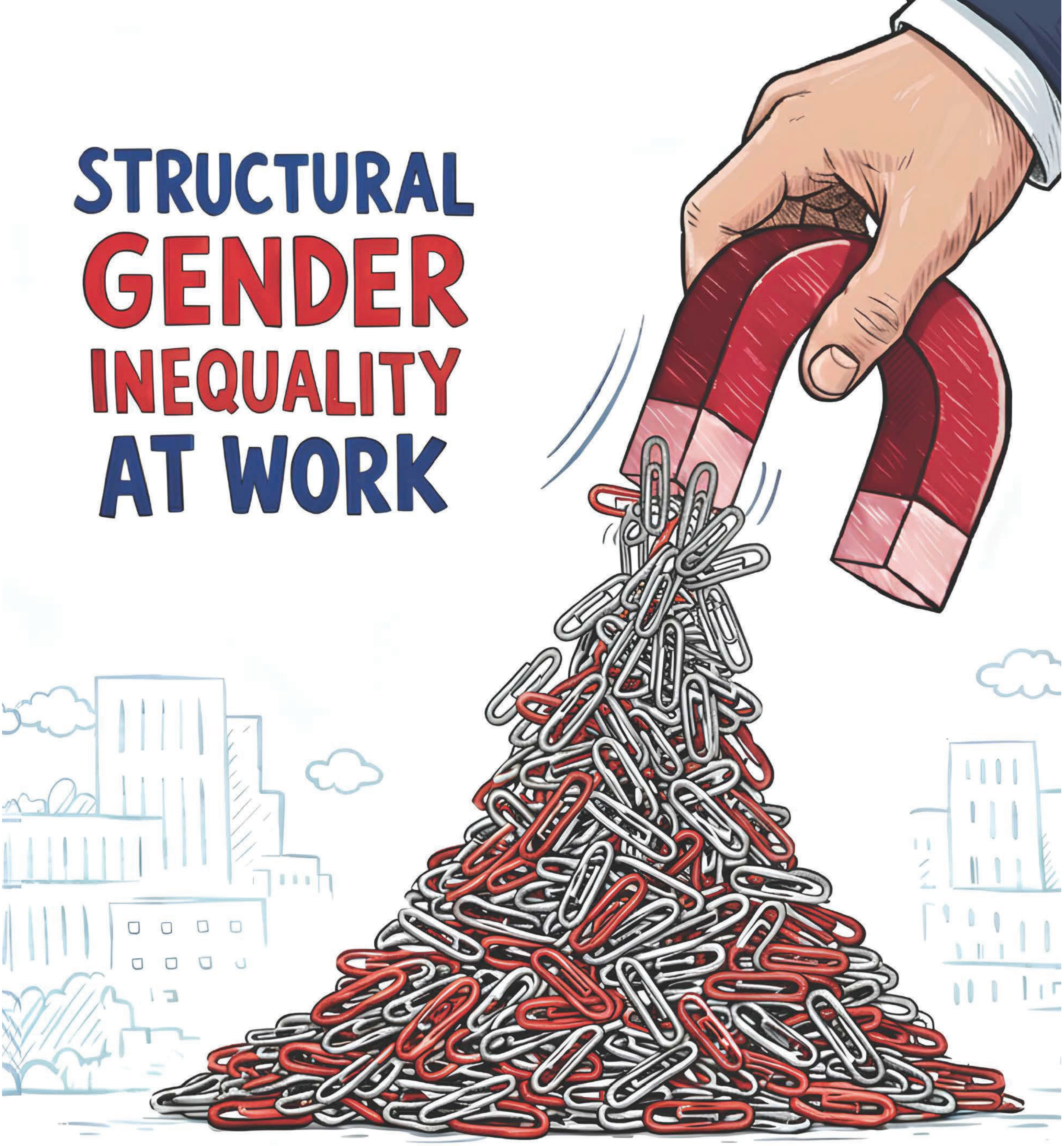
STRUCTURAL GENDER INEQUALITY AT WORK



How integration and the accumulation of advantages shape (in)equality

Sanjana Singh

Structural Gender Inequality at Work

How integration and the accumulation of advantages shape (in)equality

Sanjana Singh

Structural Gender Inequality at Work: How integration and the accumulation of advantages shape (in)equality
Sanjana Singh

ISBN: 978-90-393-8035-2
DOI: https://doi.org/10.33540/3414

Cover design: Srishti Singh
Layout and design: Srishti Singh and Sanjana Singh
Printing: Gildeprint, gildeprint.nl

Structural Gender Inequality at Work

How integration and the accumulation of advantages shape (in)equality

Structurele genderongelijkheid op het werk

Hoe integratie en de accumulatie van voordelen (on)gelijkheid beïnvloeden

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof. dr. ir. W. Hazeleger,
ingevolge het besluit van het College voor Promoties
in het openbaar te verdedigen op

woensdag 8 april 2026 des middags te 12.15 uur

door

Sanjana Singh

geboren op 20 maart 1993
te Chennai, Tamilnadu, India

Promotoren:
Prof. dr. A.G. van der Lippe
Prof. dr. E. Jaspers

Beoordelingscommissie:
Prof. dr. N. Ellemers
Prof. dr. E.L. Hamaker
Dr. R.M. Kuiper
Prof. dr. W.A.F. Maas
Prof. dr. R.P.M. Wittek

Dit proefschrift is onderdeel van het onderzoeksprogramma Sustainable Cooperation - Roadmaps to Resilient Societies (SCOOP). Dit proefschrift werd mogelijk gemaakt met financiele steun van de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) en het Ministerie van Onderwijs, Cultuur, en Wetenschap (OCW) in het kader van een in 2017 toegekende Zwaartekracht subsidie aan SCOOP (grant number 024.003.025). Dit proefschrift is afgedrukt met financiële steun van de J.E.Jurriaanse Stichting.

Structural Gender Inequality at Work

How integration and the accumulation of advantages shape (in)equality

Table of Contents

CHAPTER 1

Examining Structural Gender Inequality at Work: A Synthesis

Sanjana Singh is the sole author of this chapter.
This chapter benefited from the valuable feedback provided by Eva Jaspers and Tanja van der Lippe.

1.1. INTRODUCTION

In 1963, JoAnn Morgan, one of NASA's first woman engineers, recalls being "immersed in a man's world where everybody around me were men." She had to endure obscene phone calls, lewd comments, and was excluded from meetings and social gatherings (NASA 2019; Smith 2019). JoAnn's situation reflects how women's underrepresentation defined workplace gender[1] inequality at the time. Her experience also highlights how workplaces were structurally designed for men, creating barriers for not only women's entry, but also their integration and career progression (Stamarski and Son Hing 2015; Williams, Muller, and Kilanski 2012).

Fast-forward to the present, women's representation has significantly increased. NASA has over 2300 women engineers, making up about 24% of the engineering workforce (NASA 2021). This progress follows decades of efforts to improve women's representation by quantifying, tracking, targeting, assessing, and managing gender diversity. For example, the International Labor Organization (ILO) has tracked women's employment since the 1930s, supporting policy design and evaluation for gender equality (ILO 1931, 2016). Numerous Gender diversity management (GDM) practices and the United Nations' Millennium Development Goals (MDG) for 2015 focused on improving women's representation. For instance, over 39 countries have gender quotas to increasing the share of women in company boards and management (Belaounia, Tao and Zhao 2020; Deloitte 2019; Terjesen, Aguilera and Lorenz 2015).

However, despite improvements in women's general workplace representation and policies supporting it, their presence sharply drops with career progression. In the U.S., only 5% of Fortune 500 CEOs in 2019 were women, the same as men named James. This stark contrast between women's general representation and their dismal numbers in top leadership positions is seen across sectors and countries (Boyle, Matthew and Green, Jeff 2023). This gender inequality pattern appears not only in representation but also in other career outcomes like pay, promotion, and retention (e.g. Huang et al. 2020; Manning and Swaffield 2008). Its widespread nature is captured in familiar phrases like the "glass ceiling", "leaky pipeline" and "women's drop to the top".

NASA is an exemplary case of how workplace gender inequality has shifted from representational to a structural issue, particularly through two mechanisms: integration and career progression. Historically shaped by male-dominated norms, NASA's leadership has been described as operating through a "good ole boys" network, and its organizational language still includes gendered terms like "manned missions" (LaFrance 2015; Lovell 2021; Reynolds 2021). Such a culture makes it easier for men to integrate—forming friendships, accessing mentorship, and gaining informal support—while women are often left out. This structural inequality is also seen in career progression: although women now make up 34.9% of all NASA employees, they represent only 17.9% of senior level employees

[1] Although gender is a non-binary identity, in this book we look at the gender binary of men and women.

(NASA 2021). In fact the proportion of women at the senior most level has remained largely unchanged for over a decade, signaling a persistent lack of progress in addressing structural inequality.

The growing need to focus on structural inequality is also seen in how global policy goals have changed over time. While UN's Millennium Development Goals (MDG) focused on improving women's general workplace representation, the Sustainable Development Goals (SDG) for 2030, MDG's successor, promotes structural gender equality through equal access to career advancement opportunities and inclusion at all levels of decision-making (United Nations, 2001, 2022)[2]. Furthermore, many organizations are now documenting the impact of structural sexism on key performance indicators, and the use of tools to make such information available in real-time is growing (Tedder-King and Sherf 2024). Thus, gender inequality is increasingly recognized as a structural issue, evident not only in representation but also in career outcomes like income and productivity, as well as in related outcomes like integration and job satisfaction.

However, our understanding of structural gender inequality remains limited. Despite ample evidence of its existence and importance, it has been overshadowed by the focus on improving women's representation and its ease of measurement.

In this dissertation, I aim to understand structural gender inequality in the workplace. Specifically, I focus on examining how two structural mechanisms—employee integration and career progression—can shape gender (in)equality.

This dissertation is part of the 'Sustainable Cooperation: Roadmaps to a Resilient Society (SCOOP)' research program, which investigates how cooperation can be sustained under changing conditions to create valuable outcomes, alongside developing and testing solutions in the domains of care, inclusion, and work. This dissertation contributes to the domain of work, by examining how employee integration and GDM practices can help us understand and support sustainable cooperation between men and women in the workplace.

Cooperation allows employees to achieve more together than they could alone, but changes in the workplace can strain how people work with each other. For example, a new GDM policy that increases the number of women in a team may be seen as unfair or as a threat to meritocracy. This can reduce trust and weaken cooperation. While there are anecdotal examples of such backlash, there is currently no clear or general way to measure and track employee cooperation, and its sustainability over time.

This dissertation shows how sustainable cooperation can be assessed by focusing on employee integration, defined as the group of colleagues an employee likes to work or

[2] Empirically, the MDG tracked organizational gender inequality based on the proportion of women in non-agricultural wage employment. But the SDG now tracks the proportion of women in managerial positions (United Nations, 2001, 2022). While the former was a more static approach, the latter is argued to be more dynamic by better-capturing equality in access to career advancement opportunities, inclusion, and representation in decision-making roles across different levels. Thus organizational gender inequality research has evolved from its static focus on the gender gap in representation to a more dynamic focus on the gender gap in career progression – which remains underexplored.

socialize with. When one gender dominates these groups, it becomes harder for others to integrate due to biased perceptions. For instance, women who are viewed as "diversity hires" often struggle to be accepted. Their successes may be overlooked, while their mistakes are judged more harshly (Elaine Spector 2024). By focusing on integration, we can gain a clearer view of what supports cooperation and what outcomes it produces. In particular, we examine how GDM practices shape integration, and how integration in turn affects career outcomes such as income and job satisfaction.

Additionally, we take a closer look at integration as a key mechanism of structural gender inequality, alongside a second mechanism, cumulative (dis)advantage (CAD). Together, these mechanisms reflect the informal and formal processes that shape inequality in organizations. Using examples from STEM fields, we show how these mechanisms can lead to persistent gender gaps, even when women's representation improves.

Structural Gender Inequality: Integration and Cumulative (Dis)Advantage (CAD)

Organizations are gendered spaces, with language and practices historically developed by men that reinforce divisions of labor, positioning men as the norm and women as secondary (Acker 1990, 2006). Gender differences are structurally created in such spaces and can be challenging to detect. This is clearly illustrated in STEM fields, where women make up just 29.2% of the workforce (WEF 2023).

Women's underrepresentation in STEM is often wrongly framed as a "pipeline" issue, where too few women are qualified for or interested in STEM (Cheryan et al. 2017). Girls perform as well as, if not better than, boys on math tests (Hyde, Fennema, and Lamon 1990; Lindberg et al. 2010), and in countries like the U.S. women represent the majority of STEM graduates (Fry, Kennedy, and Funk 2021). This underrepresentation is then clearly not a pipeline issue.

Research indicates that many talented women either never enter or quickly leave STEM occupations due to barriers like implicit bias, harassment, poor work-life balance, and toxic work cultures (Seron et al., 2016; Hall et al., 2015). Two structural mechanisms can help explain how these inequalities persist despite women's high representation among graduates: (a lack of) integration and CAD.

Our first mechanism, integration, is described as a web of informal bonds including outside socializing, and co-worker friendships (Lincoln and Kalleberg 1985; Lincoln and Kalleberg 1992; Wallace 1995). As offices are social places, integration is essential to access opportunities and reduce gender inequalities (Cullen and Perez-Truglia 2019), which is found to be challenging for women in STEM. For instance, women engineers often face negative interactions with male colleagues which is systematically found to make them feel incompetent or excluded and may lead to burnout. Such effects are not observed in men's interactions with women or among men. (Hall, Schmader, and Croft 2015). Even before graduating, women engineering students are frequently made to feel unwelcome, assigned menial tasks during group projects or internships, stereotyped, and subject to harassment. Male students, in contrast, report that such projects and internships build their confidence and reinforce their sense of belonging (Seron et al. 2016).

