female juniors employ inferior wagering strategies when randomly assigned to male opponents.

Nevertheless, a series of evidence, ranging from low-stake lab studies to high-stake field experiments suggest otherwise. [Moely et al., 1979; Conti et al., 2001] documented that girls perform better when competing against boys than girls. Most recently, the best-of-five repeated contest by [Mago and Razzolini, 2019] found that women exert significantly higher effort only when competing against other women, while women are just as competitive as men in mixed-gender sessions. In the field, across five sequential elementary math contests, [Cotton et al., 2013] found that the male advantage is at best short-lived, while females even outperform males in later periods. In TV shows, in contrast to findings of [van Dolder et al., 2020], [Jetter and Walker, 2018] and [Antonovics et al., 2009] found that women are more competitive when facing men in the US *Jeopardy!* version and the high-stake rounds of the *Weakest Link* show, respectively. The closest work to my paper is [De Paola et al., 2015] on midterm exam performance of Italian students competing in pairs of equal predicted ability but different gender composition. Similar to their work and [Mago and Razzolini, 2019], I find that on average, the performance of *neither* men or women is affected by the gender composition of opponents.

Secondly, this research expands the empirical evidence on real-world contest literature with a piece of novel evidence in high-stake debate tournaments. To the best of my knowledge, other than school exams or TV shows, empirical studies on the gender differences in competitive performance are mostly restricted to one-on-one settings e.g. expert chess or tennis tournaments. Specifically, in chess tournaments, [Backus et al., 2016], [Dilmaghani, 2020] and [Gerdes and Gränsmark, 2010] consistently confirm that conditional on ELO ratings, the gender composition effect is driven by women performing worse against men, rather than by men playing better against women. Furthermore, the largest gender performance gap is among elite players. Comparatively, in debate tournaments, I find supporting evidence for a larger gender gap in higher-ranked debates. Yet, in contrast to chess tournaments, female debaters fare comparatively worse when facing more female opponents in higher-ranked debates. In same-sex only tennis tournaments where [Wozniak, 2012] studied the tournament entry decision given relative past performance feedback, he found that such information feedback has gender-specific effects. Since recruitment or promotion decisions in firms are often drawn on a pool of similarly able candidates across multiple rounds, insights from these mixed-sex, multi-round debate competitions, where participants compete head-to-head based on previous rankings, are more relatable to real-life competitions.

Finally, the literature on team gender composition and performance provides possible mechanisms to explain the descriptive result of the concentrated gender score gap among female speakers in female-only teams, and not those in mixed-gender teams. Since participants compete in their chosen teams of two, my descriptive finding that women-only teams perform worse than mixed-gender or male-only teams is in line with the observational study of [Apesteguia et al., 2012] in high-stake, online business game contests. [Dargnies, 2012] offers a likely explanation for this overall gender score gap, based on differential self-selection: low-performing women are more likely to enter tournaments with similar others in two-person tournaments. The descriptive finding that male-dominated teams perform similarly to mixed-gender teams is also in line with the causal result in the larger 12-person business team field experiment by [Hoogendoorn et al., 2013].

This chapter proceeds as follows. Section 4.2 summarizes the debate competition setup. Section 4.3 provides data set overview and summary statistics, followed by empirical strategies in Section 4.4. Section 4.5 highlights the main results, with extension findings on higher- vs. lower-ranked debates and competition performance given teammate's gender choice in Section 4.6. Section 4.7 concludes with discussions on future research avenues.

4.2 Institutional Setup

Tournament Format. Participants in these tournaments are undergraduate or graduate students who are active and dedicated in their respective debate societies. Debaters participate in weekly meetings and travel to various local and international tournaments to sharpen their debate skills. Every year, around 200+/- two-person-teams across Europe attend the European Universities Debate Championship (EUDC); 450+/- teams across the world participate in the World Universities Debate Championship. They represent their institutions to compete across nine preliminary rounds (i.e *in-rounds*) with exogenously assigned controversial topics, speaking positions, judges, and opponents in every round. All debates are conducted in British Parliamentary (BP) Debate style.[5] After each round, a panel of judges submit two results of each individual to the score tabulation organizer:

[5] For more details on BP debate style and format, please check Appendix 4.9.

(1) team ranking [6] and (2) individual speaker scores.[7] Within a debate, individual speaker scores must reflect the ordinal team ranking i.e. the cumulative score of two speakers whose team ranked first must be higher than that of the team ranked second. The total team points and speaker points[8] across all preliminary rounds determine the top $10 - 15\%$ performing teams to enter elimination rounds (i.e *out-rounds*)[9]. Since evaluation scores are only given in preliminary rounds, this research focuses exclusively on these rounds, and not the out-rounds.

Team Matching & Performance Feedback Mechanism. Every debate consists of four teams. In Round 1, team matching is unconditionally randomized. From Round 2 onward, teams are power-matched i.e. in N^{th} round, teams debate teams with similar cumulative team *and* speech evaluation points from $(N - 1)$ rounds.[10] In other words, *within* each team score bracket, the teams with the highest speaker points in $(N - 1)$ rounds will meet one another in N^{th} round. Hence, the universal individual speech score scale aims to ensure consistent evaluation across rooms.[11] Regarding performance feedback to speakers, from Round 1 to Round 6 (*open rounds*), teams receive only their team ranking results and relative performance feedback after the debate and judge deliberation discussions. From Round 7 to Round 9 (*closed rounds*), no results are communicated to speakers right after the debate. Once all elimination rounds are completed, speakers receive team ranking results and feedback from judges. Finally, speakers will receive the public results of their evaluation scores across rounds when the tournament ends.

Judge Allocation Mechanism & Fairness. Every tournament has an appointed Chief Adjudicator (CA) team of four to six internationally accomplished debaters who are in charge

---

[6] i.e. team that ranks 1^{st} gets 3 points, 2^{nd} gets 2 points, 3^{rd} gets 1 point and 4^{th} gets no point.

[7] 50-to-100 score scale, with 50 as the lowest. See Appendix 4.3 for a speaker score scale example of European Universities Debate Championship 2017.

[8] Speaker points are used for: (i) award best performing speakers in the form of top 10 speaker awards; and (ii) determine teams advancing to elimination rounds in case of ties.

[9] In these rounds, teams that are ranked 1^{st} and 2^{nd} advanced into further rounds, whereas those on 3^{rd} and 4^{th} place are eliminated. In the final debate, the best team becomes the champion. The best speaker of the tournament is an individual with the highest cumulative individual speech scores across all preliminary rounds.

[10] Note that *within* a debate, speech evaluation points must reflect team rankings i.e. the cumulative speech scores of two speakers whose team ranked first must be higher than that of the team ranked second. For more information about power-pairing, see this discussion thread on Monash Debate Review.

[11] i.e. winning in a lower-ranked room does not necessarily mean higher individual speaker scores than, for instance, taking a 2^{nd} or 3^{rd} in a higher ranked room

of judge recruitment, quality screening, monitoring, and overall panel allocation throughout the tournament. Three mechanisms are set in place to ensure fairness in judgment across rounds. First, no judges who come from the same institutions, in the past or present, as any debaters in the room can be allocated to judge that debate. Second, before the competition, judges and debaters are required to disclose any potential conflicts with other participants.[12] Third, the intensive nature of a 3-day, 9-round competition makes it difficult for any strategic collusion to be formed between judges, the CA team, and speakers from different institutions. Appendix 4.10 provides more details on the judge's tasks, check-and-balance feedback mechanism, and adjudication procedure throughout these tournaments.

4.3 Data & Descriptive Statistics

4.3.1 Data

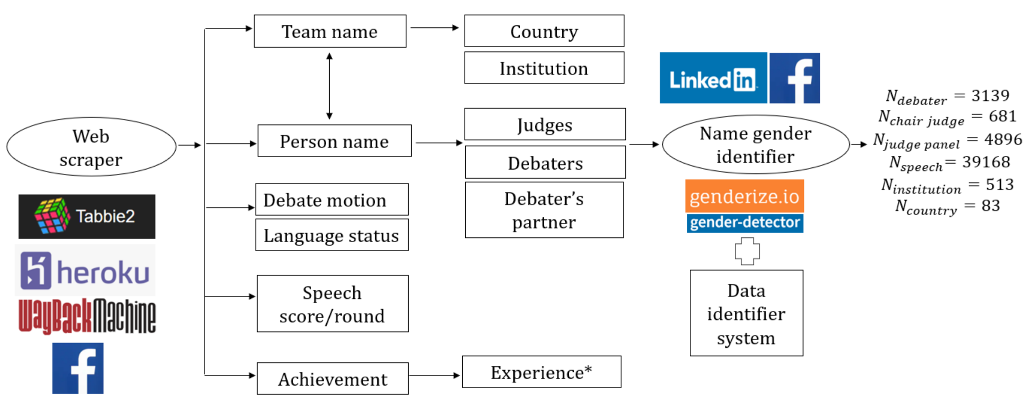

Figure 4.1: Overview: Data collection and construction procedure

This section describes the data set construction procedure and key descriptive statistics of speakers and judges. Figure 4.1 illustrates the entire data collection process. In general, names of individuals, judges and institutions, the roles of judges[13] and opponents;[14] individual evaluation scores for every debate, language skill status of speakers and debate

[12]Legitimate clashing reasons include, among others, close friendship/partnership, past romantic encounters, or negative experiences. To disincentivize strategic clashing, an independent committee conducts confidential interviews with the requested persons to verify their reasons.
[13]i.e.chair judge, wing judge and trainee judge.
[14]i.e. Opening Government, Opening Opposition, Closing Government, Closing Opposition

motions are available from tabulated archival sources.[15] Detailed data collection procedure on other control variables such as judge panels, debate topics, language skills and institutions is provided in Appendix 4.11. Section 4.3.2 then gives descriptive statistics on score differentials across speakers given their characteristics.

4.3.1.1 Outcome Variable: Speech Scores

The main outcome variable is the speech evaluation scores given by the adjudication panel for every debate speech. Across the total of 5081 debates from eight competitions, 185 debates are omitted due to missing identity of speakers or speech scores. These are because of either of the following reasons: (i) swing speakers (i.e. last-minute fill-in volunteers in case speakers cannot speak); (ii) speakers who redacted their identity after the tournament; (iii) one speaker spoke for both roles, since the other speaker excused him/herself from speaking in that respective round.[16] Since full information about gender composition of speakers and judges is crucial for the analysis, all rooms with at least one of such issues are omitted. This procedure results in 39168 speeches across 4896 debates, which is documented per competition in Appendix 4.5, along with omitted debates per competition.

4.3.1.2 Speakers

Full names of speakers[17] and matching their identities across the years is done given the tabulation tournament data archive in Tabbie2 and Tabbycat. To avoid discretionary personal judgment as much as possible, we used a conservative method: a person is considered a duplicate only if their name, institution, EUDC language status, and WUDC language status are the same. Next, to identify gender of speakers, I ran gender inference algorithms: gender guesser and genderize.io[18] on their first names. Both algorithms return the most likely gender, given its hand-coded data label, and a frequency count of such names in their database as male or female.[19] This procedure results in 89.23% of names

[15] Such data is released given the consent of speakers and judges, unless otherwise redacted, in which case they are omitted from the sample.

[16] In this case, the missing speaker receives 0 point, whereas his/her partner who gave both speeches receive the higher score of the speeches he/she gave.

[17] We first clean out: strange characters from non-English names, reversed first and last names, abbreviated names are properly restored across tournaments by matching with their institutions and social media profile (where applicable).

[18] This API contains 216286 distinct names across 79 countries and 89 languages

[19] For a comparison of features and performance of different gender inference algorithms, please refer to the report of [Menéndez et al., 2020] and [Santamaría and Mihaljević, 2018].

assigned gender with certainty. The remaining 10.77% names, which consist of mostly African, South Asian, Israeli, and Eastern European names, were manually checked using social media.

Altogether, after omitting 27 unisex names without any social media sources and possible confirmation from tab masters, we have $N = 3153$ unique speakers for analysis: whereby $N_{MaleSpeaker} = 1949$ and $N_{FemaleSpeaker} = 1190$. Figure 4.5 shows the proportion of speakers by gender for each competition. Across all competitions, female speakers account for 35% to 41% of all participants. Furthermore, the world map distribution of speakers given their gender in Figure 4.6 shows that most countries sent disproportionately more male speakers than female speakers, except for China/Hong Kong. The US, UK, and Australia sent the highest number of speakers, understandably so, given their established debate training culture and civic participation.

4.3.2 Descriptive Statistics

Relationship between Gender of Speakers, Opponents, and Room Characteristics. Table 4.8 provides a comprehensive breakdown of the proportion of speeches by male and female speakers, while Figure 4.8 gives the Spearman correlation coefficient heatmap across various characteristics of speakers, debate room, and judges. Table 4.8 shows that there does not appear to be any differences in terms of the proportion of speeches by male vs. female speakers across these characteristics. Most importantly, Figure 4.8 shows no correlation between the number of female opponents and any observable characteristic, including the speaker's gender.[20] Apart from a very mild positive correlation between the speaker's gender and the gender of their chosen debate partner, there is virtually no relationship between the speaker's gender and other characteristics. Regarding the distribution of female opponents in a debate (excluding partner's gender), Figure 4.5 notes that speakers face only one to three female opponents, thus reflecting the male-dominated nature of competitive debate tournaments.

Speech Scores: Male vs. Female Speakers. Table 4.6 reports the descriptive statistics of scores across all tournaments. The t-test statistics in Table 4.7 and the kernel density

[20]This is confirmed in the Spearman's correlation coefficient test between speaker's gender and their opponents in a room (excluding debate partner), with $\rho = -0.0015$ and $p = 0.768$).

of speech scores in Figure 4.12 show that male speakers scored slightly higher than female speakers. This pattern holds regardless of whether it is male- or female-dominated debates,[21] the language skill statuses or whether the speaker belongs to top 50-ranked institutions. Across rounds, Figure 4.9 plotting the mean standardized evaluation scores of men vs. women shows persistently lower scores from women than men, except for Round 3s.