Our second mechanism, CAD, is described as the tendency for initial advantages or disadvantages to accumulate over time. It is a well-established mechanism for inequality and reflected in popular sayings like “poor get poorer”. STEM reflects a similar widening pattern with women’s poor representation dropping further with career progression. For instance, women hold 29.4% of entry-level STEM roles, but this drops with career progression to17.8% at the VP level and just 12.4% at the C-suite (WEF 2023). Several factors can contribute to the widening gender gap, affording men a relative advantage and women a disadvantage, which accumulates over their careers. For example, even in the absence of technical degrees men can progress into prestigious technical roles through informal coaching and support. Women without credentials, however, are at a relative disadvantage as they are typically seen as unqualified and denied similar opportunities. Women further accumulate disadvantages as they are often expected to take on unpaid diversity work, such as organizing implicit bias trainings and collecting demographic data. This further limits their time and energy to build technical skills necessary for career progression, widening the gender differences (Luhr 2024).

As integration considers informal ties and CAD considers formal career progression, together they can offer a comprehensive understanding of workplace gender inequality, by capturing both informal and formal organizational processes. However, as gender inequality has mainly been seen as a representational issue, such structural mechanisms have been largely overlooked. Below, we discuss the role of these mechanisms, how they have been historically overlooked, and highlight their potential to both explain and address workplace gender inequality.

Integration

Informal relationships are crucial for career advancement, but they can also reinforce structural gender inequality by creating barriers for women. This issue was clearly exposed by Ellen Pao’s landmark 2012 case, where she sued her employer after being passed over for promotion because of her gender. Her trial, along with a wave of similar sexual discrimination lawsuits, revealed how women’s poor integration limits their career progression. These cases exposed practices like excluding women from work trips because male colleagues thought they might “kill the buzz,” holding all-male dinners and ski trips that blocked women from key opportunities, widened compensation gaps, devalued women’s accomplishments, and outright sexual harassment (Hinchliffe 2017; Moore 2019).

Integration’s role in fostering gender inequality stems from how historically, women were legally excluded from most professional fields until the early 20th century, and even now organizational cultures frequently block career progression for disadvantaged groups (Amis, Mair, and Munir 2020; Muzio and Tomlinson 2012) as seen with women engineers.

At the same time, better integration can reduce turnover and support women’s career progression (e.g. Nie, Lämsä, and Pučėtaitė 2018). Many GDM practices aim to foster women’s networking and mentorship to enhance career outcomes. However, their effectiveness remains mixed. For example, networking with men, who often hold higher social status, can provide access to valuable resources that support career advancement. However, studies have found the proportion of men in women’s networks to be positively, negatively or even unrelated to the their career outcomes (Woehler et al. 2020).

Overall, integration can shape men's and women's career outcomes in both positive and negative ways. Yet, it remains largely overlooked because GDM has traditionally focused solely on increasing women's representation. Gender diversity or equality is frequently equated with, and measured as, a group's gender ratio (Williams and Meân 2004). Consequently, most tools for studying and managing gender inequality continue to rely on these representation-based metrics (e.g. Adams and Ferreira 2009; Fine, Sojo, and Lawford-Smith 2020).

Some gender diversity measures have tried to account for integration indirectly, but they fail to capture actual workplace relationships. For example, a common metric is the proportion of women in management, based on the assumption that reaching managerial positions implies strong integration. However, this assumption is flawed. Women can hold managerial roles without being well integrated, especially if they leave those roles quickly. In fact, women in top management teams turn over at nearly twice the rate of men, despite increased representation (Krishnan 2009). This shows that representation alone does not guarantee equality, and both representation and integration must be considered.

Fortunately, integration can be measured more directly using tools from social network research. This field has long studied how actual workplace relationships affect men's and women's careers (Woehler et al. 2020). Specifically, integration can be assessed through employees' affective cooperative networks, capturing real ties between men and women, such as those formed with colleagues an employee chooses to work or socialize with.

The current exclusion of integration from gender diversity metrics necessitates a re-evaluation of how gender diversity truly associates with outcomes like performance, income, and turnover, as well as the effectiveness of GDM practices. For example, while the gender diversity and performance link is widely studied (e.g. Post and Byron 2015), these analyses rely solely on representation-based metrics, which can yield ambiguous or contradictory results. In contrast, diversity measures based on social network variables offer more precise and direct insights into the causal mechanisms (Reagans, Zuckerman, and McEvily 2004). The criticality of this oversight is further evidenced by GDM sometimes improving gender diversity and some organizational outcomes, while simultaneously leading to widespread backlash and costly litigation (e.g. Leibbrandt, Wang, and Foo 2018). Gender diversity is thus not just measured by who is present, but also by how people connect.

Cumulative (dis)advantage (CAD)

CAD's role in widening workplace gender inequality with career progression is reflected in popular phrased like the "glass ceiling" and "leaky pipeline". Men and women start out equally represented in entry-level jobs, but women's presence declines with career progression, with only 29% of U.S. C-suite executives being women (Krivkovich et al. 2024). This widening gap is not limited to leadership roles but appears in career outcomes across industries and countries. For instance, in the UK, the gender pay gap is negligible at labor market entry but widens to 8% within ten years (Manning and Swaffield 2008).

Understanding this widening gap requires a cumulative approach. For instance, the UK's gender pay gap would appear nonexistent if measured only at labor market entry. However, the consistent widening of inequality that can be observed across many settings, suggests a structural pattern in which men gain and women lose relative advantage over time. Specifically, men's higher social status can give them a head start, which compounds into greater access to resources, recognition, and opportunities, as compared to women. Social scientists describe such self-reinforcing dynamics as CAD, a mechanism for inequality.

Although CAD originated in studies of career inequality, such as Merton's (1968) work on the Matthew effect in science, its application in organizational research has diminished. Methodological challenges have led most theoretical and empirical developments of CAD to emerge outside the workplace context. In 2006 DiPrete and Eirich highlighted the significance of these challenges, noting that most career researchers in sociology focus elsewhere, with CAD being implicitly studied at most. However, the situation has not evolved much since then, largely due to the difficulty in separating CAD effects from unmeasured heterogeneity – the endogeneity challenge. For instance, consider the earlier gender pay gap example, which widened from almost 0 to 8% in 10 years. It would be difficult, if not impossible, to determine how much of the 8% comes from the cumulative disadvantage faced by women (the CAD effect) and how much comes from difference in education, intelligence or other characteristics (unmeasured heterogeneity).

One way of overcoming this methodological limitation is through experimental randomization[3], which is challenging to implement and rarely occurs naturally within organizations. As a result, CAD research has mostly developed outside organizational studies, particularly in education, gerontology, and psychology, where causal effects are easier to identify using methods like random assignment or twin studies (e.g. Stienstra et al. 2021).

However, despite advancements, current approaches to studying CAD still lack the sophistication to examine how different CAD factors interplay to create inequality. For example, they cannot fully explain how CAD from gender and race together shape income inequality among employees.

Summarizing, integration and CAD can improve our understanding of how gender inequality is not just structurally reinforced but also how it can be reduced based on employee networks and early career dynamics. This however requires developing tools that extend the study of gender diversity from a representational issue to one that also captures actual integration. In addition, to study the role of CAD, we will not only have to break down and address the methodological challenges that have limited its research, but also provide

[3] When individuals are randomly assigned to groups, the groups will have the same average resource level, i.e., there will be no difference in their resources. Thus, if the groups accumulate diverging resource levels, it can be causally attributed to group membership via CAD. If the earlier gender pay gap example was studied through such an experiment, we would start by considering identical employees who have the same pay at entry, i.e., a pay gap of 0%. To control unmeasured heterogeneity here, we would then randomly assign a gender to these employees, creating two identical groups, differing only in gender. This means that thereon any pay difference between the two groups would be caused by their gender. Hence, 10 years later, when women are seen to earn 8% less than their male counterparts, we can clearly identify that there is not only widening gender inequality but that the inequality follows from the accumulation of disadvantages by women and advantages by men. Alternatively, we could have randomly assigned different pay to all employees irrespective of their gender. We could then study the change in average pay in the ten years to causally understand how men and women differently accumulate (dis)advantages.

a foundation for future research in this area. In this light, this dissertation seeks to address the research question:

How is gender inequality structurally created through employee integration and their accumulation of (dis)advantages?

1.2. Theoretical and Methodological Foundation

This section introduces four key premises underpinning the dissertation. Each premise begins with a core idea related to workplace gender equality, followed by a summary of its strengths and limitations. Subsequent chapters build upon these premises.

Integration is Essential for Gender Diversity

Gender equality is equated with improving gender diversity, typically measured by women's representation (e.g. Giannetti and Wang 2023). Historically, this made sense, as gender inequality was viewed primarily as an issue of underrepresentation. However, simply increasing the number of women does not guarantee equality, especially if women remain excluded from informal networks critical for collaboration, learning, and career advancement.

This limitation for instance is illustrated through California's gender quota law, which effectively boosted women's board representation but also caused unintended consequences. It reduced support for female board members, worsened women's treatment at work, increased their turnover, and sometimes led to fewer women being hired and shareholder backlash (Bian, Li, and Li 2024; Gertsberg, Mollerstrom, and Pagel 2021).

As workplaces are social environments, employee integration is crucial for gaining access to resources and opportunities. For example, women's chances of securing board positions significantly increase if they participate in informal activities such as golfing (Agarwal et al. 2016). Similarly, informal interactions like cigarette breaks between men and their male managers can explain a third of the gender gap in promotions (Cullen and Perez-Truglia 2023). While such seemingly casual interactions can greatly enhance career prospects, women often have limited access to them, exacerbating gender inequalities in for example pay, promotions, and job satisfaction (Cullen and Perez-Truglia 2023; Grissom, Nicholson-Crotty, and Keiser 2012).

This dynamic is captured by the "old boys' club" hypothesis, suggesting men advance more easily by networking and interacting with powerful male colleagues in ways typically less accessible to women (Cullen and Perez-Truglia 2019). Supporting this, a survey of over 100,000 employees found that 81% of women felt excluded from workplace socialization, attributing their exclusion to repeated behaviors such as being ignored in meetings, excluded from client lunches, or sidelined from informal events like cigar nights or club visits (Annis and Gray 2013). Ironically, 92% of men in the same study did not perceive themselves as excluding women, highlighting how normalized male-centered workplace norms can unintentionally reinforce gender inequality.

Recognizing the role of integration, some researchers have attempted to study it. However, they often rely on indirect measures, such as assuming participation in golf or smoking equates to meaningful relationships. This ignores actual relationships between non-participants and can falsely assume connections among participants who may not interact closely.

Efforts to include integration in gender diversity metrics also rely on flawed proxies, like the share of women in management roles (e.g. Kalev, Dobbin, and Kelly 2006). This approach assumes that managers are inherently well-integrated, but it does not capture actual relationships. In fact, women in top management leave at nearly double the rate of men (Krishnan, 2009). This shows that representation does not always translate to true inclusion, highlighting the need to study integration alongside representation.

Although current diversity measures are limited, *network research* offers tools to directly assess integration, through employees' affective cooperative networks: the group of colleagues an employee likes to work or socialize with. This literature also provides a strong foundation for understanding how social ties influence access to resources and opportunities, shedding light on gender inequalities.

Networks research in organizations has found men and women to differ in both what their networks look like, and their network returns. Given men often hold higher-status positions, network gender composition significantly impacts access to opportunities and career success. The earlier golfing example highlights this: playing a traditionally male sport provided women access to valuable networks in male-dominated environments.

Similar patterns persist across professions, with male-dominated networks consistently excluding women and impeding their career progress. For instance, male scientists often exclude women from professional networks or create unwelcoming environments (Amis et al. 2020; Blickenstaff 2005; Cockburn 1988; Etzkowitz, Kemelgor, and Uzzi 2000). Two common ways to measure network gender composition are:

- Homophily: proportion of same-gender colleagues in one's network.
- Heterophily: proportion of opposite-gender colleagues in one's network.

Despite being the most commonly used measures to study how networks influence career outcomes, they produce mixed results (Woehler et al. 2020). For example, there is empirical evidence on women's careers benefitting from having networks that are both predominantly homophilous (e.g. Ibarra 1997)and heterophilous (e.g. Aten, DiRenzo, and Shatnawi 2017).

The case for homophily follows from theories of knowledge and exchange, arguing that working with similar others increases trust, cooperation, and resource sharing, leading to better career outcomes (Ertug et al. 2018; Hansen 1999; Reagans and McEvily 2003). But, as noted earlier, women also benefit from ties with men, especially in male-dominated environments. This case for heterophily follows from expectation states theory, whereby status beliefs from gender stereotypes associate men with greater status and competence in the workplace (Ridgeway 1991, 2001).

However, neither homophily nor heterophily directly measure integration or the true gender diversity of a network. A mixed-gender team with strong, inclusive relationships will score low on both measures, even though it is well integrated. This shows a disconnect between how network research is conducted and how gender diversity is understood. As a result, integration is often overlooked or misrepresented, limiting our ability to fully understand and address workplace gender inequality.

This gap is compounded by the methodological challenges of network research. Unlike traditional survey methods, full social network analysis of a group requires complete data from all members of that group, not just a random sample. While this improves internal validity and helps reveal how relationships influence outcomes, such as through Stochastic Actor-Oriented Models (SAOM), it also makes data collection far more demanding.

As random sampling is often unsuitable, collecting full, large-scale network data is difficult. Many studies rely on small samples or single organizations, limiting the generalizability of their findings. Network data collection is also resource-intensive and frequently affected by missing responses or incomplete participation, which can introduce bias (Huisman 2014). These challenges become even greater in longitudinal research, where tracking networks over time is complicated by staff turnover, department changes, and limited access to organizations.

Gender Diversity Assessment in GDM

Relatedly, for the comparative study on the effectiveness of GDM, many organizations need to be sampled. Organizations are increasingly adopting GDM to address workplace gender inequality. This adoption is supported by legislative actions, such as board gender quotas now in over 39 countries since it was first introduced by Norway in 2008 (Belaounia, Tao and Zhao 2020; Deloitte 2019; Terjesen, Aguilera and Lorenz 2015). Private investment in GDM has also surged, with Fortune 500 companies spending over $16 billion on diversity management in 2017 alone (Staley 2017).

Broadly, GDM follows *three main approaches* (Kalev et al. 2006). The first focuses on “establishing organizational responsibility”, placing formal pressure on institutions to hire or promote more women, such as through quotas. The second focuses on “moderating managerial bias” through training and performance evaluations. Unlike the first, this approach recognizes the role of integration, as managers significantly influence both workplace relationships and career outcomes. For example, male managers often show in-group favoritism by socializing more with male subordinates, such as during smoking breaks, which can accelerate those subordinates’ pay grade progression (Cullen and Perez-Truglia 2023). The third approach addresses integration most directly by “reducing women’s social isolation”, often through networking and mentoring programs.

However, there is limited empirical evidence on the actual effectiveness of specific GDM practices. Scholars often assume practices addressing known causes of inequality will naturally reduce it, but this assumption remains largely untested (Kalev et al. 2006). This false assumption is highlighted by GDM’s widespread criticism from multiple stakeholders, including frustration from minority activists over slow progress (Köllen, Kakkuri-Knuuttila, and Bendl 2018), “reverse discrimination” lawsuits from majority groups, and concerns that organizations use GDM as mere propaganda to avoid litigation, making discrimination harder to identify (Dover, Kaiser, and Major 2019).

These issues likely stem from a lack of evidence-based design or limited evaluations on GDM's effectiveness. Moreover, these evaluations narrowly focus on whether GDM improves representation, rather than examining whether women are actually included in workplace networks. This is partly because representation is easier to measure and has been traditionally used to measure gender diversity. As a result, practices that only improve numbers may appear effective at improving such 'gender diversity', even if they worsen gender inequality in other areas.

While GDM can have both positive and negative effects, most evaluations fail to capture the latter, particularly on integration. For instance, while gender quotas successfully increase women's presence, they may unintentionally reinforce negative stereotypes or trigger resentment among male colleagues. These perceptions can lead to social exclusion, negative stereotypes, foster resentment among male colleagues and is known to create forums on "Ever work for a woman? Share your horror story" (Bowles 2017). Women are also often excluded from informal networking opportunities critical for career advancement, increasing gender gaps in pay, promotion rates (Cullen and Perez-Truglia 2019), job satisfaction and turnover (Grissom, Nicholson-Crotty and Keiser 2012).

Yet most current evaluations of GDM effectiveness still rely on representation metrics (e.g. Dobbin and Kalev 2013; Kalev, Dobbin and Kelly 2006; Timmers, Willemsen and Tijdens 2010). As a result, existing evaluations of GDM are often incomplete and lack external validity. They typically assess one practice, often not grounded in theory, and are usually conducted in firms already committed to diversity. This introduces selection bias, overlooks integration, and limits generalizability to less committed or resource-constrained organizations. Additionally, these evaluations rarely account for contextual factors like organizational culture, industry norms, or national labor laws, limiting the generalizability of results.

The reliance on representation-based gender diversity metrics has created an ineffective incentive structure for GDM. Many practices are designed to meet legal requirements rather than address underlying social dynamics, often becoming symbolic gestures that may unintentionally reinforce inequality (Dover et al. 2019). For instance, gender quotas may fail if recruiters continue to rely on traditional and oft male-dominated hiring channels (Bowen, Swim, and Jacobs 2000; González, Cortina, and Rodríguez 2019). The narrow focus on representation not only hides ineffective or harmful practices, but also overlooks more inclusive approaches like recruiting from minority-serving institutions or using open calls, which could structurally improve gender equality (Dover et al. 2019).

CAD in Career Progression

CAD is a well-established theory used to explain persistent inequalities across many fields. It is described as the systemic tendency for interindividual divergence in a valued resource over time. This divergence is not about individual traits, but about how initial advantages or disadvantages accumulate to widen resource gaps. CAD's broad acceptance and applications make it a strong theoretical foundation for understanding workplace inequalities, including gender differences in career progression.

CAD typically arises when performance is ambiguous or hard to measure directly. In such situations, decision-makers rely on indirect indicators or "signals" of competence to allocate resources and opportunities. Depending on whether these signals are perceived as positive or negative, they shape how resources are distributed over time. Broadly, CAD operates in three forms, each defined by the kind of signal that initiates the accumulation of a relative (dis)advantage:

Path-dependent CAD was the first form to be identified and is triggered by early signals of success, reflected in the popular saying "success breeds success". This dynamic, also known as the Matthew Effect, was first described by Merton (1968) in his influential work on scientific careers. He observed that a scientist's early recognition led to increased access to resources, which in turn reinforced future success. These insights were generalized to produce a dynamic theory of stratification and mathematically systematized to show that CAD predictions were consistent with observed characteristics of scientific careers (DiPrete and Eirich 2006). More recent evidence of this CAD form is seen again in scientific careers, with scientists who win an initial grant being more likely to secure additional grants, regardless of abilities (Bol, de Vaan, and van de Rijt 2018).

Status-dependent CAD emerged later in the literature and is triggered by signals from ascribed status characteristics, like gender, race, or age, rather than past achievement or path (Berger, Rosenholtz, and Zelditch 1980; Dannefer 2020; DiPrete and Eirich 2006). The actor's relative status, high vs low, shapes the resources and opportunities they receive. For example, if being white is viewed as higher status than being a person of color, income gaps between the two groups can widen over time—creating cumulative advantage for whites and disadvantage for the colored group. This was shown in Blau & Duncan's (1967) study on race and occupation. Gender inequality follows a similar pattern, with men seen as higher status and more competent than women (Berger et al. 1977; Ridgeway 2001), systematically granting men greater access to career opportunities which consequently widens gender inequality with career progression.

CAD interactions are the most recent and least explored form, triggered when actors have multiple signals. For example, women experiencing early career setbacks might face larger disadvantages than men, as they are penalized both for their gender and their initial failures. Men, by contrast, may more easily overcome similar setbacks due to their higher social status. This interplay means the effect of one CAD factor depends on the other, producing outcomes that cannot be independently explained by each factor alone.
Early research on CAD interactions typically involved simple comparisons of only selected groups. For instance Cole & Singer's (1991) theory of limited differences looked at gender and productivity in scientific publishing by comparing high-achieving women vs. high-achieving men, neglecting other important comparisons like high-achieving men vs. low-achieving men. This limits a comprehensive understanding of CAD by failing to show how resources diverge across all groups formed by the interaction of gender and success, like say high-achieving and low-achieving men. While recent research has advanced, it continues to overlook certain group comparisons, often emphasizing status differences at similar resource levels but not exploring fully how outcomes differ within the same status category.

Significant theoretical and empirical advances have been made since Merton's pioneering work, offering valuable frameworks for studying how early signals influence long-term outcomes. The emerging research on CAD interactions is particularly promising because it helps examine how multiple (dis)advantages, like gender and early (un)success, combine to both widen and reduce inequalities.

Unlike the former two CAD forms, recent research shows that CAD interactions between gender and career outcomes can both reinforce and reduce inequality (e.g. Pedulla 2016; Rivera and Tilcsik 2016). This highlights CAD interactions' untapped potential for explaining and addressing workplace inequalities. Its growing relevance is recognized across both the CAD and intersectionality literatures, particularly for questions on inequality (e.g. Bauer et al. 2021; Cho, Crenshaw, and McCall 2013; Erola and Kilpi-Jakonen 2017; Ross and Mirowsky 2006; Stienstra et al. 2021).

These gaps are further compounded by fragmentation in the CAD literature, as much of the progress has occurred outside organizational research. This is largely due to the difficulty of causally identifying CAD effects in workplace settings. Experimental methods often used for causal identification are more feasible in fields like education and psychology than in organizational research.

Another challenge in organizational contexts is the lack of standardized, long-term data that tracks employee careers across firms using consistent outcome measures. For example, to study the career progression of a sales employee, we would need performance records that follow them as they move between firms—despite each company using different systems to evaluate success, such as revenue targets, client retention, or peer feedback. Without comparable data across time and settings, it becomes difficult to assess how career outcomes evolve.

Finally, the absence of clear, structured guidelines for studying and testing CAD interactions prevents researchers from developing consistent approaches. This limits the generalizability of findings and slows overall progress in understanding and addressing structural gender inequalities in workplaces.

Methods and Measures for Studying Gender Diversity

Traditional gender diversity measures and evaluations were not designed to capture structural inequality. It is therefore unsurprising that gender inequality has decreased in terms of representation while increasing in integration, and with career progression. These trends are partly shaped by limitations in commonly used empirical approaches, which influence what researchers can observe and assess.

The widespread use of *regression mod*els has helped quantify conditional associations between gender and career outcomes. For example, an OLS regression might estimate how future income differs by gender, controlling for background variables like age or education. However, in the absence of experimental or quasi-experimental designs, as is often the case with organizational data, these associations are not causal and may be biased due to endogeneity. It is difficult to separate the effect of gender from other unobserved factors like cognitive ability, particularly in CAD research. For example, when the gender pay gap widens over time, it's hard to tell how much is due to women's accumulated disadvantage versus unmeasured traits like IQ.

Null hypothesis significance testing (NHST) remains a dominant method for evaluating statistical effects. While useful for falsifying hypotheses of no difference, e.g., testing whether men and women earn the same income, NHST is poorly suited to questions involving multiple comparisons, inequality hierarchies, or evidence in favor of a null effect. This is often due to issues like multiple testing and low statistical power (Gu, Mulder, and Hoijtink 2018; Hoijtink 2011), which hinder evaluating more complex gender dynamics. For example, identifying the most effective GDM practice (e.g., diversity training vs. networking vs. mentorship programs), assessing whether network gender diversity benefits employees, or determining whether men and women have equal career outcomes.

Social network analysis offers rich insight into how individuals integrate within teams, making it a valuable tool for studying workplace gender inequality, such as through employees' cooperative networks. For example, network studies have shown that men and women differ in how they build workplace networks and the benefits they receive from them (Woehler et al. 2020). However, collecting such data is resource-intensive, requiring full information on all individuals in a network, including their gender, collaboration patterns, and future outcomes. Ideally, the data should be collected longitudinally to track how networks and inequalities evolve, making the process prone to missing data, which can introduce bias and reduce reliability.

Fortunately, *methodological innovations* can support more rigorous analysis of causal relationships, enable the evaluation of more informative hypotheses, and address missing data. Yet these tools remain underused in gender diversity research due to disciplinary silos. For example, economists often use techniques such as Two-Stage Least Squares Instrumental Variables (2SLS-IV) and difference-in-differences regressions to identify causal effects. Social psychologists often use Bayesian informative hypothesis evaluation (BAIN) to explore more nuanced questions. Statisticians have shown that missing data can be effectively handled using robust techniques like Multivariate Imputation by Chained Equations (MICE), which performs well even with complex models like multilevel regressions—a common approach for addressing nested data structures in organizational research.

While these methods have advanced the study of structural gender inequality, several limitations persist. First, experimental and network-based data are often drawn from a single organization or group due to access challenges, which limits external validity, i.e. how much we can generalize the results to other contexts. Second, studying structural gender inequality faces the same endogeneity problem found in CAD research. While experimental methods can help, they are rarely feasible in real-world workplace studies.

Third, these studies often analyze interaction effects, such as how networks or early career experiences affect men and women differently. But such interaction effects are prone to misinterpretation, especially if researchers fail to treat interactions as a combined or "global effect". This increases Type-IV errors, where a hypothesis is correctly rejected but its meaning is misunderstood or misapplied.

Finally, current approaches largely adopt a static view, showing only a snapshot of gender inequality rather than how it accumulates or evolves over time. This is a key limitation, since gender inequality often varies over employee careers. For example, the gender pay gap is consistently found to widen as careers progress (Barth, Kerr, and Olivetti 2021).

1.3. Our Contribution

This dissertation advances the understanding of structural gender inequality in workplaces by examining two critical yet understudied mechanisms: integration and CAD. By focusing on how these processes shape individual careers and workplace outcomes, we provide novel tools and frameworks for studying gender inequality. This dissertation makes three key contributions.

The first contribution shifts the focus of gender diversity from a simple group-level measure, like counting how many women are in teams or management, to the individual level by explicitly studying employee integration. Integration refers to how men and women form friendships, collaborate, and support each other at work. Traditional diversity metrics overlook these critical social dynamics despite research highlighting their importance for employee productivity and creativity. For example, teams with similar proportions of men and women may produce vastly different results depending on how well employees interact and form supportive relationships. Yet integration is rarely measured or analyzed directly.