Speech Score: Higher- vs. Lower-ranked Debates. At any given N^{th} round (except for Round 1s), I split the sample based on the *median* average cumulative $(N-1)$ round speech scores of *two speakers in a team*. Specifically, higher-ranked debates are those where the score is higher than or equal to the *median* speech score and vice versa. Comparing speeches of male and female speakers in higher- vs. lower-ranked debates in Figure 4.14 shows notably lower scores of female speakers only in higher-ranked debates. A further breakdown given partner's gender in the histogram and kernel distribution in Figure 4.17 found slightly higher scores for women in mixed-gender teams than those in women-only teams, yet the pattern is more pronounced in lower-ranked debates.

Speech Score: Team Gender & Round Dynamics. Since speakers choose their respective partners to enter the tournament together, this subsection gives some descriptive graphs on score differentials across rounds given the team gender composition. Figure 4.11 gives a descriptive overview of the average speaker's score across rounds across male-only, mixed-gender, and female-only teams. We note that the gender score gap found in Figure 4.9 is predominantly driven by female speakers in female-only teams. For women in mixed-gender teams, compared to their male partner, except for Round 5s and 7s where they scored on average lower than their male partner, in the rest of the rounds, they either scored similarly or slightly higher than their partner.

4.4 Empirical Strategies

To understand whether the gender composition of opponents affects speech performance of male and female speakers, I run linear and fixed effects regression on standardized speech

[21] see the histogram in Figure 4.13.

score, interacting the indicator variable of speaker's gender with the number of female opponents in the debate, as shown below:

$$S_{sk} = \alpha \mathbb{I}_{FemS} + \theta \mathbb{I}_{FemO} + \beta \mathbb{I}_{FemS} \mathbb{I}_{FemO} + \sum_{i=1}^{n} \gamma \mathbf{Y}_{sk} + \eta_s + \varepsilon_{sk}$$

The dependent variable S_{sk} is the standardized evaluation score of the speech of speaker s in debate k. The coefficients of interest are θ and β, where θ measures any significant relationship between speech performance of male speakers and the number of female speakers in debate k, and β checks for any significant differences between male and female speakers therein.

$\mathbb{I}_{FemS}$ is the gender of the speaker, whereas $\mathbb{I}_{FemO}$ refers to the number of female opponents for speaker s in debate k. Given the male-skewed distribution of speakers in a room shown in Figure 4.7, I use both a linear specification of $\mathbb{I}_{FemO}$, where $\mathbb{I}_{FemO}$ is the number of female opponents, and a non-parametric specification, where I add dummy variables for each possible number of female opponents i.e. $\mathbb{I}_{FemO} \in \{0, 6\}$. ε_{sk} is the error term of the speech given by speaker s in debate k. Throughout all analyses, standard errors are clustered at debate level.

Speaker fixed effect η_s is included to take care of any unobserved heterogeneity on the speaker's characteristics. Other control variables $\mathbf{Y}_{sk}$ are as follows:

1. η_J is the chair judge fixed effect.[22]
2. language skill level (non-native or native English speaker).
3. institution ranking group (i.e. whether if the speaker represents a top-50-ranked institution).
4. gender of speaker's debate partner.
5. group competition type (EUDC or WUDC).
6. Speaking position (1^{st} to 8^{th}) in any given debate.

[22] Since chair judges have decisive power in determining the team and speaker outcomes in a debate, this fixed effect captures unobserved heterogeneity on chair judge's characteristics.

7. Motion topic type (17 topics) in any given debate.[23]

8. Debate round (1 to 9) for any given debate.

9. whether the majority of wing judges are women.

In debate tournaments, the power-matching mechanism makes teams debate teams with a comparable cumulative performance from previous rounds, starting from Round 2 onward. Therefore, as an attempt to control for the average team standing from previous rounds i.e. selection effect on the interested variable, I include in some regression analyses the average cumulative speech scores over $(N-1)$ rounds of *two speakers in a team* in the analysis of N^{th} round, for Round 2s to Round 9s.

4.5 Results

4.5.1 Overall

Column (1) of Table 4.1 shows that, unconditionally, female speakers get 5.6 percentage point (p.p) standard deviation (SD) lower scores compared to male speakers. On average, an additional female opponent is associated with a reduction of 2.0 p.p SD in speech scores, as noted in Column (2). Columns (3) and (4) show that there is no difference between male and female speakers in the relation between the number of female opponents and speech scores.

[23] See Figure 4.4 for the list of motions.

Table 4.1: Regression Analysis: Gender of Speakers and Opponents (N = 39 168)

	Dependent Variable: Score (standardized)					
	(1)	(2)	(3)	(4)	(5)	(6)
Female Speaker	-0.056***		-0.053**	-0.051***		
	(0.01)		(0.02)	(0.02)		
Number of Female Opponents		-0.020**	-0.020**	-0.013*	0.002	0.004
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Female Speaker × Number of Female Opponents			-0.001	-0.001	-0.003	-0.002
			(0.01)	(0.01)	(0.01)	(0.01)
Speaker Controls				✓		
Room Controls				✓		✓
Speaker FE					✓	✓
R^2	0.001	0.001	0.001	0.318	0.568	0.642
Observations	39168	39168	39168	39168	39157	39157

Speaker controls include: (i) language skill status and (ii) institution ranking. Room controls include: (i) chair judge fixed effect, (ii) debate partner's gender, (iii) wing gender composition, (iv) speaking position, (v) round, (vi) motion type, (vii) competition & year. Robust clustered standard errors at debate level in parentheses.
R^2 of model (5) and (6) is $R^2_{between}$. Singleton observations are dropped in model (5) and (6).
$^{*}p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Noteworthily, the score gap in rooms with more or fewer female opponents vanishes upon controlling for speaker fixed effect, as seen in Columns (5) and (6). The difference in the estimated effect of the number of female opponents between column (2) to (4) and Column (5) to (6) suggests that the relationship in Columns (2) to (4) is driven by a selection effect. In other words, within individuals, the speech performance of neither men nor women responds to the number of female opponents in the room. Therefore, the effect in columns (2) to (4) is across individuals, and potentially due to the fact that female speakers perform slightly worse. Over time, because of the power-matching mechanism, gender segregation occurs, i.e. more women cluster to lower-ranked debates in later rounds compared to earlier rounds.

Next, Table 4.2 reports the non-parametric regression results of speaker's gender on the number of female opponents in the debate against speech scores. Column (2) shows the unconditional score difference across debates given the number of female opponents that a speaker faces. Compared to debates where speakers face no female opponents, speakers in debates with only one female opponent received 5.8 p.p. SD higher scores. As the number of female opponents increases, we noted a negative, yet insignificant speaker score gap between such debates and debates with no female opponents. Yet, given the limited number of rooms with 5 or 6 female opponents (see Figure 4.7), it is difficult to draw conclusions from these numbers. Importantly, at the speaker's fixed effect level, Column (5) shows that there is no difference in the association between the number of female opponents and speech performance of both male and female speakers. This is consistent with the analysis using a continuous specification of the number of female opponents above.

Table 4.2: Regression Analysis : Gender of Speaker and Opponents (N = 39 168)

	Dependent Variable: Score (standardized)					
	(1)	(2)	(3)	(4)	(5)	(6)
Female Speaker	-0.056***		-0.052	-0.014		
	(0.01)		(0.05)	(0.03)		
1 Female Opponent		0.058*	0.059	0.051	0.000	-0.000
		(0.03)	(0.04)	(0.03)	(0.03)	(0.02)
2 Female Opponents		0.054	0.053	0.040	-0.005	0.008
		(0.03)	(0.04)	(0.03)	(0.03)	(0.02)
3 Female Opponents		0.022	0.029	0.027	-0.001	0.013
		(0.04)	(0.04)	(0.04)	(0.03)	(0.03)
4 Female Opponents		-0.052	-0.052	-0.028	0.013	0.015
		(0.04)	(0.05)	(0.04)	(0.03)	(0.03)
5 Female Opponents		-0.072	-0.074	-0.030	-0.013	-0.020
		(0.06)	(0.06)	(0.05)	(0.04)	(0.04)
6 Female Opponents		-0.120	-0.136	-0.059	0.119	0.121*
		(0.10)	(0.13)	(0.11)	(0.08)	(0.07)
Female Speaker × 1 Female Opponent			-0.001	-0.040	-0.019	-0.013
			(0.05)	(0.04)	(0.04)	(0.03)
Female Speaker × 2 Female Opponents			0.001	-0.050	-0.021	-0.028
			(0.05)	(0.04)	(0.04)	(0.03)
Female Speaker × 3 Female Opponents			-0.019	-0.053	-0.015	-0.021
			(0.05)	(0.04)	(0.03)	(0.03)
Female Speaker × 4 Female Opponents			0.003	-0.012	-0.033	-0.018
			(0.06)	(0.05)	(0.04)	(0.03)
Female Speaker × 5 Female Opponents			0.004	-0.052	-0.007	0.010
			(0.08)	(0.06)	(0.06)	(0.05)
Female Speaker × 6 Female Opponents			0.038	0.147	-0.015	-0.051
			(0.16)	(0.14)	(0.15)	(0.10)
Speaker Controls				✓		
Room Controls				✓		✓
Speaker FE					✓	✓
R^2	0.001	0.002	0.002	0.318	0.568	0.642
Observations	39168	39168	39168	39168	39157	39157

Speaker controls include: (i) language skill status and (ii) institution ranking. Room controls include: (i) chair judge fixed effect, (ii) debate partner's gender, (iii) wing gender composition, (iv) speaking position, (v) round, (vi) motion type, (vii) competition & year. Robust clustered standard errors at debate level in parentheses.
R^2 of model (5) and (6) is $R^2_{between}$. Singleton observations are dropped in model (5) and (6).
$^{*}p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.5.2 Round 1s vs. Round 2s to 9s

Table 4.3 provides the regression analysis using a continuous specification of the number of female opponents, among Round 1s and those of Round 2s to 9s. The split is because the power-matching mechanism of teams is applied from Round 2 onward, whereas in Round 1s, the allocation is unconditionally random. In an attempt to capture the selection effect of the power-matching mechanism, the final rows in Table 4.3 control for the average team standing of $(N-1)$ round from Round 2s to Round 9s, in the analysis of N^{th} round at a particular tournament. For Round 1s, no speaker fixed-effect model applies since there is only one observation per speaker for every tournament, and within-speaker comparisons across Round 1s of different tournaments are not comparable to other round analyses.

Overall, no significant difference in speech performance of men and women given the number of female opponents, which is consistent with the findings in Table 4.1. For Round 1s where room allocation is unconditionally random, Column (2) shows a similar, albeit insignificant, unconditional score gap of 1.9 p.p SD for speakers who face more female opponents, to the overall finding in Table 4.1. It is important to note that upon controlling for speaker and room characteristics and interacting the gender of speakers with the number of female opponents, I find a significant and negative relationship i.e. speakers who face more female opponents get 4.2 p.p SD lower scores.

Comparing the gender speech score gap between Round 1s and Round 2s to 9s, Column (1) shows that this gap from 8.3 p.p SD in Round 1s to only 5.3 p.p. SD across Round 2s to 9s. Upon controlling for team standing, this gap remains significant but shrinks to 3.0 p.p SD. This pattern illustrates the functioning power-matching mechanism, whereby teams of comparable ability compete against one another. Regarding the number of female opponents, Column (2) shows an unconditional score gap of 2.0 p.p SD for speakers who face more female opponents in Round 2s to Round 9s. Upon interacting speaker's gender with the number of female opponents, Column (3) finds that this relation is similar between male and female speakers, yet it vanishes upon controlling for speaker and debate room characteristics and speaker fixed effects. Noteworthily, once team standing is taken into account, speakers who face more female opponents get 4.0 p.p SD lower scores. A qualitatively similar gap of 3.5 p.p SD remains upon interacting with speaker's gender shows up, only to disappear upon further controls in Columns (3) to (5).

Table 4.3: Regression Analysis: Gender of Speaker and Opponents, Round 1s vs Round 2s - 9s (N = 39 168)

	Dependent Variable: Score (standardized)					
	(1)	(2)	(3)	(4)	(5)	(6)
Round 1s						
Female Speaker	-0.083***		-0.061	-0.085		
	(0.03)		(0.05)	(0.05)		
Number of Female Opponents		-0.019	-0.014	-0.042*		
		(0.01)	(0.02)	(0.02)		
Female Speaker × Number of Female Opponents			-0.009	-0.001		
			(0.02)	(0.02)		
R^2	0.002	0.001	0.003	0.361		
Observations	4376	4376	4376	4376		
Round 2s to 9s						
Female Speaker	-0.053***		-0.054**	-0.045**		
	(0.01)		(0.03)	(0.02)		
Number of Female Opponents		-0.020**	-0.021**	-0.009	0.005	0.006
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Female Speaker × Number of Female Opponents			0.000	-0.001	-0.003	-0.003
			(0.01)	(0.01)	(0.01)	(0.01)
R^2	0.001	0.001	0.001	0.337	0.572	0.648
Observations	34792	34792	34792	34792	34786	34786
Round 2s to 9s (controlled debate room quality)						
Female Speaker	-0.030***		-0.005	-0.025		
	(0.01)		(0.02)	(0.02)		
Number of Female Opponents		-0.040***	-0.035***	-0.009	0.005	0.006
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Female Speaker ×Number of Female Opponents			-0.012	-0.006	-0.003	-0.003
			(0.01)	(0.01)	(0.01)	(0.01)
R^2	0.171	0.173	0.173	0.387	0.572	0.648
Observations	34792	34792	34792	34792	34786	34786
Speaker Controls				✓		
Room Controls				✓		✓
Speaker FE					✓	✓

Speaker controls include: (i) language skill status and (ii) institution ranking. Room controls include: (i) chair judge fixed effect, (ii) debate partner's gender, (iii) wing gender composition, (iv) speaking position, (v) round, (vi) motion type, (vii) competition & year. Robust clustered standard errors at debate level in parentheses. R^2 of model (5) and (6) is $R^2_{between}$. Singleton observations are dropped in model (5) and (6). $^{*}p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$.

Table 4.4 investigates speech performance given the number of female opponents across Round 2s to Round 9s separately, controlling for team standing, speaker, and judge characteristics. We note that in Round 3s and 4s, female speakers facing more female opponents get comparatively lower scores compared to male speakers, at 3.5 and 5.0 p.p SD, respectively. As the rounds and the power-matching mechanism progress, no significant difference in speech performance of male and female speakers given the number of female opponents is detected. Overall, the changes in magnitude could be due to the varying distribution of the number of female opponents across rounds. A non-parametric estimation of the effect of the number of female opponents in Round 1s of Table 4.9, and Round 2 to 9s in Table 4.10 and 4.11 (controlling for average team standing) in Appendix 4.13.6 show consistent findings with those in Table 4.3.