We address this gap using unique, detailed network data from the European Sustainable Workforce Survey (2017–2018), covering more than 4,000 employees across nine European countries and six sectors. Unlike typical network data, which focuses on a single context, this rich nested dataset captures detailed employee networks and organizational gender diversity practices across diverse environments. Such data allow us to comprehensively study gender diversity by accounting for individual, team, and organizational dynamics, along with evaluating the effectiveness of diversity practices in ways previously not possible.

The second contribution challenges the conventional view of CAD as a mechanism that always increases inequality. CAD traditionally explains how early success or high status reinforces further advantage, captured in phrases like "the rich get richer." However, we demonstrate that CAD interactions, such as between gender and early career success, can both amplify and reduce inequality. We propose a nuanced theoretical framework to examine how multiple forms of advantage and disadvantage interact to shape future outcomes, moving beyond single-factor explanations.

To do this, we reintroduce CAD into organizational research, where it has remained underexplored due to difficulties in experimentally identifying causality. Using large-scale administrative data from the United States Patent and Trademark Office (USPTO), which covers 2.8 million patent applications from 2001 to 2020, we overcome this challenge. The quasi-random assignment of patent examiners allows us to causally examine how early career outcomes shape long-term trajectories, providing rare and valuable insights into gender disparities over professional careers.

To rigorously analyze CAD interactions, this dissertation uses simulation studies to examine how common analysis methods perform. Simulations allow the creation of datasets with known characteristics, helping to isolate the effects of factors like small effect sizes, group similarity, and model selection. This approach makes it possible to identify when statistical errors are most likely and to develop clear guidelines for detecting and interpreting CAD interactions accurately. These insights are essential for improving how inequality is studied and informing effective policy design.

Our third contribution is the development of tools to systematically measure and analyze both integration and CAD. For integration, we use established insights from social network research to develop a new, generalizable measure reflecting actual employee relationships. We also employ advanced Bayesian methods to rigorously evaluate competing hypotheses, resolving previous inconsistencies in the network literature. This approach allows detailed analysis of how different network types distinctly benefit men and women and reveals the impact of team gender composition, including phenomena like tokenism. These insights are essential for both individual career strategies and organizational policies, especially in the context of recent backlash against diversity initiatives.

For CAD, we offer practical guidelines on how researchers can effectively theorize, analyze, and interpret CAD interactions. Despite CAD's widespread recognition, its interaction effects remain poorly understood and frequently mishandled in empirical analyses. Our guidelines help researchers across disciplines avoid common analytical pitfalls and provide reliable insights into how inequalities develop or diminish over careers.

Together, these contributions offer new tools, perspectives, and robust evidence for understanding why gender inequality persists structurally in workplaces. By centering integration and CAD, leveraging diverse and high-quality data, and refining analytical approaches, this dissertation significantly advances knowledge of structural gender inequality and concretely supports efforts to create fairer workplaces.

1.4. Chapter Overview

Table 1.1 Chapter Overview

Ch.	Main Research Question	Main Findings	Sub-questions	Findings	Data	Method
2	How does networking with men and women relate to career success?	Diverse networks benefit both men and women. Gender differences in network returns disappear when women are not severely underrepresented.	Do employee careers benefit from gender-diverse networks?	Yes.	Second wave of the European Sustainable Workforce Survey, 2017-2018.	Analysis of Covariance (ANCOVA) multilevel model and Bayesian informative hypotheses evaluation
			Are diverse, homophilous, or heterophilous networks better for men's and women's careers?	Gender-diverse networks are generally better for both.		
			How does team gender composition affect network benefits?	When women are not the minority, men and women have similar network returns.		
3	Do organizational GDM practices actually reduce gender inequality?	The effectiveness of GDM depends on its evaluation.	Is having a GDM practice better than not having any GDM practice?	No, except practices with an individual approach which appear effective at improving representation.	Second wave of the European Sustainable Workforce Survey, 2017-2018.	Linear regression and Bayesian informative hypotheses evaluation
		Integration must be considered to account for GDM's negative unintended consequences.	Does the evaluation of a practice's effectiveness depend on the gender diversity measure used?	Yes.		
			Do practices appear more effective on just representation-based measures, than those including integration?	Yes.		

4	How is the interaction of CAD factors studied?	Research remains fragmented and focused on the single CAD mechanism. Interaction effects are often neglected (Type-II error ~68%) or misinterpreted (Type-I error ~47%). These errors can be addressed through standardized guidelines for theorizing, analyzing, and interpreting CAD interactions.	Are all relevant group comparisons being considered? Can CAD interactions be analyzed using conventional step-wise regression? How should CAD interactions be analyzed? Are results accurately understood? How can non-significant results be better interpreted?	No. Visualizing the theory using line graphs can resolve this No, it leads to biased and invalid results. Full-model regression should be used to estimate all possible group comparisons, without adjusting p-values. Not always. There's a high risk of false positives for main effects and missed interactions. Scrutinize using graphs, data, and power. More likely with small effects and high homogeneity.	441,000 CAD estimates from 441 unique combinations of group sizes, main effects, and interaction effects—each simulated with 1,000 datasets of 1,000 observations.	Type I and Type II error estimation: comparing CAD estimates from regression simulations to predefined scenarios
5	Can CAD explain workplace gender (in)equality?	CAD can widen and reduce inequality: successful women and unsuccessful men catch up, but unsuccessful women fall far behind, producing only half as much.	Can employees recover from initial setbacks? Can initial success reduce women's disadvantages? Does the gender–success CAD effect change over time?	Men can but women cannot. Yes. Yes. Most disparities shrink over time, except for initially unsuccessful women, who continue to fall further behind all other groups.	Population data of over 2.8 million patent applications filed with the USPTO, 2000-2020	2SLS-IV regression with fixed-effects

Chapter 2: Should Birds of a Feather Flock Together? How Networking with Men and Women Relates to Employees' Career Success

This chapter explores how gender diversity in employee networks shapes men's and women's career outcomes. While much of the research on workplace diversity focuses on group-level representation, we shift attention to the individual level by studying integration—the informal relationships employees form at work. We measure this through affective cooperative networks, which include the colleagues employees prefer to work or socialize with.

Past studies on network gender composition have produced mixed findings. Some suggest that homophily, or same-gender networks, supports career advancement, while others find heterophily, or opposite-gender networks, to be more beneficial. And, some studies report no clear effect at all (Woehler et al. 2020).

Novel approach to studying network gender composition

To address these inconsistencies, we draw on the gender diversity literature. Rather than focusing only on the proportion of same or other-gender colleagues, we introduce a third possibility: gender-diverse networks that include both men and women. These balanced networks may offer distinct advantages that neither homophilous nor heterophilous networks can provide.

We take a new approach by categorizing networks into three types: homophilous, diverse, and heterophilous. This allows us to test how each network type relates to two distinct measures of career success: job outlook and income. This follows from the implication of network characteristics varying depending on how career success is measured (Woehler et al. 2020). While job outlook is a forward-looking subjective measure based on the employees' opinion on their prospects in the firm, income provides a static objective measure for career success.

We also examine how a team's gender composition influences the link between network type and career success. Since network composition reflects both personal preferences and interaction opportunities, we draw on Kanter's tokenism theory (1977c) to assess whether women's minority status in a team shapes interactions.

We simultaneously analyze both career success measures using an uncentered ANCOVA multilevel model and apply Bayesian hypothesis evaluation to understand how men's and women's careers relate to homophilous, diverse, and heterophilous networks. We use data from the European Sustainable Workforce Survey (2017–2018), covering 4,345 employees across 401 teams, 113 organizations, 9 countries, and 6 industries. This multilevel dataset captures organizational structure while allowing analysis of individual networks and outcomes.

Findings and implications

Despite most employees forming same-gender networks, our findings show that gender-diverse networks are generally most beneficial. The exception is women's income, which is more strongly linked to male-dominated (heterophilous) networks. Additionally, in teams where women were not severely underrepresented, there were no gender difference in how the three network types benefit careers.

These results highlight the importance of looking beyond representation to how employees connect, as integration plays a key role in shaping career outcomes. This approach can explain inconsistencies in past network and diversity research and inform more effective GDM. In the next chapter, we further examine the effectiveness of GDM practices at improving representation and integration using different diversity measures.

Chapter 3: Are Organizational Gender Diversity Management Practices Effective?

This chapter revisits how gender diversity is measured and evaluates the effectiveness of GDM practices. Diversity measures serve a dual role: they track inequality and guide how institutions respond to it. For this reason, diversity measurement carries important policy implications.

Traditional measures focus on representation, leading to GDM practices that often overlook integration. However, as discussed in Chapter 2, representation alone is insufficient. Women's poor integration, even when their numbers improve, can reinforce inequalities. For example, when women are hired under diversity policies, it may trigger perceptions of lower competence, discourage collaboration, and hinder their integration. While such unintended consequences deepen gender divides, representation-based measures fail to capture them and may misleadingly suggest improved gender equality.

Measuring integration alongside representation

Building on the diversity measure introduced in Chapter 2, we propose a novel network-based measure of gender diversity that quantifies integration within teams. Given the critical role of integration, earlier studies have tried to capture it indirectly by focusing on women's representation among managers (e.g. Kalev et al. 2006). This approach assumes that managerial roles are harder to integrate. Yet this does not capture actual social dynamics, and serves only as an untested proxy. In contrast, we focus on employees' affective networks, using real interaction data.

We evaluate GDM effectiveness in teams using both the current representation-based measure, focused on women's share in management, and a new integration measure built from employees' cooperative networks described in the previous chapter.

We examine how these two measures differ in evaluating the effectiveness of the three most common GDM approaches, which operate at different levels. At the individual level, practices aim to reduce isolation through tools like networking and mentoring programs. At the managerial level, the focus is on moderating bias, often through training and performance evaluations. At the organizational level, practices aim to institutionalize responsibility by creating diversity task forces, appointing diversity officers, or adopting affirmative action plans.

GDM practices fall short, especially on integration

We found GDM practices to be largely ineffective on both measures. Nevertheless, they appeared more effective at improving representation than integration. This could be because most policies prioritize increasing representation, sometimes at the expense of integration.

When we examined the three types of GDM approaches separately, only the individual level approach was effective at improving representation. However, none improved integration. This discrepancy highlights the danger of relying solely on headcounts when evaluating gender equality.

Implications for Research and Policy

Our findings have important implications for how GDM is evaluated, implemented, and understood. Integration should be a central goal, not an afterthought. Simply hiring more women is unlikely to reduce inequality if there are no affective connections between men and women.

Institutions are therefore advised to prioritize retention and integration before expanding recruitment efforts. Using network-based measures of diversity can offer a more accurate picture of workplace dynamics and highlight areas where exclusion persists. It is also important to recognize and address GDM's unintended consequences, such as reinforcing gender stereotypes or undermining women's credibility.

The proposed network-based measure offers a practical and scalable tool for assessing GDM effectiveness. It can be implemented by any organization using digital communication systems and requires only modest resources.

Chapter 4: Studying Interaction of Cumulative (Dis)Advantage Factors: What Have We Done?!

While integration matters, other forces also shape gender inequality at work. We explore CAD as one such mechanism. Chapter 4 develops practical guidelines for studying CAD interactions, which we apply in Chapter 5 to analyze gender inequality among inventors using large-scale data.

Since CAD research remains fragmented across disciplines, developing a general guideline for CAD interaction research requires both a broad methodological perspective and an understanding of CAD's cross-disciplinary relevance. As a result, the approach in Chapter 4 differs considerably from the other chapters in this dissertation.

Why CAD interactions matter and are challenging to study?

CAD explains how early advantages accumulate over time. Its importance in understanding inequality is reflected in sayings like "the rich get richer" and "success begets success." These phrases also illustrate how CAD has traditionally focused on a single factor, dividing people into two groups: those with cumulative advantage and those with disadvantage. However, this two-group view oversimplifies reality, because individuals often experience multiple, interacting CAD factors, leading to more than two groups.

For instance, a successful person from a low-status household experiences both "poor get poorer" and "success breeds success" effects. To understand how these two CAD processes interact to shape inequality, we cannot rely on simple two-group comparisons, such as between successful and unsuccessful peers. Instead, we must consider all comparisons among the four possible combinations of success and social status: (i) successful individuals from low-status backgrounds, (ii) unsuccessful individuals from low-status backgrounds, (iii) successful individuals from high-status backgrounds, and (iv) unsuccessful individuals from high-status backgrounds. To fully capture how inequality develops and evolves, CAD interactions must be studied by comparing how future outcomes differ across all these groups. However, traditional approaches to analyzing interaction effects often fail to capture this complexity.

This is crucial as studies like Bernardi (2014) and Pedulla (2018) show that CAD interactions can sometimes reduce inequality, rather than always widen it. This dual potential makes CAD interactions especially valuable for both research and policy. Yet, they remain understudied, fragmented across disciplines, and vulnerable to errors—largely due to incomplete comparisons and the lack of standardized methods.

How do we study CAD interactions?

To address these challenges, we used a four-group interaction model combining success and status: SL, UH, UL, and SH. This setup reflects the two most common CAD types in the literature—path-dependent (early success) and status-dependent (social position). The 2x2 structure is also widely applicable across fields, including in the seminal works of Bernardi (2014) and Pedulla (2018).

Since CAD effects reflect how resources diverge among groups, it can be estimated through seven group comparisons: six direct pairwise and one indirect or difference-in-difference comparisons. These can be analyzed using two complementary methods. The first is using regression coefficients, which estimate three comparisons: the two main effects and the interaction effect. The second is through marginal means estimation, which offer an alternative for estimating all seven comparisons.

All seven comparisons can be visualized in a single resource distribution graph, forming the basis of our framework to examine how CAD interaction is currently theorized, analyzed and interpreted. Specifically, we identified pitfalls and developed resolutions using simulation analysis.

What did simulations reveal?

We tested common analysis methods using simulations. Stepwise regression, one of the most widely used methods, performed poorly. It missed true interaction effects in 68% of cases and falsely detected main effects in 47%—comparable to flipping a coin. These errors were even worse when sample sizes were small or groups were homogeneous—conditions common in inequality studies, such as those on gender gaps in male-dominated fields like STEM or female-dominated care work.

Implications for Research and Policy

A comprehensive CAD interaction study should theorize all relevant group comparisons. Researchers should use full regression models, regardless of the interaction term's statistical significance. Since all group comparisons are part of the hypothesis, unadjusted significance testing is appropriate, avoiding unnecessary corrections. Finally, non-significant results should be interpreted with caution, especially when effect sizes are small or groups are homogeneous—both of which increase the risk of Type II errors and are common in inequality research.

This study also shows that CAD might not always widen inequality. Interactions between CAD factors like status and success can also reduce disparities. These findings offer practical guidance to help researchers avoid common mistakes and improve how CAD interactions are theorized, analyzed, and interpreted—supporting more accurate and equitable policy responses.

Chapter 5: Cumulative Advantage as a Mechanism for Career (In)Equality: Can Some Women Benefit?

Despite extensive research, gender inequality in career outcomes like pay, productivity and innovation remain poorly understood, with inequalities widening over time. This chapter examines whether this could be explained by CAD.

Although CAD originated from studies on career inequality (e.g. Merton 1968), it has remained relatively understudied in organizational contexts. One reason is the difficulty of studying CAD effects rigorously, given challenges with both internal and external validity.

Internal validity issues stem from CAD's endogeneity problem, given the difficulty in identifying the effects of cumulative processes, as against unobserved factors like talent or motivation. While experimental randomization could address this, it is rarely feasible in real workplaces and can limit the external validity or generalizability of findings. Given that organizations are shaped by complex norms, incentives, and power dynamics, even well-designed experiments may fail to capture how CAD actually unfolds across careers. In contrast, External validity is typically supported through long-term observational data that tracks employee careers across firms and sectors. Ideally, such data would also include standardized outcome measures. Unfortunately, this kind of data is rarely available.

To address these challenges, this study uses large-scale patent data from the United States Patent and Trademark Office and a quasi-experimental design. Internal validity is improved by using the USPTO's random assignment of patent examiners. The strictness of an inventor's first examiner serves as an instrumental variable to identify how the outcome of their first patent affects future innovation (Bohnet, van Geen, and Bazerman 2016; Sampat and Williams 2019). External validity is ensured by tracking inventor careers across firms and sectors, using standardized outcomes. Since all inventors start with a first patent application reviewed under the same rules, it provides a clear and comparable trigger for studying how early outcomes shape long-term careers.

Extending CAD theory through interaction analysis

To our best knowledge, this is the first study that compares all groups in CAD interaction systematically. We build on Chapter 4's approach by theorizing and evaluating the interaction of two CAD factors, gender and initial success. These factors are often studied separately, or at best together, without fully considering their interplay. We address this gap by comparing all possible combinations across the four interaction groups: successful-men, unsuccessful-men, successful-women, and unsuccessful-women.

We found that early success is crucial for women's careers but has limited long-term impact on men's. Despite successful-men being the most advantaged group, both unsuccessful-men and successful-women were able to catch up over time, ultimately achieving similar outcomes. This shows that CAD interactions can reduce or even eliminate inequality for certain groups.

In contrast, unsuccessful women emerged as the most disadvantaged group. Their career outcomes not only lagged behind but continued to worsen over time. They were the only group to suffer large and irreversible setbacks.

Overall, while early setbacks can seriously impede women's careers, they tend to matter less for men. The CAD from men's gender helps them recover, completely eliminating any effects of those setbacks on their career in the long-term.

Implications for Research and Policy
CAD offers a powerful lens to understand how gender inequality becomes structurally embedded over the course of a career. This perspective helps draw on the broader CAD literature to better explain current workplace gender inequality

For policymakers, CAD can highlight when and how interventions are most effective by uncovering the mechanisms that drive inequality. In innovation careers, for example, improving how early work is evaluated may reduce inequality more than simply hiring more women.

Since CAD effects often take root early, late-stage solutions like gender quotas on corporate boards may come too late, treating symptoms rather than the cause. Our findings instead suggest targeted early interventions such as mentorship, resource support, or legal assistance programs. These efforts can stop small setbacks from snowballing into lasting disadvantage.

1.5. CONCLUSION

Gender-diverse networks benefit both men and women, but current Gender diversity management (GDM) practices are not effective in creating these.
An employee's level of integration can be seen in their affective network through the colleagues they choose to work or socialize with. Our research shows that having a gender-diverse network is generally more beneficial for employee careers than having a network of colleagues with mostly the same gender. The only exception was for women's income gains, particularly in male-dominated teams. Thus, to advance gender equality, we need to look beyond headcounts and ask how men and women actually integrate.

Most workplace diversity efforts focus on counting how many women are on the team or in leadership. But numbers alone don't tell the whole story. We find integration to be crucial for not just employee career outcomes, but also general workplace gender inequality.

Current gender diversity efforts tend to overlook integration as it is difficult to measure, as against traditional gender diversity metrics which simply require counting women. However, without measuring integration we miss key dynamics. For instance, GDM practices like gender quotas can increase representation while unintentionally reinforcing exclusion, triggering backlash, or reinforcing harmful stereotypes. These issues will go unnoticed if its effectiveness is evaluated only by representation.

We find GDM practices to be evaluated as more effective on traditional gender diversity measures. But when integration is factored in, they are found to be far less effective. In other words, companies may be celebrating surface-level wins while missing deeper structural problems.

Early career setbacks hurt women, but not men.
Setbacks early in a career can disadvantage anyone. But while men often recover, women can face growing penalties over time. This is especially true in male-dominated fields like innovation, where men benefit from being part of the dominant group (Ibarra 1992; Ibarra, Carter, and Silva 2010; Kanter 1977c). Over time, this support helps men fully recover from early setbacks, allowing them to reach similar career outcomes as those who had successfully started off. Yet time is a surprisingly underdeveloped dimension in research on gender and careers. Much of the existing literature focuses either on women's early career entry—examining barriers to access—or on their representation in senior management, by which point many have already exited or plateaued. What happens in between remains poorly understood. Yet this middle stage, where career integration, cumulative advantages, and recovery from early setbacks unfold, is critical for understanding enduring gender gaps.

For women, the same setbacks carry higher costs, raising questions on not just their ability, but also their commitment. Employers may assume that women's struggles are due to personal responsibilities, like childcare or family demands (Kelly et al. 2010). This has been noted by researchers as a "gender punishment gap," where setbacks are judged more harshly for women than men (e.g. Rivera and Tilcsik 2016). For example, venture capitalists judged failed female entrepreneurs more negatively than equally unsuccessful male ones (Pistilli et al. 2023).

However, if the same women instead had early success, they can recover from their gender disadvantages. Their future outcomes can match or even exceed those of men. This mirrors research on pay gap reversals, where women are not uniformly disadvantaged. Particularly in male-dominated fields, high-performing women sometimes earn more than men, because only the most capable and determined women remain in these environments (e.g. Huang, Mayer, and Miller 2024; Kumar 2010). But we find that these advantages are not driven by a select group of high-performing women. Even women with identical capabilities can have very different career outcomes depending on whether they had success or setback early in their careers. While early success could make a woman's future career outcomes even exceed those of her male counterparts, a setback could instead lead to an early end for her career or lead to considerably lower outcomes than her peers.

In short, a rough start can close doors for women in ways it does not for men. Yet most gender diversity efforts ignore this early stage. Many current initiatives focus on later-career roles, such as increasing the number of women in management or on company boards (OECD 2022). Such practices mainly help women who were initially successful, further widening the gap between them and those who has initial setbacks. To truly reduce workplace gender inequality, we must look earlier. Interventions that support women after setbacks, such as mentorship, training, or fair performance evaluations, could more effectively improve gender equality.

Widening gender inequality is alarmingly neglected or misunderstood.
Our research highlights a major issue in how growing gender gaps in careers are studied. A key mechanism behind this is cumulative (dis)advantage (CAD), whereby early advantages

or disadvantages, like success or gender, can build up over time to create large career gaps. While we focus on CAD interaction between gender and early career success as an example, our findings apply more broadly to how CAD interactions are typically studied.

The most common method used to study interaction effects, step-wise regression, is highly unreliable. In our tests, it failed to detect true interaction effects in 68% of cases and falsely found independent effects 47% of the time. That is nearly the same as flipping a coin. However, these errors dropped sharply when using a full regression model, where the interaction term is kept in regardless of its significance.

Inequality research is especially prone to such errors, as the effects are often small and the group sizes are highly uneven. This is common in gender research, where one gender often dominates. For instance, men still make up most top leadership roles, and some jobs remain clearly male- or female-dominated, such as construction and nursing. Similar issues arise in migration research, where migrants typically constitute a small share of the population. In both cases, it becomes harder to tell whether outcomes like income or promotion are truly affected by gender or migrant status, or just missed due to skewed group sizes.

Another common mistake is focusing on only a few subgroups, which misses the full picture of how two CAD factors interact over time. For example, when studying how gender and early career success shape outcomes, researchers may compare men and women who succeed early on, but overlook how setbacks affect each gender differently.

To move forward, researchers need to change how CAD interactions are studied. This means avoiding step-wise methods, testing across all relevant groups, using full models, and interpreting results with care. By doing so, we can better uncover the roots of gender and other forms of inequality and support stronger, evidence-based policies.

1.6. LIMITATIONS AND FUTURE RESEARCH

While this dissertation provides new insights into structural mechanisms underlying workplace gender inequality, it also has limitations that future research could address. One limitation is that we only focused on one type of employee network—affective cooperative networks, which include colleagues someone likes to work with or socialize with. While this gives a good sense of both formal and informal ties, there are other types of networks, like advice networks or friendship networks, that may also affect career outcomes differently. Unfortunately, we did not have the data to study these separately. Future research could compare how gender diversity across different types of networks, such as instrumental versus affective, shape access to resources and affect career success.

Another limitation is related to how we studied GDM practices. To make comparisons across many countries and industries, we grouped GDM practices into three broad approaches: helping individuals feel less isolated, reducing bias among managers, and making organizations more accountable. While this helped ensure consistency, it meant we could not account for specific variations in the design and implementation of each practice. Future research could evaluate more narrowly defined practices within these categories or study their effects independently. For example, researchers might isolate the effects of mentoring programs or diversity training within a single organization, and assess their separate impacts on both representation and integration.

While our findings reveal strong associations between network gender diversity and career outcomes, they are based on cross-sectional data and cannot confirm causal relationships. Particularly for individual-level effects, network structure and career success likely influence each other over time. For instance, consider a successful employee with a large network of colleagues at a given time point. Without information on the coevolution of their success and network we cannot know if they used their large network to get successful, or if their success subsequently attracted more connections to form a large network. To answer that, we would need data collected over time. Future research could use methods like Stochastic Actor-Oriented Models (SAOM) to better understand how men's and women's networks and career outcomes evolve together.

Finally, while this dissertation focuses on structural gender inequality, it does not explicitly adopt an intersectional lens. Specifically, we do not consider how overlapping identities—such as race, age, and ethnicity—interact with gender to shape employee experiences. Future research could explore how these intersecting identities compound or alleviate inequality in the workplace. For instance, although women are often penalized for assertiveness in negotiations (Bowles, Babcock, and Lai 2007), this does not hold true for all women. Studies have found that Black women, due to different stereotypes, may actually do better in certain negotiations than White women or Black men (Leigh and Desai 2023). Another example is how white women, but not women of color, experience something similar to a "glass escalator" where they are promoted into management in male dominated fields (Alegria 2019). Thus, an intersectional approach could offer a deeper understanding of how multiple dimensions of inequality interact—and how to design more just and inclusive workplace policies. While we do not adopt a traditional intersectional lens, this dissertation takes an interactional approach by examining how CAD factors like gender and early career outcomes combine to shape long-term inequality. This framework can be extended in other research to include other intersecting identities.

1.7. Advancing SCOOP's Vision: Inclusion and Early Intervention for Sustainable Workplace Cooperation

This dissertation contributes directly to SCOOP's mission of fostering sustainable cooperation, especially within the domain of work. It provides new insights into how gender inequality persists in the workplace and how cooperation can be better supported through improved diversity practices.

A central insight is that true cooperation cannot be sustained by simply increasing the number of women in organizations. For cooperation to last, women must also be integrated into the social and professional fabric of the workplace. This dissertation shows that focusing only on representation overlooks this critical aspect.

It introduces a new way to measure cooperation using employee integration, the colleagues an employee prefers to work or socialize with. Integration can become harder when one gender dominates these circles, as it can reinforce stereotypes and exclusion. For example, women seen as "diversity hires" may find their achievements undervalued and their mistakes judged more harshly. By studying how integration affects outcomes like job satisfaction and income, the dissertation shows how to measure and track sustainable cooperation at work. This tool can also be used to assess the effects of broader workplace changes on cooperation, such as the rise of remote work or temporary contracts, especially after COVID-19. It also has potential for studying cooperation across other social divisions, including race and migrant status.