Table 4.4: Round-by-round regression Gender of Speaker and Opponents, controlling for room quality, speaker and judge characteristics (N = 34 792)

	Dependent variable: Score (standardized)							
	R2s	R3s	R4s	R5s	R6s	R7s	R8s	R9s
Female Speaker	-0.094	0.093*	0.119**	-0.058	0.010	-0.093*	0.065	0.030
	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)
Number of Female Opponents	0.021	0.028	-0.016	0.002	-0.053**	0.007	0.015	0.041
	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)
Female Speaker × Number of Female Opponents	-0.000	-0.035*	-0.050**	-0.019	-0.011	0.032	-0.020	0.005
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
R^2	0.421	0.531	0.519	0.561	0.585	0.592	0.609	0.679
Observations per round	4408	4376	4304	4320	4352	4336	4344	4352

All models control for team standing (i.e. average cumulative speech scores of two speakers in a team up to the respective round.). Speaker controls include: (i) language skill status and (ii) institution ranking. Room controls include: (i) chair judge fixed effect, (ii) debate partner's gender, (iii) wing gender composition, (iv) speaking position, (v) round, (vi) motion type, (vii) competition & year.
Robust clustered standard errors at debate level in parentheses.
$^{*}p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$.

4.6 Extensions

4.6.1 Higher vs. lower-ranked debates

Given that the stakes are significantly higher in higher-ranked debates, for every N^{th} debate after Round 1, I split the sample based on the *median* average cumulative $(N-1)$ round speech scores of *two speakers in a team*. Since the power-matching mechanism in these tournaments matches teams with not only similar team records but also speech performance records, this split provides a good approximation of team standings in any respective N^{th} round. I then run the regression analysis in Table 4.12. For lower-ranked debates, controlling for average debate standing, I find no significant gender score gaps. An additional female opponent is associated with a reduction in speaker score of 2.2 pp SD. However, this score gap vanishes upon interacting with the speaker's gender and controlling for relevant speaker and room characteristics.

In contrast, among higher-ranked debates, female speakers get 3.1 p.p SD lower scores than their male counterparts. With respect to the number of female opponents, Column (2) shows that speakers who face more female opponents receive 2 p.p SD per additional female opponent. Upon interacting with the speaker's gender and controlling for speaker, judge, and room characteristics, I find that male speakers perform similarly when there are more or fewer female speakers. However, female speakers get significantly lower scores when faced with more female opponents. At the speaker fixed effect level controlling for relevant characteristics, one more female opponent leads to 2.2 p.p SD lower scores for female speakers, compared to their male counterparts.

To better understand the relationship between the speaker's gender and the specific number of female opponents, Table 4.13 gives the non-parametric regression analysis in higher-ranked debates. Column (2) shows that, compared to speakers who face no female opponents, speakers facing 4 and 5 female opponents get 11.1 p.p and 10.8 p.p SD lower scores respectively. Upon interacting speaker's gender with the number of opponents, Column (3) to (6) show that the number of female opponents only affect female speakers, but not male speakers. Specifically, compared to male speakers facing 1 to 4 female opponents, female speakers facing the same number of opponents receive robustly lower scores. Given the limited number of female speakers facing 5 or 6 female opponents in higher-ranked debates ($N_{5female} = 181, N_{6female} = 11$), there is insufficient power to draw firm conclusions from these results.

4.6.2 Speaker gender ft. partner's gender

Given the difference in speech performance between male-only (MM), mixed-gender (MF), and female-only teams (MF) shown in Figure 4.11, to better understand the relationship between speaker's performance given their partner's gender choice and the number of female opponents, Table 4.14 reports the regression analysis with a continuous variable of the number of female opponents. In this case, since teams are fixed within a tournament, models with speaker fixed effects estimate *across* tournaments. Therefore, the results in these models in Column (5) and (6) of Table 4.14 and 4.15 should be interpreted as cross-tournament estimation. Since many speakers compete in multiple tournaments with varying partner's choices, such analysis gives an insight into the overall scoring patterns across tournaments on the basis of the partner's gender choice.

Across all rounds, an important overall finding is that the gender score gap is mainly observed among female speakers in female-only teams; but the gender composition of opponents in a debate largely plays no role. A split analysis of Round 1s and Round 2s to 9s in Table 4.15, whereby the latter rounds controls for average cumulative team standings, find similar results across all rounds. An important note here is that these findings only serve as *descriptive evidence* on speaker's performance given their team gender composition because debate partner's choice is endogenous.

4.7 Conclusion

This chapter contributes a causal finding on the impact of the gender composition of opponents on the competitive performance of 3153 debaters in highest-profile debate tournaments. The multi-dimensional, complex debate task together with the multi-round, power-matching mechanism in these tournaments adds a piece of useful empirical evidence to the contest and gender competitive performance dynamics. The key finding is that the performance of *neither* male nor female debaters is affected by the gender composition of opponents. In higher-ranked debates, women perform comparatively worse when facing more *female* opponents. Descriptively, the raw gender score gap is mainly found among women in female-only teams, not those in mixed-gender teams. Overall, if these findings carry over to other real-life settings, they indicate that having more women competing for high-profile careers and positions does not necessarily reduce the thickness of the glass ceiling.

Three limitations potentially restrict the generalizability of these results. First, since the power-matching mechanism is a known feature among participants in debate tournaments, they have some certainty over the previous performance records of their opponents. In other real-world contexts, beliefs about the performance or ability of opponents are possibly biased or dependent on the gender of opponents, which in turn may affect one's own performance. Second, participants who select themselves into these debate competitions are young, talented university students with significantly public speaking training and international exposure. Since these competitions are held Europe-wide or worldwide, it is not directly applicable to local competition contexts. Third, while this paper can study the causal impact of the gender composition of opponents on individual speech performance, endogenous team formation prevents any causal interpretation on how within-team gender composition interacts with the gender composition of opponents. As teams train and prepare speeches together, I cannot disentangle whether women who select into mixed-gender teams have a higher innate ability, or simply that their team dynamics differ from that of female-only teams. Finally, to enrich the analyses and descriptive evidence on partner's choices, future research can take into account previous debate experiences of participants and the progression of different team gender compositions in elimination rounds.

4.8 Appendix

4.9 British Parliamentary Style Debate Format

British Parliamentary (BP) is the most widely adopted debate format in top-tier intercollegiate competitions worldwide.[24] BP debate topics relate to a broad range of current issues in politics, animal rights or social justice. With respect to motion types, the motion is either a policy which changes the status quo (e.g. This House Would Provide All Police Officers With Firearms) or a statement, the truth or falsehood of which is examined in the debate (e.g. This House Regrets the Decline of Marxism in Western Liberal Democracies).[25]

In terms of the debate format, participants enter the competition in fixed teams of two, whereby they will be *randomly allocated* : (i) a TEAM speaking position (Opening Government (OG), Opening Opposition (OO), Closing Government (CG), Closing Opposition (CO)) and (ii) opponent teams, in every debate round. After given a topic, teams are given fifteen minutes to prepare; usage of online resources are prohibited. During preparation time, speakers *within* a team can strategically decide who takes each which roles in their assigned team position. Afterwards, everyone gives a 7-minute speech, sequentially, as shown in Figure 4.2 below:

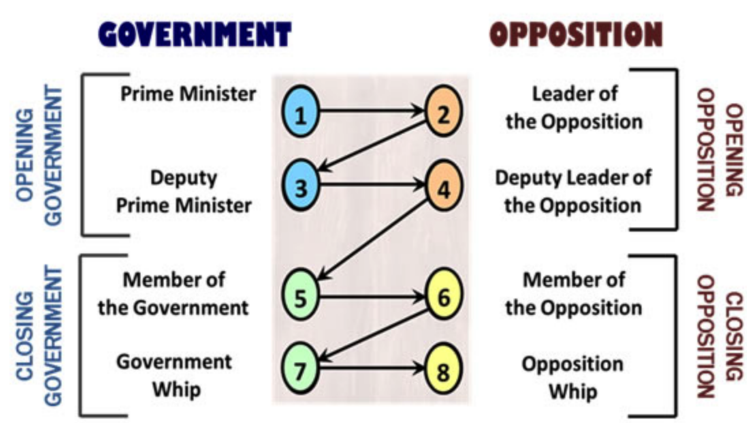

Figure 4.2: British Parliamentary debate speech order

[24]For instance, the World Universities Debating Championship, Pan African Universities Debate Championship and European Universities Debating Championship and numerous regional tournaments in Europe, Canada, United States, Hong Kong, Shanghai, Philippines, Australia, New Zealand and Africa.

[25]For a list of motions and topic pools in debate tournaments, see Hello Motions and European Debating Blogspot

Speech duration is capped at 7 minutes per person. The first and last minute of the speech is *protected*, i.e no opposition teams could offer a point of information (POI). The POI is a formalized interjection of any speaker from the opposite side, which often lasts *no* longer than fifteen seconds. The speaking debater can choose to hear the POI or to dismiss it politely. It is generally considered good practice to accept at least one POI during a speech. After the debate, the judging panel, often consists of highly accomplished debaters, [26] will discuss in 15-20 minutes and decide upon: team ranking (from 1^{st} to 4^{th} place), individual speech evaluation scores and justifications for the ranking decision. In BP debating, speech evaluation refers primarily to the comparative strength of the argument analysis, with respect to logical proofs and rebuttals to substantive materials of opponents.

4.10 Adjudication structure and deliberation rules

A judge's role.[27] Judges need to act as an *informed global citizen*, who evaluate the argumentative cases *holistically*, given their relevance and plausibility. Judging is done comparatively, i.e. decide which team, when weighed against another team, gave the most persuasive case for their side, given one's *impartial* reading of the entire debate. The standard is only on general knowledge, found in the front pages of major articles in the national or international newspapers.[28] A qualified judge must accurately weigh what was *actually said* by teams in the debate, without inserting one's preconceptions or expert knowledge into their decisions.

Adjudication panel structure. A debate room is adjudicated by a *chair (C)* judge (one chair/room) and several *wing (W)* judges. Typically two to four wing judges are allocated in preliminary rounds, whereas four to eight in elimination rounds. All judges are responsible for keeping track of the key arguments and determine the team ranking, speaker scores and justifications thereof. The chair (C) judge has the *ultimate power* and responsibility to assign the definitive ranking,[29] and speaker scores, as well as delivering verbal ranking

[26]A judging panel consists of three or five judges in a round, and anywhere from five, seven or nine judges in elimination rounds, depending on the size of the debate tournament.

[27]For a detailed description of judge's role, please refer to page 4 - 10 of Novi Sad EUDC 2018 Judge Briefing.

[28]For instance, discussing the reparations for WWII, the Iraq conflict or AI ethics would be a fair game, not on the technical or esoteric knowledge about these issues.

[29]In case of ranking conflict, the vote of the chair judge will be the tie-breaker vote.

explanations to debaters after panel discussion.

Deliberation procedure. During the debate, all judges carefully take notes of the speeches and determine their ranking of teams without interacting with one another. Upon making up their mind, judges reveal their decisions to one another. The chair judge then moderates the deliberations, with the unanimous voting rule if time permits. If the panel exceeds the given time, majority voting i.e. 'split' rule kicks in. In case of a tie among wing judges,[30] the chair judge has the tie-breaking vote to determine the final call. Afterwards, chair judges determine the individual speech scores given the team ranking. The chair judge then orally justifies the team ranking decision of the entire panel to competing teams. Noteworthily, speakers only learn their individual scores after the tournament has ended. Finally, all judges and teams can give feedback on one another via confidential forms, which go directly to the Chief Adjudicators and Tab Master of the tournament. These subjective feedback forms the strong basis to determine the highest ranked judges to adjudicate in the elimination rounds.

4.11 Data collection

4.11.1 Judges & Evaluation Panels

Upon scraping from archival tabulation data, we obtained the full names of chair (C), wing (W) and trainee (T) judges for each debate. To determine their identity uniqueness and represented gender, we first sorted the names of all judges per tournament by their function: C, W or T. We temporarily stored names of all judges in a different file to codify gender and deleted their names afterwards. For chair (C) judges, we managed to determine unique identity and gender for everyone, since: (1) they hold the most power in speech evaluation; and (2) their identity is easily tracked given their high-profile statuses and social media presence in debating channels. For wing (W) and trainee (T) judges, I combined results from gender inference algorithms on their first names and information on their affiliated institutions, countries and region. Using the gender inference algorithms similar to with speaker's names, we identified gender of 92% of (W) and (T) judges. For the remaining 8%, which either are: (i) African, South East Asian, Indian and Israeli and gender-neutral names or (ii) conflicting gender assignment, I manually checked them using social media

[30] e.g. one wing gives a second to Team X, and the other gives a fourth to Team X.

connections. Finally, the completed gender list for each judge is confirmed with respective tab directors.

4.11.2 Language skill status

In EUDC/WUDC debate tournaments, individuals are classified into different language categories by an appointed independent language committees. This classification is meant to provide an inclusive playground to speakers with limited exposure to English language, which enables participants to break into open and/or non-native (ESL) speaker's league in knock-out rounds. The evaluation criteria are based on individual survey applications regarding: (i) the age at which they were exposed to English; and (ii) the content, structure and quality of English used for any relevant instruction or exchange.[31] From the archival tab data, I documented 46.65 % of speeches given by non-native English speakers.

4.11.3 Debate topics

Across the 4896 debates, 72 unique debate motions discussed across a wide range of topics. All topics provided a balanced, in-depth but polarized distribution of views, as empirically tested by chief adjudicators in earlier regional competitions. I manually classified these motions into 17 debate topics, based on the classification at International Debate Education Association, which are summarized by the distribution of debate speeches in Figure 4.4. Topics on society, international relations and military policies are the three most popular debate motions at these tournaments, followed closely by debates on the economy, law and justice systems, as well as topics on health, feminism and digital freedom.