The dissertation also explores how early career experiences shape long-term outcomes through cumulative (dis)advantage. This means that small advantages or setbacks at the start of a career can snowball into larger gaps over time. This idea is particularly important to SCOOP's goal of understanding how societies can remain resilient in the face of negative cycles—such as repeated patterns of exclusion, stalled career mobility, or persistent gender and class inequalities—and external shocks, including economic recessions, or shifts in labor supply. Examining how individuals and institutions respond to these cycles and shocks provides insight into the mechanisms that either entrench disadvantage or enable recovery and resilience.

Importantly, the research shows that cumulative advantage does not always increase inequality. In some cases, it can help reduce it. For example, both women who succeed early and men who struggle at first can catch up to the most advantaged group over time. But for women who start with a setback, the long-term consequences are often far more severe. These findings suggest that policies aimed at supporting women only later in their careers may come too late. Instead, early interventions can prevent setbacks from becoming permanent disadvantages.

Methodologically, the dissertation aligns closely with SCOOP's emphasis on interdisciplinary, multi-method research. It leverages social network analysis to directly measure integration, furthering how gender diversity is evaluated. It also applies Bayesian hypothesis evaluation to address conflicting findings in the literature and study complex hypotheses more reliably.

Simulation studies are used to test how cumulative advantage interactions are typically analyzed. Recognizing the methodological challenges and fragmentation in this research across disciplines, the dissertation provides structured guidelines for theorizing, analyzing, and interpreting these interactions. It identifies pitfalls, showing that common methods such as step-wise regression often fail to detect real effects or identify effects that do not exist. The dissertation recommends full regression models and provides practical guidance to avoid these errors. These contributions enhance the reliability and validity of evidence-based policy, a key goal of SCOOP.

Although the primary focus is on gender and early career outcomes, the findings help illustrate a broader form of intersectionality. The research shows how different factors interact to shape outcomes over time, offering insights that also apply to other dimensions like race or age.

In sum, this dissertation supports SCOOP's broader mission by providing a clearer understanding of how workplace cooperation can be sustained. It offers new tools to track integration, shows how early outcomes shape inequality, and presents better methods for studying these processes. These insights offer practical guidance for building more inclusive, resilient workplaces and societies.

1.8. POLICY REFLECTION

We know that gender inequality is structurally reinforced. While we do not know if GDM policies can effectively reduce it, their very presence is known to have negative unintended consequences. This is evident from both the widespread backlash against GDM and their use by organizations to justify discriminatory outcomes or reduce litigation risks (Brady et al. 2015).

The best way to eliminate structural inequality would be to eliminate stereotypes, but that could take years. For instance, women are now stereotyped as more competent than they were 60-70 years ago, and belief in gender equality in competence has grown over time. But this change took decades and represents only one gender dimension. Women are still stereotyped as inherently more warm and compassionate and men as more competitive and ambitious (Eagly et al. 2020). So, if we cannot change people's gender stereotypes, what can we do?

Reframing GDM for Broader Support

One way to broaden support for GDM is by rethinking what it means to be a successful or deserving employee—and what it means for an organization to be meritocratic. Institutions often distance themselves from policies like gender quotas out of concern that such measures conflict with the idea of selecting on merit. But this assumes we are already operating in a true meritocracy. In reality, many people benefit from structural advantages related to gender, race, class, or family background, while others face systemic barriers that limit their opportunities.

For example, 72% of Fortune 500 senior executives are white men (Jones 2017). Moreover, among 587 male CEOs in the U.S., over 71% had fathers in white-collar jobs, about half attended private colleges, and nearly 20% came from families in the top 1% of the income distribution (Duchin, Simutin, and Sosyura 2021). This extreme concentration of privilege makes claims of a purely merit-based system unconvincing. Rather than viewing GDM as undermining meritocracy, it can instead be framed as correcting an existing imbalance.

Making the case for diversity without backlash.
Our research shows that gender-diverse networks benefit both men and women, providing a compelling reason to support diversity at the individual level. Emphasizing this shared benefit can reduce resistance to gender diversity initiatives and aligns with recommendations to reframe diversity initiatives as opportunities rather than threats (Dover et al. 2019).

This individual-level framing also sidesteps problems with the traditional business-case for diversity, which argues that diversity is good for organizational performance. While this argument can open doors for initial conversations about equity, it paradoxically undermines belonging for minority groups and places undue responsibility on them to enhance organizational outcomes based on their identity. For example, in publishing, it burdens minorities with the responsibility of reaching their identity group as an audience, even if they don't necessarily have the tools to do that (Saha and and van Lente 2022). To avoid these pitfalls, two alternative framings are proposed:

- The moral- or fairness-case: framing diversity simply as the right thing to do. Although not perfect, this approach substantially reduces harmful consequences compared to the business case (Georgeac and Rattan 2023).
- The “no-case” or transformational-case: encouraging organizations to internalize equity as a core value needing no explicit justification—just like innovation (Williams 2022).

While the no-case is argued to be a superior approach as it internalizes the case for diversity, considerable time and effort is be needed to get there. Moreover, both the moral-case and the no-case framings could make leaders, especially from majority groups, feel blamed or coerced, which can lead to defensiveness and disengagement. To reduce this risk, leaders should view themselves as active agents in shaping DEI (Diversity, Equity, and Inclusion) efforts, not just reacting to demands. Encouraging open dialogue and collaborative problem-solving can shift the focus from obligation to shared responsibility.

Thus, both the moral-case and the no-case for diversity can be most effective when paired with leader autonomy. This holds for all diversity advocates, whether employees, activists, or regulators. Rather than pushing one solution, offering a menu of evidence-based actions allows leaders to choose while remaining accountable. This approach encourages buy-in, reduces resistance, and supports more sustainable progress on DEI goals (Preston 2025).

Revisiting leadership traits and promotion processes benefits everyone, not just women.

Redefining what success looks like in leadership can benefit everyone. Many professional environments still value traits like assertiveness and independence—traits stereotypically linked to men—over those like empathy and collaboration, which are more often associated with women. This mismatch can lead to women being undervalued, especially in leadership roles.

One solution is prototype balancing, which expands the definition of leadership to value both sets of traits. Rather than replacing existing standards, this approach highlights the strengths underrepresented groups bring while still recognizing traditional qualities. Danbold & Bendersky (2020) demonstrate how this approach can help reduce bias without triggering backlash even in highly male-dominated fields like firefighting.

Another effective measure is redesigning promotion processes. Individual-level interventions like unconscious bias training or "lean-in" strategies have limited impact. More effective are structural changes to choice architecture. However, interventions, altering the choice architecture around the decision to apply for top positions are found to be much more effective. For instance, many organizations use opt-in systems where employees must self-nominate for leadership opportunities. Research shows that shifting to an opt-out model, where all qualified individuals are automatically considered unless they decline, significantly increases the chances that underrepresented groups are included in the talent pipeline. These changes can help organizations surface talent more fairly and build leadership that better reflects the diversity of their workforce (He, Kang, and Lacetera 2021).

Tailoring Interventions for Effective GDM

Prioritize early-career interventions over later-stage diversity targets.
Current diversity initiatives, such as board quotas (OECD 2022), often target advanced career stages when early disparities have already widened. Such late interventions primarily benefit women who have already overcome structural barriers, inadvertently creating a "female premium" at the top and exacerbating overall inequalities (Leslie, Manchester, and Dahm 2017).

Policies that target women's representation at advanced career stages mainly benefit those who are already successful. In the case of inventors, such policies would reinforce advantages for initially successful women, widening gaps with unsuccessful women.

Instead, our findings highlight the value of early-career interventions. Supporting women at the point of initial failure—rather than later in their careers—can help prevent inequality from taking root. In the case of inventors, this could mean improving the fairness of initial evaluations rather than focusing solely on hiring more women. Examples include fairer appraisal procedures or tax incentives for underrepresented inventor groups.

This case is strengthened by results from a randomized control trial conducted by the USPTO, where some patent applicants without legal representation were randomly selected to receive early assistance. Because the selection was random, the outcomes can be directly linked to the intervention, making the findings more reliable. While both men and women benefited from the support, the likelihood of women securing a patent increased by over 12 percentage points more than it did for men. This is likely because women are more often independent applicants or part of small businesses and non-profits, which typically lack legal and financial resources (Pairolero et al. 2025). Since such early, gender-neutral, merit-based interventions help a broader group of under-resourced applicants, not just women, they may also be more politically viable in today's climate of growing resistance to diversity initiatives.

Centering integration in measuring gender diversity and evaluating GDM.
Despite growing investment in diversity programs, many organizations have seen little improvement in workplace experiences for women and racialized employees (Trzebiatowski et al. 2025). A key reason is gender diversity measures focusing on representation and the subsequent limited GDM evaluations looking at changes in women's representation. However, increasing representation alone does not necessarily reduce inequality (Bellotti et al. 2022). GDM policies that focus solely on representation can have mixed outcomes, especially if they ignore how individuals connect and collaborate at work (Sabharwal 2014). For example, research shows that gender diversity without inclusion fails to enhance team creativity (Vedres and Vásárhelyi 2023). Even when GDM improves representation, it may be ineffective unless it also promotes integration. In fact, improving representation is unlikely to be sustainable without integration.

To address this, we propose that policymakers and organizations adopt a network-based gender diversity measure that captures both representation and integration. This approach offers a more complete picture by accounting for how men and women are connected in the workplace, not just how many of each are present. It also avoids the blind spots of representation-only measures by revealing how diversity actually functions within employee networks (Reagans et al. 2004; Van Knippenberg and Schippers 2007). Such a measure can offer a practical and cost-effective tool to evaluate GDM. it can be built from anonymized digital communication data such as email metadata or internal messaging platforms, aligning with data protection standards and requiring only modest resources to implement.

Network-based measures of gender diversity can also offer a useful lens for improving employee networking strategies and related programs (Forret and Dougherty 2004). For instance, organizations often promote homophilous ties in their networking and mentoring programs (Dobbin, Kalev, and Kelly 2007; Pini, Brown, and Ryan 2004). However our finding show that both men's and women's careers benefit from gender-diverse networks, and women's income is particularly benefited by heterophilous networks. Thus accounting for integration allows us to extend gender diversity research and policy to the oft overlooked micro level, which can better enable policy makers to address backlash.

The need to look at the micro level is also reflected in gender inequality being systematically reinforced even in the presence of fairer reward systems. This is because fairer reward systems are typically recommended to address systemic inequalities(e.g. Bol et al. 2018; Rivera and Tilcsik 2019). However, as we saw with inventors, gender inequality persist despite the fair patent evaluation system. Indicating the critical role of micro processes in systematically reinforcing gender inequality, say through employees' psychosocial conditions or managerial sponsorship (Biegert, Kühhirt, and Van Lancker 2023; Bol et al. 2018).

Institutions could also track gender differences in turnover across organizational levels, as high turnover among minority hires signals poor integration (McKay and Avery 2005). Aggressively increasing representation without improving integration may result in high turnover, especially among women. While traditional gender metrics focus on outcomes, like income or promotions, they overlook those who exit the system. Thus turnover rates offer a powerful complement to traditional metrics. For example, women in senior roles leave at nearly twice the rate of men (Krishnan 2009), not just undermining gender equality but also incurring costly talent losses and reducing organizational performance (Krishnan and Park 2005; Smith, Smith and Verner 2006).

In sum, tracking representation is not enough. To design and evaluate effective GDM practices, organizations must also assess whether women are integrated and retained. Including network-based gender diversity measures and turnover data can effectively improve GDM.

CHAPTER 2

Should Birds of a Feather Flock Together?

How Networking with Men and Women Relates to Employees' Career Success

Abstract

Despite extensive research on workplace gender diversity, its impact at the individual level has often been overlooked, as diversity effects are typically measured at group or team level. We study the individual level by shifting the focus from representation to integration, examining the gender composition of employee networks based on the colleagues they prefer to work with or socialize with. These colleagues are often of the same gender, making the network homophilous, but studies have shown mixed results on whether homophily helps or hinders career success. We resolve this conflicting finding by studying all three network types—homophilous, diverse, and heterophilous—and evaluating how each relates to job outlook and income for men and women. Applying Bayesian model selection to rich network data from six sectors and nine European countries, we find gender-diverse networks to be generally the most beneficial. The exception is women's income, which benefits more from male-dominated networks. However, when women are not in the minority, there was no gender difference in the benefit of all three network types. These findings offer important novel insights for diversity policies and employee networking strategies.

The authors jointly developed the core ideas of this chapter. **Sanjana Singh** wrote the main part of the manuscript and conducted the analyses. **Eva Jaspers, Tanja Van der Lippe** and **Belle Derks** substantially contributed to the manuscript. An earlier version of this chapter has been presented at *Dag van de Sociologie (2019)*, and *European Conference on Social Networks (2019)*. The chapter also benefitted from invaluable analytical advice provided by **Herbert Hoijtink**.

2.1. INTRODUCTION

Despite decades of progress in women's representation at work, gender inequality continues to shape how careers develop. Even in fields where women are the majority, they are often disadvantaged. For example, women make up 78.4% of U.S. elementary and middle school teachers but earn only 89.1% of what men earn (IWPR 2024), showing that representation alone does not guarantee fairness. This chapter explores how a key missing factor—employee integration—helps explain such persistent inequalities.

Integration refers to the informal relationships that employees build at work. These include who they choose to collaborate with, seek support from, or socialize with. Such ties shape access to opportunities, resources, and advancement. Since workplaces are inherently social, these informal bonds are essential for both men's and women's careers (Cullen and Perez-Truglia 2019).