4.11.4 Institution & Ranking

Since the academic institution that a speaker represents carries reputation/prestige that could impact evaluations, we collected institution information embedded in team name, in addition to registry data from tab masters, where possible. By pairing up speaker's identity with their team names, along with public social media and confirmation with the tab directors, we obtained 513 distinct institutions across 83 countries in this data set. Since there

[31] For more detailed criteria to be qualified as ESL for EUDC and ESL & EFL for WUDC, see the Language Status section of WUDC constitution and EUDC constitution.

exists no university ranking given their debate achievements,[32] these institutions are categorized by their average academic ranking from QS World Universities Ranking from 2013 to 2017 into two groups: top-50-ranked and the non-top-50-ranked universities. Descriptive statistics table in Appendix 4.13.5 shows that participants affiliated with top-50-ranked institutions account for roughly 10 - 20% of all participants, with the slight exception of WUDC 2017 and WUDC 2018, where this proportion is above 20%. More male speakers tend to represent top-50-ranked institutions and be native English speakers.

[32] Apart from a top 5 and top 10 list of UK & UK universities to master debate skills in the US and UK

4.12 Figures

4.12.1 Example of individual speech score scale (50 -100)

TALLINN EUDC 2017

SPEAKER SCALE[1]

The mark bands below are rough and general descriptions; speeches need not have every feature described to fit in a particular band. Throughout this scale, 'arguments' refers both to constructive material and responses. Please use the full range of the scale. Speaker marks determine many of the breaking teams, and tab finishes can be big achievements, so please give them the serious thought they require.

95-100	• Plausibly one of the best debating speeches ever given; • It is incredibly difficult to think up satisfactory responses to any of the arguments made; • Flawless and compelling arguments.
92-94	• An incredible speech, undoubtedly one of the best at the competition; • Successfully engaging with the core issues of the debate, arguments exceptionally well made, and it would take a brilliant set of responses to defeat the arguments; • There are no flaws of any significance.
89-91	• Brilliant arguments successfully engage with the main issues in the round; • Arguments are very well-explained and illustrated, and demand extremely sophisticated responses in order to be defeated; • Only very minor problems, if any, but they do not affect the strength of the claims made.
86-88	• Arguments engage with core issues of the debate, and are highly compelling; • No logical gaps, and sophisticated responses required to defeat the arguments; • Only minor flaws in arguments.
83-85	• Arguments address the core issues of the debate; • Arguments have strong explanations, which demand a strong response from other speakers in order to defeat the arguments; • May occasionally fail to fully respond to very well-made arguments; but flaws in the speech are limited.
79-82	• Arguments are relevant, and address the core issues in the debate; • Arguments well made without obvious logical gaps, and are all well explained; • May be vulnerable to good responses.
76-78	• Arguments are almost exclusively relevant, and address most of the core issues; • Occasionally, but not often, arguments may slip into: i) deficits in explanation, ii) simplistic argumentation vulnerable to competent responses or iii) peripheral or irrelevant arguments; • Clear to follow, and thus credit.
73-75	• Arguments are almost exclusively relevant, although may fail to address one or more core issues sufficiently; • Arguments are logical, but tend to be simplistic and vulnerable to competent responses; • Clear enough to follow, and thus credit.
70-72	• Arguments are frequently relevant; • Arguments have some explanation, but there are regular significant logical gaps; • Sometimes difficult to follow, and thus credit fully.
67-69	• Arguments are generally relevant; • Arguments almost all have explanations, but almost all have significant logical gaps; • Sometimes clear, but generally difficult to follow and thus credit the speaker for their material.
64-66	• Some arguments made that are relevant; • Arguments generally have explanations, but have significant logical gaps; • Often unclear, which makes it hard to give the speech much credit.
61-63	• Some relevant claims, and most will be formulated as arguments; • Arguments have occasional explanations, but these have significant logical gaps; • Frequently unclear and confusing; which makes it hard to give the speech much credit.
58-60	• Claims are occasionally relevant; • Claims are not be formulated as arguments, but there may be some suggestion towards an explanation; • Hard to follow, which makes it hard to give the speech much credit.
55-57	• One or two marginally relevant claims; • Claims are not formulated as arguments, and are instead are just comments; • Hard to follow almost in its entirety, which makes it hard to give the speech much credit.
50-55	• Content is not relevant; • Content does not go beyond claims, and is both confusing and confused; • Very hard to follow in its entirety, which makes it hard to give the speech any credit.

[1] The scale is consistent with the one used at Warsaw EUDC

Figure 4.3: Tallinn EUDC 2017 Individual Speech Evaluation Score Scale & Description

4.12.2 Distribution of speeches across debate topics

Motion type	Motion type ID
1	Culture
2	Economy
3	Education
4	Law/justice
5	Politics
6	Science
7	Sports
8	Environment
9	Free speech
10	Health
11	Philosophy
12	Religion
13	Society
14	Feminism
15	International
16	Military
17	Digital freedom

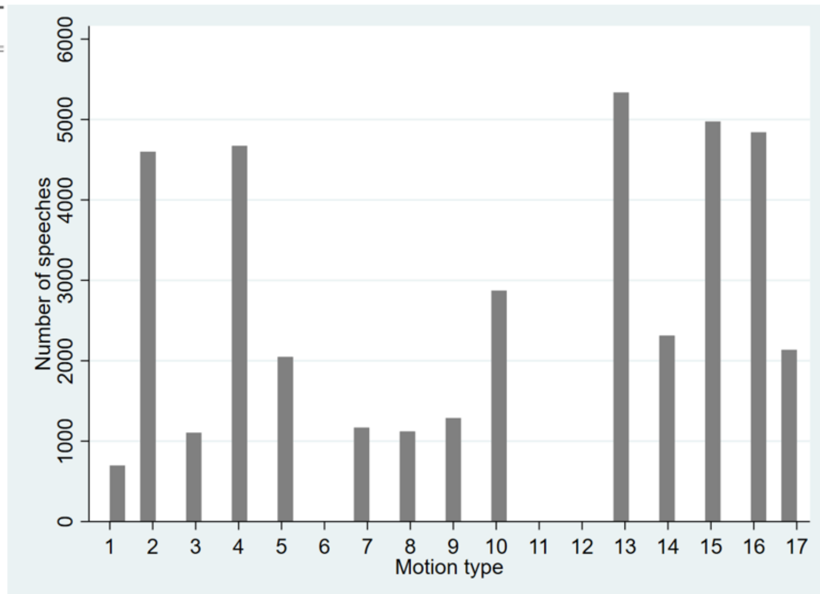

Figure 4.4: Distribution of speeches given motion types

4.12.3 Proportion of male vs. female speakers across competitions

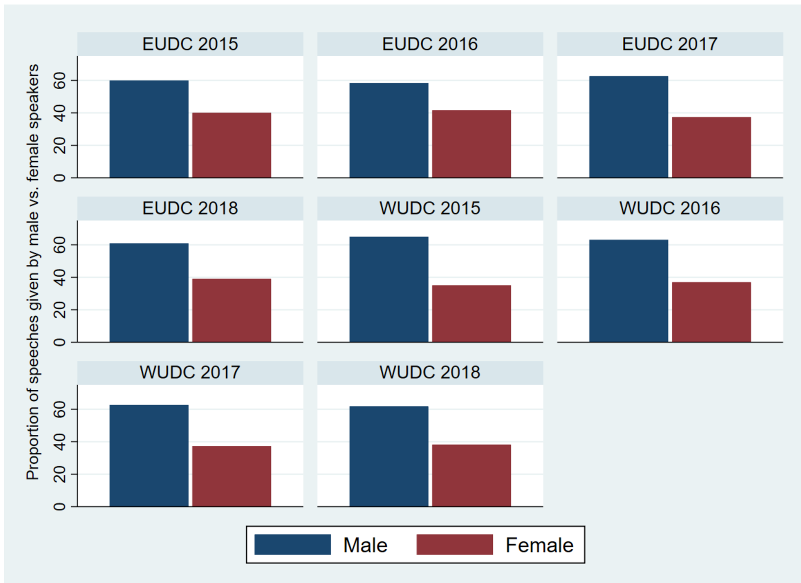

Figure 4.5: Number of male vs. female speakers per competitions ($N_{male} = 24334$, $N_{female} = 14834$

4.12.4 Proportion of male vs. female speakers across institutions

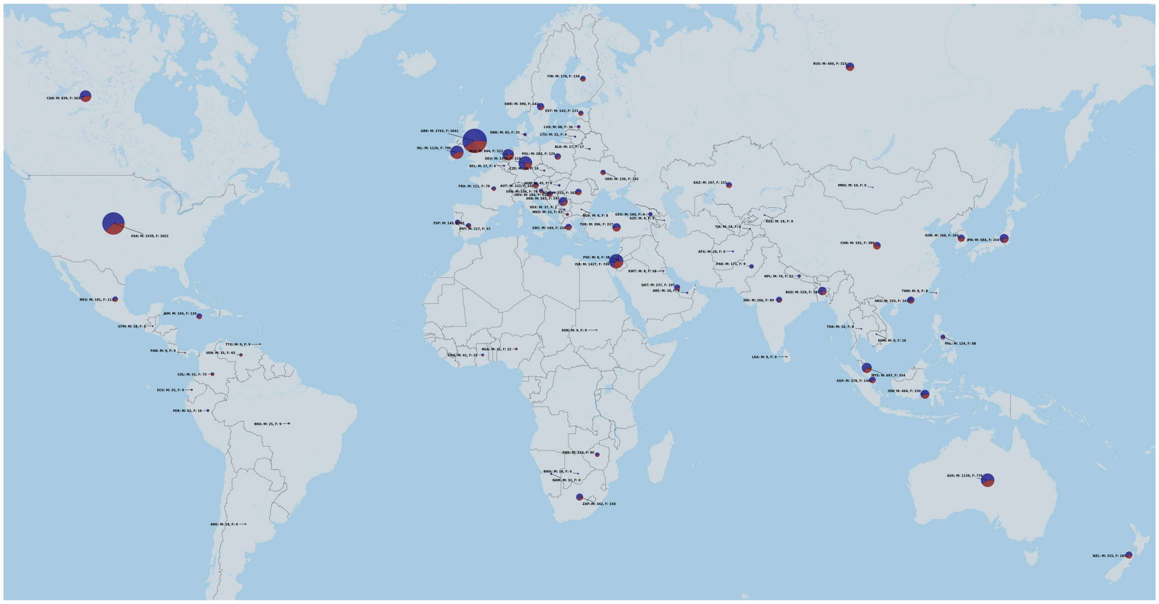

Figure 4.6: Proportion of male vs. female speakers across participating institutions worldwide. The larger the circle, the more participants and institutions from that country represented in the tournaments. Blue refers to male speakers, whereas red refers to female speakers.

4.12.5 Distribution of female opponents in a debate

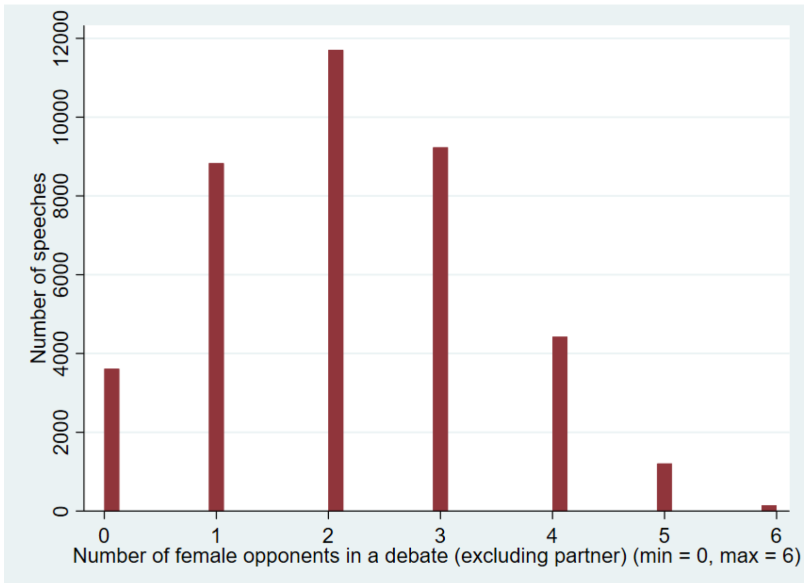

Figure 4.7: Number of female speakers in a debate room

4.12.6 Spearman's correlation coefficient heat map across characteristics of speakers, debate room and judges

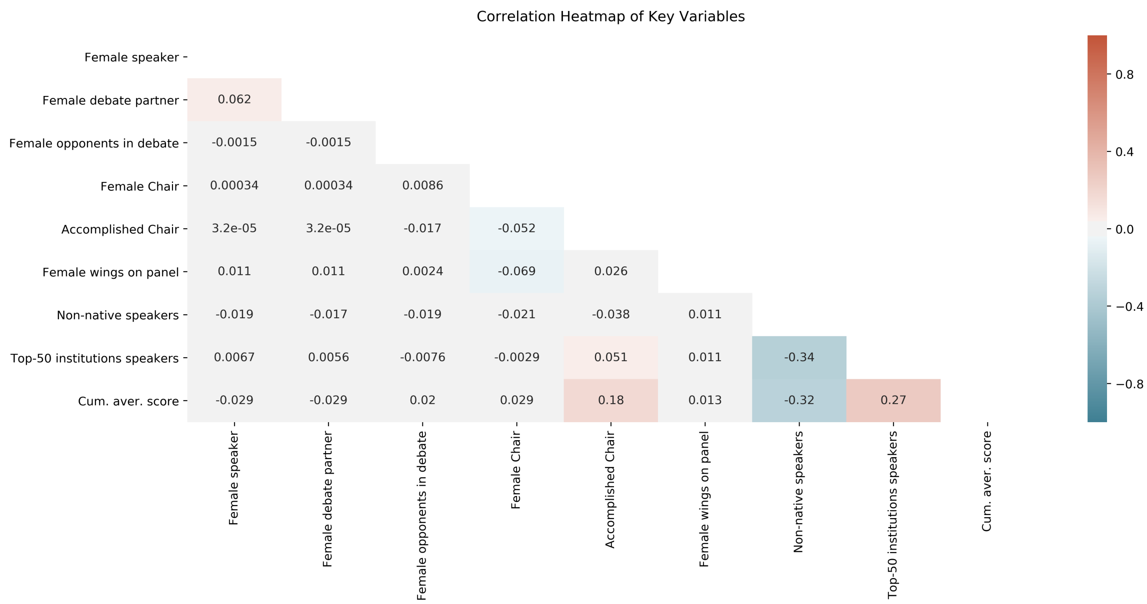

Figure 4.8: Spearman's correlation coefficient heat map across characteristics of speakers, debate room and judges. Accomplished chairs are judges who have advanced to at least one previous EUDC/WUDC tournaments.