Consider the example of golf, a traditionally male-dominated sport that remains a key networking tool in large corporations. Women who play golf are 116% more likely to serve on corporate boards because golfing gives them access to informal male social circles from which they are otherwise often excluded. This increases their visibility and credibility among senior leaders, helping their careers (Agarwal et al. 2016). A similar pattern is seen with smoking breaks. In one firm, men who regularly joined male managers for smoke breaks were promoted more often, with this informal socializing explaining nearly a third of the promotion gap (Cullen and Perez-Truglia 2019). These examples show how informal interactions shape careers, making integration essential to addressing workplace inequality.

In addition to individual careers, integration is also important for group outcomes. It underpins the business case for gender diversity. When employees from different backgrounds build strong, inclusive relationships, diversity can lead to fresh ideas, better problem-solving, innovation, stronger performance, and financial gains (e.g. Bantel and Jackson 1989; Herring 2009; Herring and Henderson 2015; Van Knippenberg, De Dreu, and Homan 2004). But when these connections are weak or absent, diversity may cause divisions that hurt team cohesion, decision-making, and performance (e.g. Garrison et al. 2010; Rogelberg and Rumery 1996; Rothman, Lipset, and Nevitte 2003; Tsui and O'Reilly 1989; Williams and O'Reilly III 1998). Thus, strong integration enables diverse groups to thrive, while poor integration can limit the potential of diversity and deepen inequality (Reagans and Zuckerman 2001). Yet studies on the effects of gender diversity often overlook this mechanism, rarely considering individual-level dynamics, leaving integration and its effect on individual outcomes largely unknown.

A key reason this gap persists is that integration is rarely measured directly. Gender diversity is typically measured using simple headcounts or gender ratios, which show who is present but not how they connect. These measures reflect only the potential for integration, not actual relations. Some studies use managerial roles as a proxy, assuming women in leadership signals integration (Stainback and Tomaskovic-Devey 2012). But this is often misleading. For example, in Korea's informal networks known as yongo, women in senior roles are still excluded from key conversations. Rooted in military and university ties among men, these networks often sideline women. In some cases, customs like expecting women or younger staff to pour drinks for senior male managers reinforce women's lower status (Horak and Suseno 2023). As a result, most studies measure gender diversity using group-level ratios and focus on group outcomes. This overlooks how well individuals are integrated and how that affects their careers.

Social network research helps fill this gap. It can directly measure integration through each employee's affective cooperative networks, or the colleagues an employee prefers to work and socialize with. These ties reflect actual relationships that shape access to opportunities. However, past research has produced mixed findings on how the gender makeup of these networks influences career success (Woehler et al., 2020).

Much of the literature focuses on the effects of having homophilous, or same-gender ties, and heterophilous, or opposite-gender ties (Woehler et al. 2020). But this overlooks gender-diverse networks, where employees have a balanced mix of male and female connections. This matters because the effects of diverse networks are both theoretically and empirically distinct from those with mostly same- or other-gender ties. Theories of knowledge and exchange suggest that gender-diverse networks can support career success by improving information flow, encouraging collaboration, and creating a more stimulating work environment (Ertug et al. 2018; Hansen 1999; Reagans and McEvily 2003).

This chapter introduces a new approach to measuring integration and examining its effects on employee careers. Rather than focusing only on homophily or heterophily, we categorize networks into three types: homophilous, diverse, and heterophilous. This allows us to assess which network type most benefits men's and women's career outcomes. In doing so, we contribute to two distinct areas of research. First, we help resolve the mixed findings in the network literature by examining the unique effects of gender-diverse networks. Second, we shift the focus of gender diversity research from team- and organizational-level outcomes to the individual level. By using job outlook and income as indicators of subjective and objective career success, we offer a more complete picture of how integration shapes careers.

To understand how integration influences career success, we examine two key factors: the employee's gender and the team's gender composition. Research shows that men and women not only form networks differently but also benefit from them in different ways (Son and Lin 2012; Williams and O'Reilly III 1998; Wolff and Moser 2009). In addition, Kanter's tokenism theory suggests that team gender balance shapes both who employees can interact with and how those interactions unfold (Kanter 1977c; Simmel 1950). For example, in a team with two women and eight men, women may have fewer networking opportunities and may be more likely to be seen as outsiders. As the majority, men are more likely to define the team's norms and identity, reinforcing gender inequalities (Kanter 1977b). In such teams, forming ties with men can play a crucial role in advancing women's careers.

We study this using a unique, extensive multilevel dataset from the European Sustainable Workforce Survey (2017–2018), covering over 4,000 employees across multiple countries and sectors. This directly addresses a major challenge in integration research: the lack of representative network data that reflects real organizational structures across diverse settings. We also tackle a second key challenge: how to reliably compare and evaluate the benefits of different networks for men and women. Traditional statistical methods often struggle when testing multiple competing hypotheses, increasing the risk of error. To overcome this, we use Bayesian hypothesis evaluation, which enables clearer and more reliable comparisons.

Summarizing, this chapter deepens our understanding of integration by examining how the gender composition of employees' networks relates to their career success. Our study makes two key contributions. First, it is among the first to assess gender diversity of relationships at the individual level and the effects thereof, filling a major gap in diversity research that has primarily focused on teams and organizations. Second, it clarifies mixed findings on the effects of both diversity and network composition by distinguishing the outcomes associated with homophilous, heterophilous, and gender-diverse networks. These insights offer practical value. For employees, they show how the composition of workplace networks can shape career success (Forret and Dougherty 2004). For policymakers, they highlight the need to look beyond representation and focus on whether people are truly included in workplace relationships. While there has been an increase in practices like gender quotas to boost representation (Leszczyńska 2018; UN Women 2018), they often overlook how these changes affect integration. Such practices often face resistance, making their real impact on gender inequality unclear (e.g. Dover et al. 2019).

2.2. NETWORKING AND CAREER SUCCESS

In this section we separately theorize which type of network—gender homophilous, diverse or heterophilous—contributes best to both men's and women's career success.

Networking can be described as the process of building up and maintaining a set of informal and cooperative relationships (Michael and Yukl 1993; Orpen 1996), generally involving exchange of affect (liking, friendship), information, benefits, and influence (Gibson, H. Hardy III, and Ronald Buckley 2014; Michael and Yukl 1993). We quantify this using the employees' affective cooperative network. These are the colleagues an employee chooses to meet outside of work or likes to work with. The network gender composition can be quantified to a three-part scale for homophily that ranges from having completely heterophilous ties (all ties in their affective cooperative network is with the other gender) to having completely homophilous ties (all ties in their affective cooperative network is with the same gender). While values in the first and last part indicate heterophily and homophily, respectively, values in the middle indicate high gender diversity in their network. Thus an employee's network could be either heterophilous, diverse or homophilous. By adopting a categorical network measure we are able simultaneously study how each of the three network types relate to career success and develop a more nuanced understanding. For instance, instead of just finding homophily to not be beneficial, this construction can help understand if it would be more beneficial to have a diverse or heterophilous network.

The novelty and necessity of this approach is reflected in Woehler et al.'s (2020) recent extensive review on "Whether, How, and Why Networks Influence Men's and Women's Career Success: Review and Research Agenda", covering 112 articles spanning 1978-2020 in a range of disciplines. They found gender homophily and the proportion of men in an employee's network to be the two most common network characteristics studied, present in more than half the articles reviewed. This reflects how the effects of heterophily on men and gender diversity on all employees have been overlooked. We address this by explicitly accounting for six groups of network gender compositions—men's and women's homophilous, diverse and heterophilous networks—which are detailed in the next section.

For career success, we use two different measures to evaluate our theories—job outlook and income. This follows from the implication of network characteristics having varying effects depending on how career success is measured (Woehler et al. 2020). While job outlook is a forward looking subjective measure based on the employees' opinion on their prospects in the firm, income provides a static objective measure for career success. This parallel use of subjective and objective measures for analyzing career success aligns with Hughes's subjective-objective career duality framework – one of the most widely accepted frameworks in career research (Mayrhofer et al. 2008). Employee careers are inherently considered as a "two-sided" concept that comprises of the subjective and objective career views. In fact the two dimensions are found to be unique, empirically distinct constructs (Arthur, Khapova, and Wilderom 2005) showing different patterns of correlations with commonly used predictor variables (Ng et al. 2005). Furthermore, it is considered essential to include both measures when studying gender-sensitive models of career success (Mayrhofer et al. 2008), as men and women perceive success differently, with women often giving more importance to subjective career success.

Is it Better for Men's and Women's Careers to have Gender Homophilous, Diverse or Heterophilous Networks?

While homophily is identified as one of the most persistent findings in social network analysis (Block and Grund 2014), it need not be the most beneficial network composition for employees' career success. All three types of networks—homophilous, diverse and heterophilous—can be argued to differently benefit both men and women.

The case for homophily benefitting employee careers follows from theories of knowledge and exchange. It suggests that employees work better with similar others. Homophily fosters higher trust and cooperation making it easier for employees to access the resources and support required to do their jobs (Ertug et al. 2018; Hansen 1999; Reagans and McEvily 2003). This is also corroborated in mentoring research, which suggests homophilous ties as necessary channels for accurate information and "psychosocial" benefits such as role modelling (Ibarra 1997; Kram 1988). For instance, Ibarra (1997) found high-potential women to rely more strongly than nonhigh-potential women on homophilous ties. Thus, both men and women with homophilous networks can expect better career success than those with heterophilous or diverse networks.

The case for diversity benefiting employee careers builds on insights from both network and organizational research. The organizational literature has shown mixed findings regarding the relationship between gender diversity and organizational outcomes. For instance, gender diversity on corporate boards has been linked to both positive (Campbell & Mínguez-Vera, 2008) and negative (Adams & Ferreira, 2009) firm performance. However, it is often argued that such mixed findings can be explained by examining gender diversity at the individual level. The network literature provides tools to do this by focusing on employees' affective cooperative networks, which capture patterns of collaboration and emotional support.

Consider a ten-member team with equal numbers of men and women. At the team level, this represents high gender diversity. But if employees are generally homophilous, i.e., choose to cooperate only with same-gender colleagues, their low network gender diversity could negatively affect outcomes by increasing social division and conflict in the team (Garrison, Wakefield, Xu, & Kim, 2010; Tsui, Egan, & O'Reilly III, 1992; Williams & O'Reilly III, 1998). Instead, if employees choose to cooperate with both same- and other-gender colleagues, creating high network gender diversity, this could positively affect outcomes by increasing information flow and collaboration (Bantel & Jackson, 1989; Van Knippenberg, De Dreu, & Homan, 2004).

Thus, the effects of gender diversity at the team level depend on the gender composition of employees' affective cooperative networks. To understand how gender diversity influences employee careers and organizational performance, we must examine not only who is present in the team, but also how they connect.

In this sense, the case for diversity refers specifically to diversity within an employee's affective network. Such networks reduce the likelihood of social division and can foster a more stimulating environment, leading to better job outlook and income. For example, studies on academic collaboration have found that mixed-gender coauthors produce higher-quality articles and publish in more prestigious journals compared to same-gender coauthors (Campbell et al. 2013; Kwiek and Roszka 2021). Therefore, employees with diverse networks are more likely to benefit in their careers.

Finally, the case for heterophily being beneficial follows from expectation states theory, but applies only to women. Gender is deeply entwined with social hierarchy, with gender stereotypes containing status beliefs that associate men with greater status worthiness and competence (Ridgeway 1991, 2001). Although these gender-based status differences arise outside the organization, they add to the structuring of controls within the organization as organizations tend to reflect the cultures they are placed in (Acker and Van Houten 1974; Fairhurst and Snavely 1983). For instance, gender was seen to independently affect an individual's status in a team, even after controlling for their competence and performance indicators (Cohen and Zhou 1991). In fact Ely and Meyerson (2000) show how gender-based status differences affect the organizational structuring of controls to perpetuate these differences.

These beliefs on gender status also shape the ability attributed to men and women, the influence they achieve, and their potential to assert themselves (Ridgeway 2001). As a result, men are not only ascribed a socially higher valued gender, but also as a higher valued network tie (McGuire 2002) for an individual's career advancement. Hence, employees with more men in their network are likely to have better career success. This aligns with the case for homophily for men. However, for women this means that heterophily, as compared to homophily or diversity, would lead to better careers. For instance, when women use male contacts in their search for employment, they are seen to obtain more success and higher income (Aten et al. 2017; O'Connor 2013). In fact professional e-networks have found women to benefit from heterophily and men to benefit from homophily (Aten et al. 2017).

This is also corroborated by the corollary, whereby women are ascribed to be the socially lower valued gender. The expectation states theory also underlines the relative legitimacy challenges faced by women to assert themselves. It argues that “the enforcement of behavioral expectations created by the status elements of stereotypes creates a legitimacy process that affects women leaders to exercise directive power and achieve compliance” (Ridgeway 2001). This could in turn make it harder for women to promote their connections within the organization, making them a lower valued connection. Thus, women with heterophilous networks can expect better career success than those with homophilous or diverse networks.

Considering the above cases for homophily, diversity and heterophily together, we have two mutually exclusive hypotheses for men and three mutually exclusive hypotheses for women. We evaluate these competing hypotheses using Bayes factor to find which network gender composition is the best at improving both men’s and women’s career success. These competing hypotheses are evaluated on the two complementary measures of career success – job outlook and income.

Which network gender composition is the best at improving men’s career success?

Competing Hypothesis 1.M1: Network homophily has the strongest association with men’s career success, as compared to network gender diversity and heterophily

Competing Hypothesis 1.M2: Network gender diversity has the strongest association with men’s career success, as compared to network homophily and heterophily

Which network gender composition is the best at improving women’s career success?