4.12.7 Average scores across rounds

4.12.7.1 Male vs. Female Speakers

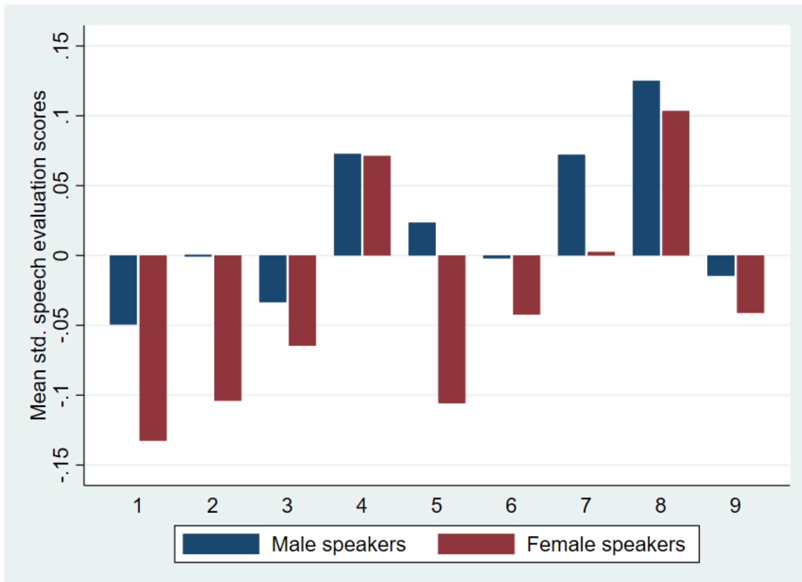

Figure 4.9: Average standardized scores of male vs. female speakers (R1 - R9)

4.12.7.2 MM vs. MF vs. FF teams

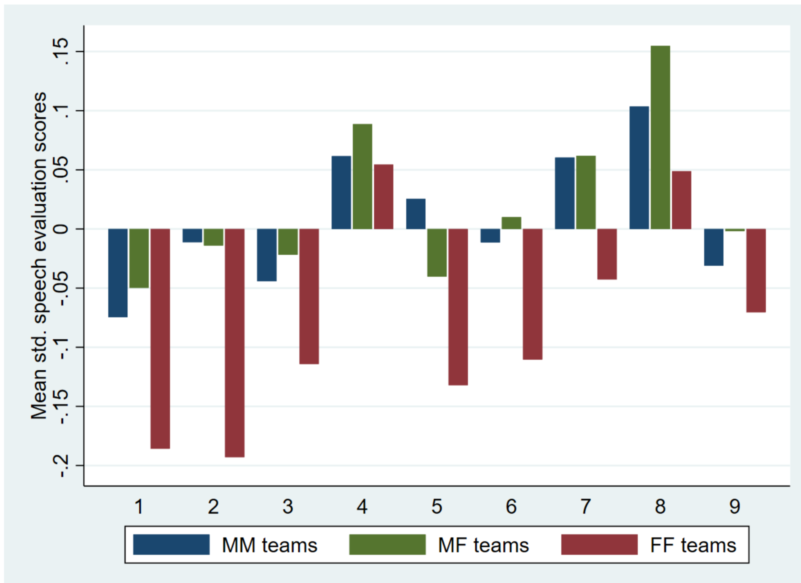

Figure 4.10: Average standardized scores of male-only vs. mixed vs. female-only teams (R1 - R9)

4.12.7.3 Speaker's gender in teams

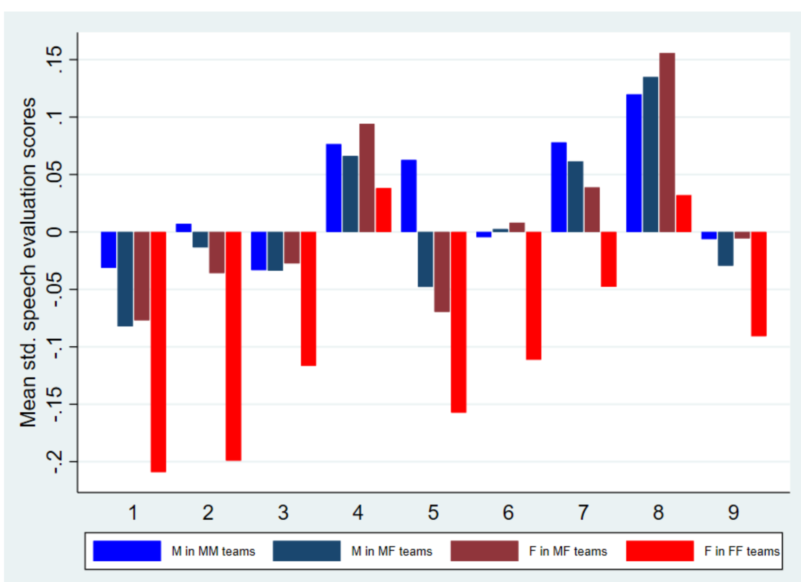

Figure 4.11: Average standardized scores speakers given teammate's gender (R1 - R9)

4.12.8 Distribution of Speech Scores

4.12.8.1 Male vs. Female Speakers (Overall)

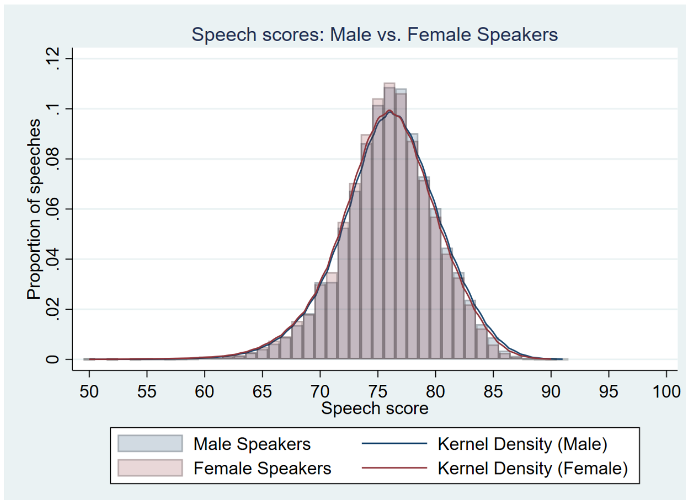

Figure 4.12: Speech score distribution by speaker's gender

4.12.8.2 Male- vs. Female-dominated debates

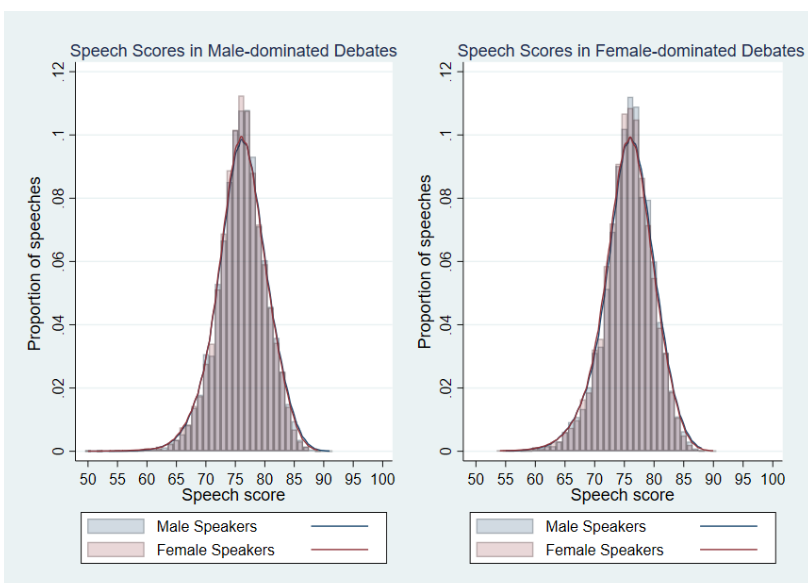

Figure 4.13: Speech score distribution by room gender composition

4.12.8.3 Higher- vs. Lower-ranked debates

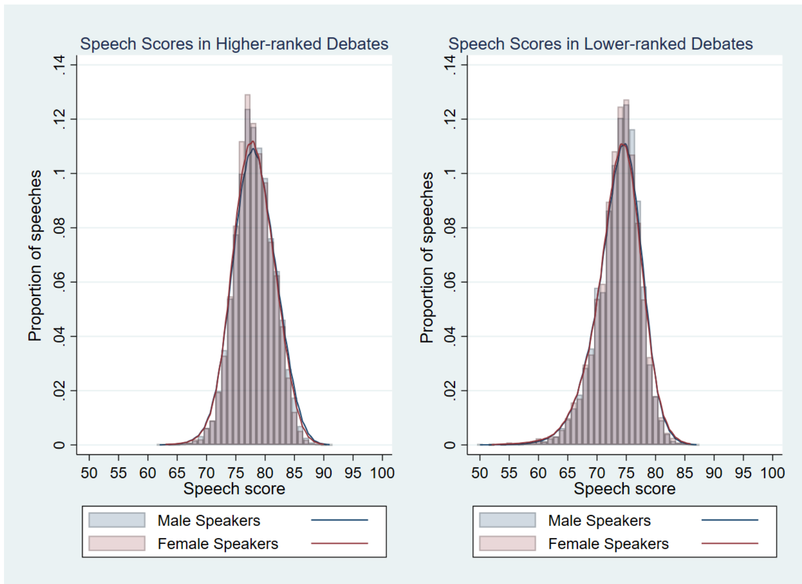

Figure 4.14: Speech score distribution by debate quality ranking

4.12.9 Speech Scores & Team gender composition

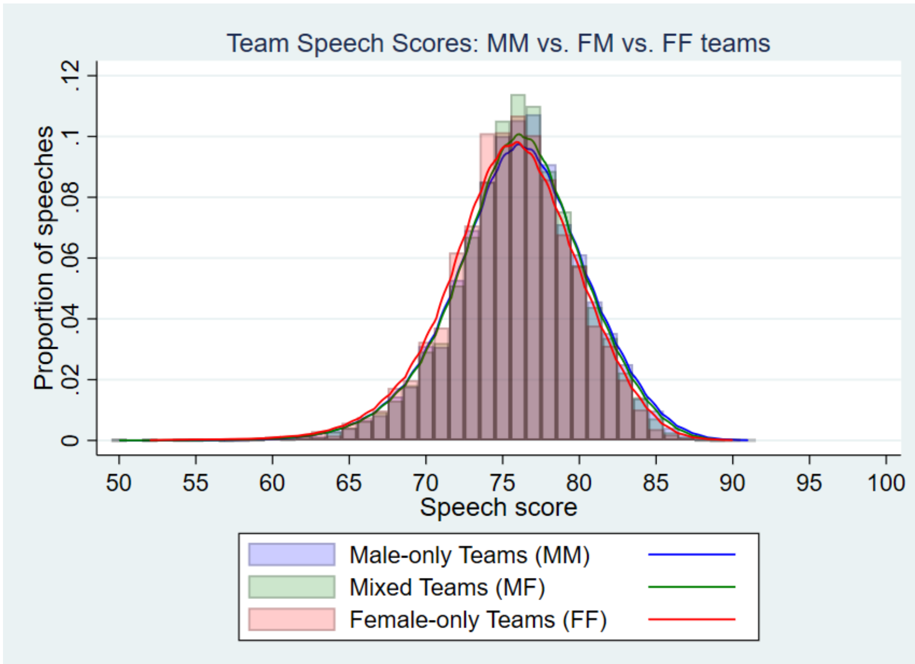

Figure 4.15: Distribution of team speech scores

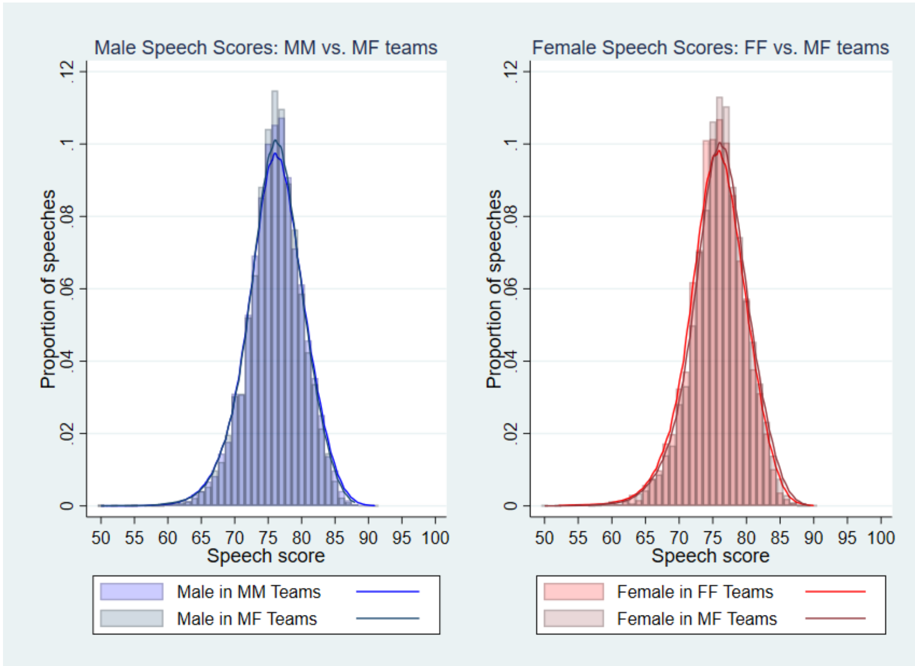

Figure 4.16: Distribution of individual speaker scores given their team gender composition

4.12.10 Speaker in Teams: Higher- vs. Lower-ranked Debates

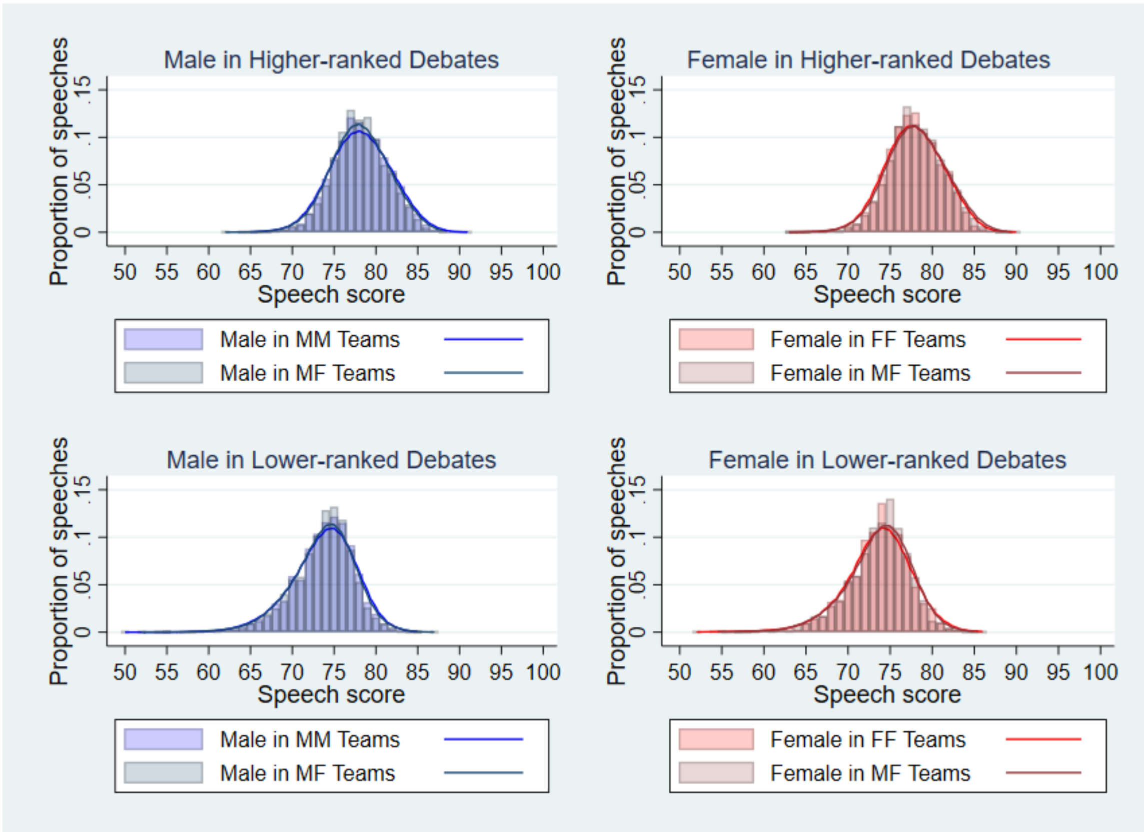

Figure 4.17: Distribution of speaker scores in higher vs. lower-ranked debates given team gender composition

4.13 Tables

4.13.1 DATA: Speeches per competition and omitted debates

Table 4.5: Speeches & debates by competition and missing debates

Tournament	Number of speeches	Number of debates	Omitted debates
EUDC 2015	3760	470	6
EUDC 2016	3968	496	9
EUDC 2017	3736	467	12
EUDC 2018	3064	383	35
WUDC 2015	6096	762	39
WUDC 2016	6776	847	16
WUDC 2017	6440	805	32
WUDC 2018	5328	666	36
Total	39168	4896	185

4.13.2 DATA: Speeches per competition type given team composition

	Competition		
Team type	EUDC	WUDC	Total
MM	6556	10248	16804
MF	5882	10654	16536
FF	2090	3738	5828
Total	14528	24640	39168

4.13.3 DESCRIPTIVE STATISTICS: Tournament score

Table 4.6: Tournament score descriptive statistics ($N = 39168$ speeches)

Competition code	Mean	Min	Max	Median	SD	Total speeches
WUDC15	76.04	58	90	76	4.09	6096
WUDC16	75.67	52	88	76	4.01	6776
WUDC17	76.47	58	88	77	3.83	6440
WUDC18	76.48	59	88	77	3.95	5328
EUDC15	74.96	54	89	75	4.46	3760
EUDC16	75.63	52	91	76	4.45	3968
EUDC17	75.98	55	88	76	4.03	3736
EUDC18	76.07	50	87	76	3.99	3064
WUDC total	76.14	52	90	76	3.99	24640
EUDC total	75.64	50	91	76	4.28	14528

4.13.4 DESCRIPTIVE STATISTICS: t-test on scores of Male vs. Female Speakers

Table 4.7: Two sample t-test with unequal variances on speech scores across demographics ($N_{MaleSpeaker} = 24334$, $N_{FemaleSpeaker} = 14834$)

Group Variable	Mean_M	Mean_F	SD_M	SD_F	t-test	p-value
Speaker Gender	76.04	75.81	4.09	4.12	5.41	0.00^{***}
Non-native Speakers	74.63	74.21	3.86	3.96	7.01	0.00^{***}
Native Speakers	77.32	77.15	3.87	3.75	3.11	0.00^{***}
Top-50-ranked Institutions	78.71	78.24	3.48	3.52	5.56	0.00^{***}
Non-top-50-ranked Institutions	75.44	75.24	3.98	4.04	4.25	0.00^{***}

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

4.13.5 DESCRIPTIVE STATISTICS: Male vs. Female Speakers

Table 4.8: Descriptive statistics of control variables by speaker's gender ($N_{speech} = 39168$, $N_{MaleSpeaker} = 1949$, $N_{FemaleSpeaker} = 1190$)

	Speeches given by...					
	male speakers		**female speakers**			
Control variables	Count	%	Count	%	Total	%
Speaker's characteristics						
Non-native speakers	11530	47.38	6742	45.45	18272	46.65
Native speakers	12804	52.62	8092	54.55	20896	53.35
Top-50-ranked institutions	4511	18.54	2830	19.08	7341	18.74
Non-50-ranked institutions	19823	81.46	12004	80.92	31827	81.26
Female debate partner	8640	35.51	6194	41.76	14834	37.87
Male debate partner	15694	64.49	8640	58.24	24334	62.13
Room & Judge Gender Composition						
Female-majority Room	5533	22.74	7427	50.07	12960	33.09
Male-majority Room	18801	77.26	7407	49.93	26208	66.91
Female chair Judge	7825	32.16	4775	32.19	12600	32.17
Male chair Judge	16509	67.84	10059	67.81	26568	67.83
Female-majority Panel	9311	38.26	5817	39.21	15128	38.62
Male-majority Panel	15023	61.74	9017	60.79	24040	61.38
Tournament & year						
EUDC 2015	2254	9.26	1506	10.15	3760	9.60
EUDC 2016	2317	9.52	1651	11.13	3968	10.13
EUDC 2017	2338	9.61	1398	9.42	3736	9.54
EUDC 2018	1866	7.67	1198	8.08	3064	7.82
WUDC 2015	3960	16.27	2136	14.40	6096	15.56
WUDC 2016	4272	17.56	2504	16.88	6776	17.30
WUDC 2017	4035	16.58	2405	16.21	6440	16.44
WUDC 2018	3292	13.53	2036	13.73	5328	13.60
Motion topic type						
Culture	424	1.74	272	1.83	696	1.78
Economy	2915	11.98	1685	11.36	4600	11.74
Education	704	2.89	400	2.70	1104	2.82
Law/justice	2876	11.82	1796	12.11	4672	11.93
Politics	1259	5.17	789	5.32	2048	5.23
Sports	710	2.92	458	3.09	1168	2.98
Environment	705	2.90	415	2.80	1120	2.86
Free Speech	766	3.15	522	3.52	1288	3.29
Health	1780	7.31	1092	7.36	2872	7.33
Society	3285	13.49	2051	13.83	5336	13.62
Feminism	1450	5.96	862	5.81	2312	5.90
International Relations	3082	12.67	1894	12.76	4976	12.70
Military	3039	12.49	1801	12.14	4840	12.36
Digital Freedom	1339	5.50	797	5.35	2136	5.45
TOTAL SPEECHES	24334	62.13	14834	37.87	39168	100

4.13.6 RESULTS: Gender of Speaker and Opponents, Round 1 vs. Round 2s to 9s (Indicator Variable)

4.13.6.1 Round 1s

Table 4.9: Regression Analysis: Gender of Speaker and Opponents, Round 1s only (N = 4376)

	Dependent Variable: Score (standardized)			
	(1)	(2)	(3)	(4)
Female Speaker	-0.083***		-0.075	-0.001
	(0.03)		(0.09)	(0.08)
1 Female Opponent		0.030	0.013	0.035
		(0.07)	(0.08)	(0.09)
2 Female Opponents		0.026	0.033	-0.130
		(0.07)	(0.08)	(0.09)
3 Female Opponents		0.028	0.052	-0.083
		(0.07)	(0.08)	(0.10)
4 Female Opponents		-0.051	-0.068	-0.225**
		(0.09)	(0.10)	(0.11)
5 Female Opponents		-0.176*	-0.155	-0.135
		(0.11)	(0.12)	(0.15)
6 Female Opponents		0.056	-0.034	-0.034
		(0.18)	(0.22)	(0.34)
Female Speaker × 1 Female Opponent			0.045	-0.167*
			(0.11)	(0.10)
Female Speaker × 2 Female Opponents			-0.028	-0.078
			(0.11)	(0.09)
Female Speaker × 3 Female Opponents			-0.056	-0.116
			(0.11)	(0.10)
Female Speaker × 4 Female Opponents			0.049	0.046
			(0.14)	(0.12)
Female Speaker × 5 Female Opponents			-0.029	-0.210*
			(0.15)	(0.12)
Female Speaker × 6 Female Opponents			0.303	0.186
			(0.25)	(0.32)
Speaker Controls				✓
Room Controls				✓
R^2	0.002	0.003	0.005	0.363
Observations	4376	4376	4376	4376

Speaker controls include: (i) language skill status and (ii) institution ranking. Room controls include: (i) chair judge fixed effect, (ii) debate partner's gender, (iii) wing gender composition, (iv) speaking position, (v) round, (vi) motion type, (vii) competition & year. Robust clustered standard errors at debate level in parentheses.

$^{*}p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.13.6.2 Round 2s to Round 9s

Table 4.10: Regression Analysis: Gender of Speaker and Opponents, Round 2s to 9s (N = 34 792)

	Dependent Variable: Score (standardized)					
	(1)	(2)	(3)	(4)	(5)	(6)
Female Speaker	-0.053***		-0.049	-0.010		
	(0.01)		(0.05)	(0.04)		
1 Female Opponent		0.061*	0.064	0.053	0.003	0.006
		(0.04)	(0.05)	(0.04)	(0.03)	(0.03)
2 Female Opponents		0.056	0.054	0.049	0.002	0.016
		(0.04)	(0.05)	(0.04)	(0.03)	(0.03)
3 Female Opponents		0.021	0.025	0.030	0.006	0.025
		(0.04)	(0.05)	(0.04)	(0.03)	(0.03)
4 Female Opponents		-0.052	-0.051	-0.011	0.028	0.034
		(0.04)	(0.05)	(0.04)	(0.03)	(0.03)
5 Female Opponents		-0.054	-0.063	-0.011	-0.005	-0.019
		(0.06)	(0.07)	(0.05)	(0.05)	(0.04)
6 Female Opponents		-0.148	-0.152	-0.027	0.125	0.142*
		(0.11)	(0.14)	(0.12)	(0.09)	(0.08)
Female Speaker × 1 Female Opponent			-0.007	-0.038	-0.023	-0.016
			(0.06)	(0.04)	(0.04)	(0.03)
Female Speaker × 2 Female Opponents			0.003	-0.047	-0.023	-0.026
			(0.06)	(0.04)	(0.04)	(0.03)
Female Speaker × 3 Female Opponents			-0.014	-0.051	-0.014	-0.023
			(0.06)	(0.04)	(0.04)	(0.03)
Female Speaker × 4 Female Opponents			-0.002	-0.012	-0.048	-0.033
			(0.07)	(0.05)	(0.04)	(0.04)
Female Speaker × 5 Female Opponents			0.023	-0.030	0.004	0.020
			(0.09)	(0.07)	(0.07)	(0.06)
Female Speaker × 6 Female Opponents			-0.002	0.088	0.015	-0.047
			(0.17)	(0.15)	(0.18)	(0.11)
Speaker Controls				✓		
Room Controls				✓		✓
Speaker FE					✓	✓
R^2	0.001	0.002	0.002	0.338	0.572	0.648
Observations	34792	34792	34792	34792	34786	34786

Speaker controls include: (i) language skill status and (ii) institution ranking. Room controls include: (i) chair judge fixed effect, (ii) debate partner's gender, (iii) wing gender composition, (iv) speaking position, (v) round, (vi) motion type, (vii) competition & year. Robust clustered standard errors at debate level in parentheses.
R^2 of model (5) and (6) is $R^2_{between}$. Singleton observations are dropped in model (5) and (6).
$^*p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.13.6.3 Round 2s to Round 9s (controlling for team standing)

Table 4.11: Regression Analysis: Gender of Speaker and Opponents, Round 2s to 9s, controlling for team standing (N = 34 792)

	Dependent Variable: Score (standardized)					
	(1)	(2)	(3)	(4)	(5)	(6)
Female Speaker	-0.030***		0.041	0.021		
	(0.01)		(0.04)	(0.03)		
1 Female Opponent		-0.094***	-0.064	0.023	0.004	0.007
		(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
2 Female Opponents		-0.133***	-0.104**	0.019	0.004	0.017
		(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
3 Female Opponents		-0.153***	-0.125***	0.005	0.008	0.026
		(0.03)	(0.04)	(0.04)	(0.03)	(0.03)
4 Female Opponents		-0.210***	-0.177***	-0.027	0.029	0.035
		(0.04)	(0.05)	(0.04)	(0.03)	(0.03)
5 Female Opponents		-0.191***	-0.156***	-0.025	-0.005	-0.018
		(0.05)	(0.06)	(0.05)	(0.05)	(0.04)
6 Female Opponents		-0.271***	-0.227*	-0.043	0.127	0.143*
		(0.10)	(0.12)	(0.11)	(0.09)	(0.08)
Female Speaker × 1 Female Opponent			-0.077	-0.063	-0.022	-0.015
			(0.05)	(0.04)	(0.04)	(0.03)
Female Speaker × 2 Female Opponents			-0.076	-0.073*	-0.022	-0.026
			(0.05)	(0.04)	(0.04)	(0.03)
Female Speaker × 3 Female Opponents			-0.075	-0.068*	-0.013	-0.023
			(0.05)	(0.04)	(0.04)	(0.03)
Female Speaker × 4 Female Opponents			-0.084	-0.051	-0.047	-0.032
			(0.06)	(0.05)	(0.04)	(0.04)
Female Speaker × 5 Female Opponents			-0.093	-0.070	0.007	0.022
			(0.08)	(0.07)	(0.07)	(0.06)
Female Speaker × 6 Female Opponents			-0.136	0.017	0.017	-0.045
			(0.16)	(0.14)	(0.18)	(0.11)
Speaker Controls				✓		
Room Controls				✓		✓
Speaker FE					✓	✓
R^2	0.171	0.173	0.174	0.387	0.572	0.648
Observations	34792	34792	34792	34792	34786	34786

Speaker controls include: (i) language skill status and (ii) institution ranking. Room controls include: (i) chair judge fixed effect, (ii) debate partner's gender, (iii) wing gender composition, (iv) speaking position, (v) round, (vi) motion type, (vii) competition & year. Robust clustered standard errors at debate level in parentheses.
R^2 of model (5) and (6) is $R^2_{between}$. Singleton observations are dropped in model (5) and (6).
$^*p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.13.7 EXTENSIONS: Higher vs. Lower-ranked Debates

Table 4.12: Regression Analysis: Gender of Speakers and Opponents, Higher vs. Lower-ranked Debates (controlling for team standing)

	Dependent Variable: Score (standardized)					
	(1)	(2)	(3)	(4)	(5)	(6)
Higher-ranked Debates						
Female Speaker	-0.031**		0.028	0.005		
	(0.01)		(0.03)	(0.02)		
Number of Female Opponents		-0.023***	-0.013	-0.008	-0.003	0.001
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Female Speaker × Number of Female Opponents			-0.028**	-0.026***	-0.023**	-0.022**
			(0.01)	(0.01)	(0.01)	(0.01)
R^2	0.040	0.041	0.042	0.243	0.423	0.527
Observations	18270	18270	18270	18270	18046	18046
Lower-ranked Debates						
Female Speaker	-0.008		-0.022	-0.038		
	(0.02)		(0.03)	(0.03)		
Number of Female Opponents		-0.022**	-0.024**	-0.007	0.016	0.006
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Female Speaker × Number of Female Opponents			0.006	0.012	0.016	0.015
			(0.01)	(0.01)	(0.01)	(0.01)
R^2	0.051	0.052	0.052	0.276	0.399	0.551
Observations	16518	16518	16518	16518	16331	16331
Speaker Controls				✓		
Room Controls				✓		✓
Speaker FE					✓	✓

All models control for team standing (i.e. average cumulative speech scores of two speakers in a team up until the respective round.). Speaker controls include: (i) language skill status and (ii) institution ranking. Room controls include: (i) chair judge fixed effect, (ii) debate partner's gender, (iii) wing gender composition, (iv) speaking position, (v) round, (vi) motion type, (vii) competition & year. Robust clustered standard errors at debate level in parentheses. R^2 of model (5) and (6) is $R^2_{between}$. Singleton observations are dropped in model (5) and (6).
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 4.13: Regression Analysis: Gender of Speakers and Number of Opponents, Higher-ranked Debates only (controlling for team standing)

	Dependent Variable: Score (standardized)					
	(1)	(2)	(3)	(4)	(5)	(6)
Female Speaker	-0.031**		0.074	0.060		
	(0.01)		(0.05)	(0.04)		
1 Female Opponent		-0.027	0.014	0.025	0.035	0.028
		(0.04)	(0.05)	(0.04)	(0.04)	(0.03)
2 Female Opponents		-0.041	-0.007	0.016	0.025	0.027
		(0.04)	(0.05)	(0.04)	(0.04)	(0.03)
3 Female Opponents		-0.050	-0.013	0.010	0.020	0.029
		(0.04)	(0.05)	(0.04)	(0.04)	(0.03)
4 Female Opponents		-0.111**	-0.044	-0.018	0.002	0.015
		(0.04)	(0.05)	(0.05)	(0.04)	(0.04)
5 Female Opponents		-0.108*	-0.049	-0.021	0.019	0.010
		(0.06)	(0.07)	(0.06)	(0.06)	(0.05)
6 Female Opponents		-0.287	-0.064	-0.090	0.010	-0.000
		(0.20)	(0.18)	(0.18)	(0.15)	(0.14)
Female Speaker × 1 Female Opponent			-0.113*	-0.111**	-0.127***	-0.113***
			(0.06)	(0.05)	(0.05)	(0.04)
Female Speaker × 2 Female Opponents			-0.095	-0.112**	-0.092**	-0.095**
			(0.06)	(0.05)	(0.05)	(0.04)
Female Speaker × 3 Female Opponents			-0.102*	-0.113**	-0.113**	-0.103**
			(0.06)	(0.05)	(0.05)	(0.04)
Female Speaker × 4 Female Opponents			-0.178***	-0.165***	-0.169***	-0.166***
			(0.07)	(0.06)	(0.05)	(0.05)
Female Speaker × 5 Female Opponents			-0.169*	-0.168*	-0.163*	-0.140*
			(0.10)	(0.09)	(0.09)	(0.08)
Female Speaker × 6 Female Opponents			-0.860**	-0.716**	-0.423	-0.410*
			(0.38)	(0.31)	(0.27)	(0.24)
Speaker Controls				✓		
Room Controls				✓		✓
Speaker FE					✓	✓
R^2	0.040	0.041	0.042	0.244	0.424	0.528
Observations	18270	18270	18270	18270	18046	18046

All models control for team standing (i.e. average cumulative speech scores of two speakers in a team up until the respective round.). Speaker controls include: (i) language skill status and (ii) institution ranking. Room controls include: (i) chair judge fixed effect, (ii) debate partner's gender, (iii) wing gender composition, (iv) speaking position, (v) round, (vi) motion type, (vii) competition & year. Robust clustered standard errors at debate level in parentheses. R^2 of model (5) and (6) is $R^2_{between}$. Singleton observations are dropped in model (5) and (6).
$^{*}p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.13.8 EXTENSIONS: Speaker ft. Partner's Gender

Table 4.14: Regression Analysis: Speaker ft. Partner's Gender, All rounds

	Dependent Variable: Score (standardized)					
	(1)	(2)	(3)	(4)	(5)	(6)
Male Speaker in MF Team	-0.023		-0.033	-0.055**	-0.029	-0.034
	(0.02)		(0.03)	(0.03)	(0.03)	(0.03)
Female Speaker in MF Team	-0.021		-0.021	-0.041	0.087**	0.070**
	(0.02)		(0.03)	(0.03)	(0.04)	(0.03)
Female Speaker in FF Team	-0.126***		-0.127***	-0.114***		
	(0.02)		(0.05)	(0.04)		
Number of Female Opponents		-0.020**	-0.021*	-0.015	0.007	0.007
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Male Speaker in MF Team × Number of Female Opponents			0.005	0.005	-0.012	-0.008
			(0.01)	(0.01)	(0.01)	(0.01)
Female Speaker in MF Team × Number of Female Opponents			0.000	-0.001	-0.013	-0.009
			(0.01)	(0.01)	(0.01)	(0.01)
Female Speaker in FF Team × Number of Female Opponents			0.001	-0.001	0.000	-0.003
			(0.02)	(0.02)	(0.01)	(0.01)
Speaker Controls				✓		
Room Controls				✓		✓
Speaker FE					✓	✓
R^2	0.002	0.001	0.002	0.314	0.568	0.640
F	12.2	6.6	5.4	84.7	2.7	19.2
Observations	39168	39168	39168	39168	39157	39157

Speaker controls include: (i) language skill status, (ii) institution ranking. Room controls include:(i) chair judge fixed effect, (ii) wing gender composition, iii) speaking position, (iv) round, (v) motion type, (vi) competition & year. Robust clustered standard errors at debate level in parentheses. R^2 of (5) and (6) is $R^2_{between}$. Singleton observations are dropped.
$^{*}p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4.15: Regression Analysis: Speaker ft. Partner's Gender, Round 1s vs. Round 2s to 9s

	Dependent Variable: Score (standardized)					
	(1)	(2)	(3)	(4)	(5)	(6)
Round 1s						
Male Speaker in MF Team	-0.051		-0.148*	-0.149*		
	(0.04)		(0.08)	(0.09)		
Female Speaker in MF Team	-0.046		-0.108	-0.109		
	(0.04)		(0.08)	(0.08)		
Female Speaker in FF Team	-0.178***		-0.133	-0.226*		
	(0.06)		(0.11)	(0.12)		
Number of Female Opponents		-0.019	-0.031	-0.060**		
		(0.01)	(0.02)	(0.03)		
Male Speaker in MF Team × Number of Female Opponents			0.045	0.039		
			(0.03)	(0.03)		
Female Speaker in MF Team × Number of Female Opponents			0.030	0.022		
			(0.03)	(0.03)		
Female Speaker in FF Team × Number of Female Opponents			-0.017	0.015		
			(0.04)	(0.04)		
R^2	0.005	0.001	0.006	0.361		
Observations	4376	4376	4376	4376		
Round 2s to 9s						
Male Speaker in MF Team	-0.020		-0.019	-0.039	-0.012	-0.020
	(0.02)		(0.04)	(0.03)	(0.03)	(0.03)
Female Speaker in MF Team	-0.018		-0.011	-0.029	0.089**	0.070*
	(0.02)		(0.04)	(0.03)	(0.04)	(0.04)
Female Speaker in FF Team	-0.119***		-0.131**	-0.109***		
	(0.02)		(0.05)	(0.04)		
Number of Female Opponents		-0.020**	-0.020*	-0.009	0.012	0.011
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Male Speaker in MF Team × Number of Female Opponents			-0.001	-0.001	-0.019*	-0.014
			(0.01)	(0.01)	(0.01)	(0.01)
Female Speaker in MF Team × Number of Female Opponents			-0.003	-0.006	-0.016	-0.011
			(0.01)	(0.01)	(0.01)	(0.01)
Female Speaker in FF Team × Number of Female Opponents			0.005	0.005	-0.001	-0.004
			(0.02)	(0.02)	(0.01)	(0.01)
R^2	0.002	0.001	0.002	0.337	0.572	0.648
Observations	34792	34792	34792	34792	34786	34786
Round 2s to 9s (controlling for team standing)						
Male Speaker in MF Team	-0.012		0.018	-0.026	-0.013	-0.020
	(0.02)		(0.03)	(0.03)	(0.03)	(0.03)
Female Speaker in MF Team	-0.009		0.026	-0.016	0.091**	0.072*
	(0.02)		(0.03)	(0.03)	(0.04)	(0.04)
Female Speaker in FF Team	-0.069***		-0.032	-0.065*		
	(0.02)		(0.05)	(0.04)		
Number of Female Opponents		-0.040***	-0.031***	-0.007	0.012	0.011
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Male Speaker in MF Team × Number of Female Opponents			-0.014	-0.005	-0.019*	-0.013
			(0.01)	(0.01)	(0.01)	(0.01)
Female Speaker in MF Team × Number of Female Opponents			-0.017	-0.009	-0.016	-0.010
			(0.01)	(0.01)	(0.01)	(0.01)
Female Speaker in FF Team × Number of Female Opponents			-0.017	-0.007	-0.001	-0.004
			(0.02)	(0.02)	(0.01)	(0.01)
R^2	0.171	0.173	0.174	0.387	0.573	0.648
Observations	34792	34792	34792	34792	34786	34786
Speaker Controls				✓		
Room Controls				✓		✓
Speaker FE					✓	✓

Team standing us the average cumulative speech scores of two speakers in a team up until the respective round. Speaker controls include: (i) language skill status, (ii) institution ranking. Room controls include:(i) chair judge fixed effect, (ii) wing gender composition, iii) speaking position, (iv) round, (v) motion type, (vi) competition & year. Robust clustered standard errors at debate level in parentheses. R^2 of (5) and (6) is $R^2_{between}$. Singleton observations are dropped.
$^{*}p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Summary

In this dissertation, I investigate gender disparities in speech patterns and how they matter in performance evaluation across genders, as well as how the gender composition of committees and opponents causally impact speech performance in real-life tournaments. Chapter 2 links the persuasion-relevant linguistic elements of debate speeches to speech evaluation scores, taking into account the interplay across genders of speakers and judges. Here, I find significant differences in persuasive speech patterns between men and women. Specifically, female speakers use more personal and disclosing speaking style, with more hedging phrases and non-fluencies in their speeches. In their answers to questions from opponents, women negate less, while having longer and more vague answers. On average, women receive lower evaluation scores than men. Across debates, having a less analytical speaking style and more positive sentiment is associated with higher scores for speeches by women, but not by men. Within debates, except for non-fluencies, there is no robust evidence of gender-specific evaluation standards. These findings suggest that the difference in average speech score between men and women arises because speeches of female speakers contain more score-reducing and fewer score-enhancing features, rather than discrimination.

In Chapter 3, I study how the gender composition and power hierarchy of judge committees causally impact performance evaluation patterns across male and female contestants. Committees with a female chair judge give lower scores to both male and female speakers, particularly in higher-ranked debates. Importantly, there is no difference between male and female speakers in how their scores are affected if the judge committee contains more women or is chaired by a woman. These results suggest that gender quotas on evaluation committees does not necessarily eliminate the glass ceiling for women.

Finally, Chapter 4 examines whether the gender composition of opponents affects the competitive performance of men and women in multi-round, high-stake contests. On average, neither male nor female contestants are affected by the gender composition of opponents. Nevertheless, in higher-ranked debates, female contestants perform comparatively worse in rooms with more female opponents. Therefore, these findings indicate that larger inflow of women into same competitions for high-profile positions does not necessarily reduce the thickness of the glass ceiling.

Nederlandse Samenvatting (Summary in Dutch)

In dit proefschrift onderzoek ik verschillen tussen mannen en vrouwen in spraakpatronen en hoe deze van belang zijn voor de evaluatie van prestaties van zowel mannen als vrouwen. Ook onderzoek ik hoe de samenstelling van commissies en tegenstanders in termen van geslacht een oorzakelijk effect heeft op prestaties in debat-toernooien. Hoofdstuk 2 koppelt de linguïstische elementen van toespraken die relevant zijn voor overredingskracht aan de score toegekend aan de toespraak, waarbij rekening wordt gehouden met de interactie in termen van geslacht tussen sprekers en juryleden. Hier vind ik significante verschillen in spraakpatronen tussen mannen en vrouwen. Over het algemeen geldt dat een minder analytische spreekstijl en een positiever sentiment geassocieerd is met hogere scores voor toespraken door vrouwen, maar niet voor toespraken door mannen. Binnen een debat is er echter geen robuust bewijs van genderspecifieke evaluatienormen, met uitzondering van vloeiend taalgebruik. Deze bevindingen suggereren dat het verschil in gemiddelde scores tussen mannen en vrouwen ontstaat doordat toespraken van vrouwelijke sprekers meer score-verlagende en minder score-verhogende kenmerken bevatten, in plaats van discriminatie.

In hoofdstuk 3 bestudeer ik hoe de gendersamenstelling en hiërarchie van jurycommissies van invloed zijn op hun evaluatie van mannelijke en vrouwelijke deelnemers. Commissies met een vrouwelijke juryvoorzitter geven lagere scores aan zowel mannelijke als vrouwelijke sprekers, vooral in hoger gerangschikte debatten. Een belangrijk resultaat is dat er geen verschil is tussen mannelijke en vrouwelijke sprekers in hoe hun score wordt beïnvloedt door het aantal vrouwen in de jurycommissie of het geslacht van de juryvoorzitter. Deze resultaten suggereren dat genderquota in evaluatiecommissies niet altijd een afdoende maatregel is om het glazen plafond voor vrouwen te doorbreken.

Ten slotte wordt in Hoofdstuk 4 onderzocht of het geslacht van tegenstanders de competitieve prestatie van mannen en vrouwen beïnvloedt in toernooien met meerdere rondes en hoge belangen. Ik vind dat gemiddeld mannelijke noch vrouwelijke deelnemers beïnvloed worden door het geslacht van hun tegenstanders. Niettemin presteren vrouwelijke deelnemers in hoger gerangschikte debatten relatief slechter in kamers met meer vrouwelijke tegenstanders. Daarom geven deze bevindingen aan dat een grotere instroom van vrouwen in competitieve trajecten voor belangrijke posities niet vanzelf leidt tot een vermindering van de dikte van het glazen plafond.

Bibliography

Alan, S., Ertac, S., Kubilay, E., and Loranth, G. (2020). Understanding gender differences in leadership. *Economic Journal*, 130(626):263–289.

Antonovics, K., Arcidiacono, P., and Walsh, R. (2009). The effects of gender interactions in the lab and in the field. *The Review of Economics and Statistics*, 91(1):152–162.

Apesteguia, J., Azmat, G., and Iriberri, N. (2012). The impact of gender composition on team performance and decision making: Evidence from the field. *Management Science*, 58(1):78–93.

Ash, E., Chen, D. L., and Lu, W. (2017). Polarization of us circuit court judges: A machine learning approach.

Backus, P., Cubel, M., Guid, M., Sanchez-Pages, S., and Mañas, E. (2016). Gender, competition and performance: Evidence from real tournaments.

Blau, F. D. and Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, 55(3):789–865.

Bohnet, I. and Bowles, H. R. (2008). Gender in negotiation. *Negotiation Journal*, 24(4):389.

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

Buser, T. and Yuan, H. (2020). Public speaking aversion. *Available at SSRN 3724341. Tinbergen Institute Discussion Paper TI 2020 -074/I.*

Chikeleze, M., Johnson, I., and Gibson, T. (2018). Let's argue: Using debate to teach critical thinking and communication skills to future leaders. *Journal of Leadership Education*, 17(2).

Conti, R., Collins, M. A., and Picariello, M. L. (2001). The impact of competition on intrinsic motivation and creativity: considering gender, gender segregation and gender role orientation. *Personality and individual differences*, 31(8):1273–1289.

Cotton, C., McIntyre, F., and Price, J. (2013). Gender differences in repeated competition: Evidence from school math contests. *Journal of Economic Behavior & Organization*, 86:52–66.

Dargnies, M.-P. (2012). Men too sometimes shy away from competition: The case of team competition. *Management Science*, 58(11):1982–2000.

Datta Gupta, N., Poulsen, A., and Villeval, M. C. (2013). Gender matching and competitiveness: Experimental evidence. *Economic Inquiry*, 51(1):816–835.

De Paola, M., Gioia, F., and Scoppa, V. (2015). Are females scared of competing with males? results from a field experiment. *Economics of Education Review*, 48:117–128.

Delfgaauw, J., Dur, R., Sol, J., and Verbeke, W. (2013). Tournament incentives in the field: Gender differences in the workplace. *Journal of Labor Economics*, 31(2):305–326.

Dilmaghani, M. (2020). Gender differences in performance under time constraint: Evidence from chess tournaments. *Journal of Behavioral and Experimental Economics*, 89:101505.

Eckel, C. C., Gangadharan, L., Grossman, P. J., and Xue, N. (2020). *The Gender Leadership Gap: Insights from Experiments*. Monash University, Monash Business School, Department of Economics.

Exley, C. L. and Kessler, J. B. (2019). The gender gap in self-promotion. Technical report, National Bureau of Economic Research.

Gentzkow, M., Shapiro, J. M., and Taddy, M. (2019). Measuring group differences in high-dimensional choices: Method and application to congressional speech. *Econometrica*, 87(4):1307–1340.

Gerdes, C. and Gränsmark, P. (2010). Strategic behavior across gender: A comparison of female and male expert chess players. *Labour Economics*, 17(5):766–775.

Gneezy, U., Niederle, M., and Rustichini, A. (2003). Performance in competitive environments: Gender differences. *Quarterly Journal of Economics*, 118(3):1049–1074.

Gneezy, U. and Rustichini, A. (2004). Gender and competition at a young age. *American Economic Review*, 94(2):377–381.

Goldin, C., Kerr, S. P., Olivetti, C., and Barth, E. (2017). The expanding gender earnings gap: Evidence from the lehd-2000 census. *American Economic Review*, 107(5):110–14.

Green III, C. S. and Klug, H. G. (1990). Teaching critical thinking and writing through debates: An experimental evaluation. *Teaching Sociology*, pages 462–471.

Hansen, S., McMahon, M., and Prat, A. (2018). Transparency and deliberation within the fomc: a computational linguistics approach. *Quarterly Journal of Economics*, 133(2):801–870.

Hoogendoorn, S., Oosterbeek, H., and Van Praag, M. (2013). The impact of gender diversity on the performance of business teams: Evidence from a field experiment. *Management Science*, 59(7):1514–1528.

Iriberri, N. and Rey-Biel, P. (2019). Competitive pressure widens the gender gap in performance: Evidence from a two-stage competition in mathematics. *The Economic Journal*, 129(620):1863–1893.

Ivanova-Stenzel, R. and Kübler, D. (2011). Gender differences in team work and team competition. *Journal of Economic Psychology*, 32(5):797–808.

Jetter, M. and Walker, J. K. (2018). The gender of opponents: Explaining gender differences in performance and risk-taking? *European Economic Review*, 109:238–256.

Leibbrandt, A. and List, J. A. (2015). Do women avoid salary negotiations? evidence from a large-scale natural field experiment. *Management Science*, 61(9):2016–2024.

Li, X., Han, Z., Fu, J., Mei, Y., and Liu, J. (2019). Debate: A new approach for improving the dialectical thinking of university students. *Innovations in Education and Teaching International*, pages 1–12.

Mago, S. D. and Razzolini, L. (2019). Best-of-five contest: An experiment on gender differences. *Journal of Economic Behavior & Organization*, 162:164–187.

Menéndez, D. A., González-Barahona, J. M., and Robles, G. (2020). Damegender: Writing and comparing gender detection tools. In *SATToSE*.

Moely, B. E., Skarin, K., and Weil, S. (1979). Sex differences in competition—cooperation behavior of children at two age levels. *Sex Roles*, 5(3):329–342.

Niederle, M. (2017). A gender agenda: a progress report on competitiveness. *American Economic Review*, 107(5):115–19.

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

Niederle, M. and Vesterlund, L. (2011). Gender and competition. *Annu. Rev. Econ.*, 3(1):601–630.

Pennebaker, J. W., Boyd, R. L., Jordan, K., and Blackburn, K. (2015). The development and psychometric properties of liwc2015. Technical report.

Santamaría, L. and Mihaljević, H. (2018). Comparison and benchmark of name-to-gender inference services. *PeerJ Computer Science*, 4:e156.

Säve-Söderbergh, J. and Sjögren Lindquist, G. (2017). Children do not behave like adults: Gender gaps in performance and risk taking in a random social context in the high-stakes game shows jeopardy and junior jeopardy. *The Economic Journal*, 127(603):1665–1692.

Shurchkov, O. (2012). Under pressure: gender differences in output quality and quantity under competition and time constraints. *Journal of the European Economic Association*, 10(5):1189–1213.

Shurchkov, O. and Eckel, C. C. (2018). *Gender differences in behavioral traits and labor market outcomes*. Oxford, UK: Oxford University Press.

van Dolder, D., van den Assem, M. J., and Buser, T. (2020). Gender and willingness to compete for high stakes. *Available at SSRN 3537678*.

Villeval, M. C. (2012). Ready, steady, compete. *Science*, 335(6068):544–545.

Wozniak, D. (2012). Gender differences in a market with relative performance feedback: Professional tennis players. *Journal of Economic Behavior & Organization*, 83(1):158–171.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and VU University Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

739. R.J. DÖTTLING, *Essays in Financial Economics*
740. E.S. ZWIERS, *About Family and Fate: Childhood Circumstances and Human Capital Formation*
741. Y.M. KUTLUAY, *The Value of (Avoiding) Malaria*
742. A. BOROWSKA, *Methods for Accurate and Efficient Bayesian Analysis of Time Series*
743. B. HU, *The Amazon Business Model, the Platform Economy and Executive Compensation: Three Essays in Search Theory*
744. R.C. SPERNA WEILAND, *Essays on Macro-Financial Risks*
745. P.M. GOLEC, *Essays in Financial Economics*
746. M.N. SOUVERIJN, *Incentives at work*
747. M.H. COVENEY, *Modern Imperatives: Essays on Education and Health Policy*
748. P. VAN BRUGGEN, *On Measuring Preferences*
749. M.H.C. NIENTKER, *On the Stability of Stochastic Dynamic Systems and their use in Econometrics*
750. S. GARCIA MANDICÓ, *Social Insurance, Labor Supply and Intra-Household Spillovers*
751. Y. SUN, Consumer Search and Quality
752. I. KERKEMEZOS, *On the Dynamics of (Anti) Competitive Behaviour in the Airline Industry*

753. G.W. GOY, Modern Challenges to Monetary Policy

754. A.C. VAN VLODROP, Essays on Modeling Time-Varying Parameters

755. J. SUN, *Tell Me How To Vote, Understanding the Role of Media in Modern Elections*

756. J.H. THIEL, *Competition, Dynamic Pricing and Advice in Frictional Markets: Theory and Evidence from the Dutch Market for Mortgages*

757. A. NEGRIU, *On the Economics of Institutions and Technology: a Computational Approach*

758. F. GRESNIGT, *Identifying and Predicting Financial Earth Quakes using Hawkes Processes*

759. A. EMIRMAHMUTOGLU, *Misperceptions of Uncertainty and Their Applications to Prevention*

760. A. RUSU, *Essays in Public Economics*

761. M.A. COTOFAN, *Essays in Applied Microeconomics: Non-Monetary Incentives, Skill Formation, and Work Preferences*

762. B.P.J. ANDRÉE, *Theory and Application of Dynamic Spatial Time Series Models*

763. P. PELZL, *Macro Questions, Micro Data: The Effects of External Shocks on Firms*

764. D.M. KUNST *Essays on Technological Change, Skill Premia and Development*

765. A.J. HUMMEL, *Tax Policy in Imperfect Labor Markets*

766. T. KLEIN, *Essays in Competition Economics*

767. M. VIGH, *Climbing the Socioeconomic Ladder: Essays on Sanitation and Schooling*

768. YAN XU, *Eliciting Preferences and Private Information: Tell Me What You Like and What You Think*

769. S. RELLSTAB, *Balancing Paid Work and Unpaid Care over the Life-Cycle*

770. Z. DENG, *Empirical Studies in Health and Development Economics*

771. L. KONG, *Identification Robust Testing in Linear Factor Models*
772. I. NEAMŢU, *Unintended Consequences of Post-Crisis Banking Reforms*
773. B. KLEIN TEESELINK, *From Mice to Men: Field Studies in Behavioral Economics*
774. B. TEREICK, *Making Crowds Wiser: The Role of Incentives, Individual Biases, and Improved Aggregation*
775. A. CASTELEIN, *Models for Individual Responses*
776. D. KOLESNYK, *Consumer Disclosures on Social Media Platforms: A Global Investigation*
777. M.A. ROLA-JANICKA, *Essays on Financial Instability and Political Economy of Regulation*
778. J.J. KLINGEN, *Natural Experiments in Environmental and Transport Economics*
779. E.M. ARTMANN, *Educational Choices and Family Outcomes*
780. F.J. OSTERMEIJER, *Economic Analyses of Cars in the City*
781. T. ÖZDEN, *Adaptive Learning and Monetary Policy in DSGE Models*
782. D. WANG, *Empirical Studies in Financial Stability and Natural Capital*
783. L.S. STEPHAN, *Estimating Diffusion and Adoption Parameters in Networks New Estimation Approaches for the Latent-Diffusion-Observed-Adoption Model*
784. S.R. MAYER, *Essays in Financial Economics*
785. A.R.S. WOERNER, *Behavioral and Financial Change – Essays in Market Design*
786. M. WIEGAND, *Essays in Development Economics*
787. L.M. TREUREN, *Essays in Industrial Economics - Labor market imperfections, cartel stability, and public interest cartels*
788. D.K. BRANDS, *Economic Policies and Mobility Behaviour*