Competing Hypothesis 1.W1: Network homophily has the strongest association with women’s career success, as compared to network gender diversity and heterophily

Competing Hypothesis 1.W2: Network gender diversity has the strongest association with women’s career success, as compared to network homophily and heterophily

Competing Hypothesis 1.W3: Network heterophily has the strongest association with women’s career success, as compared to network gender diversity and homophily

How does the Team’s Gender Composition Affect Men’s and Women’s Network Benefits?

The above competing hypotheses show how men and women mostly share similar expectations on their network benefits. The only exception is women’s differential hypothesis on heterophily, competing hypothesis 1.W3. This hypothesis was driven by the expectation states theory, where men are ascribed to be the socially higher valued gender and connection. Thus contexts where men have an advantage can increase gender differences.

Such contexts can be quantified using the team’s gender composition, that is whether it is a token or non-token team. Token teams—where women occupy minority or token position and men form the majority or dominant group—provide an environment where gender differences are increased. Conversely, non-token teams—where women comprise a greater proportion of the team and do not just hold token positions—provide an environment where gender differences are lowered. In this section we theorize on the gender difference in network benefits across token and non-token teams, through the second and third set of hypotheses, respectively.

Tokenism theory details how teams generate certain perceptions of the tokens by the dominants, which determines interaction dynamics between them. These perceptions include polarization and assimilation. The former refers to the dominants exaggeration of their commonality and the token's difference, which could result in the token's isolation. For instance, Kanter (1977c) observed this when an all-male sales team introduced token women. Their presence led to the dominant men highlighting what they could do, as men, in contrast to women, e.g. recounting stories in which masculine prowess accounted for personal, sexual, or business success. The latter, assimilation, refers to the distortion of token characteristics to fit pre-existing generalizations or stereotypes about their category, often leading to the token's role entrapment. For instance, in the above sales team despite knowing that the token saleswomen were not secretaries, they were still assigned secretarial tasks. Both these perceptions add to the legitimacy deficit faced by women and heighten gender differentials (Kanter 1977c). Thus, in token teams, both men and women would find it more beneficial to have ties with men. This means that, women would find heterophily to be more beneficial than homophily, while men would find homophily to be more beneficial than heterophily.

> *Hypothesis 2.M: In token teams, network homophily has a stronger association with men's career success, as compared to network heterophily.*

> *Hypothesis 2.W: In token teams, network heterophily has a stronger association with women's career success, as compared to network homophily.*

For non-token teams, women are likely to have more legitimacy and depend less on male colleagues. This should reduce the gender differences seen in network benefits, allowing women to network in the same way as men to improve their career success (Schoen, Rost, and Seidl 2018).

> *Hypothesis 3.1: In non-token teams, men and women have a similar association between their homophilous network and career success.*

> *Hypothesis 3.2: In non-token teams, men and women have a similar association between their diverse network and career success.*

> *Hypothesis 3.3: In non-token teams, men and women have a similar association between their heterophilous network and career success.*

Again, as the implication of network characteristics is seen to vary based on the measure for career success, we evaluate both Hypothesis 2 and Hypothesis 3 on two measures – job outlook and income.

Although Kanter (1977a) originally described the tokenism theory as gender-neutral, later studies found that while women are typically affected negatively by token positions, men are not (Schoen et al. 2018). Thus, we only look at teams where women occupy token positions.

2.3. METHOD

Data Source and Sample

We evaluate our hypotheses using data from the second wave of the European Sustainable Workforce Survey conducted in 2017-2018 (Van der Lippe et al. 2018). This data was collected from 113 organizations across nine European countries: Bulgaria, Finland, Germany, Hungary, the Netherlands, Portugal, Spain, Sweden and the United Kingdom, and six sectors: Manufacturing, Transport, Financial Services, Health Care, Higher Education and Telecommunication. We used the second wave as it included information on employees' affective cooperative networks, which is essential for studying gender diversity at the individual level. Organizations were invited based on their participation in the first wave in 2015-2016, where they were approached using stratified random sampling based on their sector (financial services; health care; higher education; manufacturing; telecommunications; transportation) and organization size (<100 employees; 100-249 employees; 250+ employees). This approach was complemented by a matching strategy to replace non-participating organizations by organizations from the same strata; hence, it is not possible to calculate between-organization response rates. The organizations that agreed to participate in a second wave (101 out of 259, 39%) were complemented by 12 new organizations that were selected by matching strata.

When an organization agreed to participate, three groups were asked to complete questionnaires: HR managers, team managers, and employees. These surveys gathered information on organizational, team, and individual-level aspects, and were administered in each country's native language. Employees were selected from three to six teams or departments, as defined by the organization.

To encourage broad participation, organizations received a customized benchmark report, which motivated them to ensure strong representation across selected teams. As a result, missing network data is unlikely to reflect patterns of nonresponse, since all employees were nudged to participate regardless of role or level. Additional incentives were also provided. In several countries—such as the United Kingdom, Germany, Finland, Sweden, the Netherlands, Portugal, and Spain—employee participation supported charitable donations. In Hungary, employees were entered into a lottery, and in Bulgaria, they received direct cash transfers. The response rate for HR managers was 89%, for team managers it was 54% and for employees it was 68%. At the organizational level, it is not possible to estimate the response rate because their purposeful sampling is partially non-random (van der Lippe and Lippényi 2019). This resulted in information on 4345 employees from 401 teams in 113 organizations. We make use of the entire sample by addressing the missing observations with multiple imputations. This procedure has been detailed later in the section.

Outcome Variables: Career Success

Job Outlook: Information on each individual's job outlook was collected using the individual level questionnaire, by asking employees to indicate the extent to which they agreed or disagreed with the statement "I am likely to get a better job in this organization in the next three years". The employees could respond on a five point scale that ranged from 'strongly agree' to 'strongly disagree'. We rescaled the responses such that a value of one referred to 'strongly disagree' and a value of five referred to 'strongly agree'. Thus a higher value would indicate better job outlook.

Income: Information on employees' income was collected using individually reported earnings in the local currency. This was standardized to reflect Purchasing power parity (PPP) in terms of the USD, using the World Bank's International Comparison Program (ICP) PPP conversion factors for the reference year 2017. As this value was found to be heavily skewed, we take its log for the analysis.

The need to simultaneously consider subjective and objective measures of career success is reflected in the low and statistically insignificant correlation between job outlook and income, indicating that they capture different aspects of employee careers. This can be seen in Table A2, in Appendix A, along with all supplementary materials for this study.

Independent Variables

Network Gender Composition: As discussed earlier, we develop a three part network measure for each individual that reflects whether they have a heterophilous, diverse or homophilous network. In following with the team and organizational level gender diversity studies that look at the gender ratio in a group, we continue to look at the gender ratio in an individual's affective cooperative network. This was based on employee responses to the following two questions:

"Whom in your department do you also see outside work?"

"Whom do you like to work with in your department?"

Employees could nominate at most three colleagues for each question. We combined the responses to both these questions to constitute the network of colleagues (alters) an individual (ego) cooperated with. We then identified the gender of all alters using a three step strategy. First, in cases where the surveyed organization provided a name list, it was possible to identify alters within the data and trace their responses to the survey. Specifically, the alter's response to the gender question in the survey was used to identify their gender. Second, in cases where organizations did not provide a name list, a fuzzy match algorithm was developed in R 3.8.0 to identify an alter's gender by matching the name entered by the ego to a roster of names and their corresponding gender by country in the World Gender-Name Dictionary –a roster of 6.2 million names from 182 different countries (Lax Martínez, Raffo, and Saito 2016). Finally, the third step was taken for cases where the second step did not result in a strong match. These were manually checked to identify gender through native speakers and through searches on social and professional networking sites. The network gender composition was then estimated only for those egos where the gender of all alters in their network was identified across both questions. Specifically, we combined the responses to both these questions to constitute the network of colleagues an individual (ego) cooperated with. This follows from findings that affective cooperative networks of employees span formal and informal channels (Gupta and Govindarajan 2000), with both channels influencing employee careers (Cullen and Perez-Truglia 2019). Additionally, as responses to the above questions saw high sparsity, combining their responses also helped improve the robustness of our analysis.

To construct the three part network measure, we first calculated the proportion of both genders, men and women, in an ego's network. However, as the measure can be seen as a scale for homophily, it additionally depends on the ego's gender. In fact we either use the proportion of men or the proportion of women, depending on whether the ego was a man or a woman, respectively. For example, if an ego had a network of six alters of whom one was a woman, then approximately 17% of their network would be women and 83% would be men. If the ego was a woman, the network measure would be based on the 17% and if they were a man, this would be based on the 83%. The next step was to categorize the selected proportion into three equal parts. The first part, ranging from 0% to 33%, would be considered heterophilous networks as these predominantly consist of members of the other gender. The second part, ranging from 34% to 66%, would be considered diverse as these networks have a mix of both men and women. Finally, the third part, ranging from 67% to 100%, would be considered homophilous networks as these predominantly consist of members of the same gender. Thus in our earlier example, if the ego is a woman, using the 17% we see that she has a heterophilous network. However if the ego was a man, using the 83% we see that he has a homophilous network.

Employee's Gender: As outlined above, each employee or ego's own gender is used to calculate their network gender composition measure. Additionally, we focus on the critical role of employee's gender, in the relationship between their network and career success. This is done by separately theorizing and analyzing the association between men's and women's network gender composition and their career success.

Team's Gender Composition: We use the team's gender composition to split the data into token and non-token teams. When the proportion of women in the team is less than 20%, i.e. women occupy a token position, we categorize it as a Token team. This aligns with Kanter's (1977c) structure of Token teams and is widely used across disciplines. Conversely, non-token teams are those where the proportion of women in the team is 20% or more. We use the data subsets with token and non-token teams to study the second and third hypotheses, respectively.

Control Variables

Given the multilevel setup of organizations, controls were used at the individual, team and organizational level. At the individual level we controlled for the employees' human capital or their educational, personal, and professional experience (Becker 1964) which can enhance their career success (Ng et al. 2005; Whitely, Dougherty, and Dreher 1991). This included indicators on employees' *educational years, tenure* (months working in the organization) and job *autonomy* (ranging from 1-5; 1=low autonomy, and 5=high autonomy). Educational years was estimated using the individual educational levels that were recorded as per the International Standard Classification of Education (ISCED) categories and recoded based on the PISA 2012 technical report to reflect the official number of years an individual spent receiving education (OECD 2014). The questions for autonomy was adapted from different organizational surveys (e.g. Karasek Jr (1979); University of Essex et al. (2019)) and can be found in Appendix A. The scale was found to be reliable with a Cronbach's alpha of 0.89 and was recoded such that a higher value on the scale reflected higher autonomy.

At the team level we controlled for the *proportion of female employees* in a team, as it provides the opportunity structure for an employee's network. This is accounted using the scaled question on the proportion of female employees in a team (1=None, 2=1% to less than 10%, 3=10% to less than 20%, 4=20% to less than 40%, 5=40% to less than 60%, 6=60% to less than 80%, 7=80% to less than 90%, 8=90% to less than 100%, 9=All).

Finally, we include the organizational level, as it can influence outcomes such as performance, turnover, salaries (Judge et al. 1995; Melamed 1996; Pfeffer 1991) and thus an individual's career success. Specifically, we included the *number of employees,* as larger organizations have greater vacancies and promotion opportunities which could alter employees' career prospects (Whitely et al. 1991). As this value was found to be skewed, we take its log when including it in the analysis. Additionally, we controlled for structural attributes that could generally influence employees' career success. For instance, as employees exist within a broader labor market, differences in their job outlook could also be driven by geographic differences in pay and career patterns (Judge et al. 1995). Thus we include the *GDP per capita* of the country in which an organization is located*.* To ensure comparability we used the purchasing power parity (PPP) values in constant 2017 international dollars, as provided by the World Bank's International Comparison Program (ICP).

Table 2.1 outlines the summary statistics at the individual, team and organizational level, using all available information per variable. The natural inclination for homophily can be seen here with about 67% of employees in our data having homophilous networks.

As discussed earlier, the network measure was constructed by using information on the proportion of women and men in an employees' affective cooperative network. This meant that the measure could only be computed for an employee when the gender of all alters in their network was identified. As the network would be incomplete if the gender of even one alter was unidentified or missing, almost 84% of the employees in our data had incomplete network information. Consequently, the network measure saw a high proportion of missing information. To address this we performed multiple imputation using Fully Conditional Specification through the MICE algorithm in R 3.8.0 (van Buuren et al. 2015; van Buuren and Groothuis-Oudshoorn 2010) with 100 imputations and 20 iterations. The active dataset used for the imputation was selected in line with the predictor variable selection strategy outlined by van Buuren (2018). This included all available information on the above variables of interest along with other variables that could explain a considerable amount of variance or related to nonresponse, such as whether they were surveyed online or using paper-and-pencil. Despite including the network measure, which saw a high missing proportion, only about 14% of the cases in the active dataset were missing. This implies that our use of 100 imputations not only complies with the suggested de-facto standards but also ensures that the essential quantities in our analysis should be reproducible within some limit (Von Hippel 2009; White, Royston, and Wood 2011). Given the multilevel nature of the dataset we modeled the team and organizational clustering of the data as a fixed effect during imputation and applied the predictive mean matching (PMM) method. This method has been noted to have comparable performance to dedicated methods for multilevel data (Vink, Lazendic, and van Buuren 2015) while simultaneously facilitating imputation for the whole dataset which is often challenging using dedicated methods for multilevel data.

Table 2.1 Aggregated Descriptive Statistics

	Mean/ Prop.	SD	Min	Max	Prop. Missing
Key Variables					
Job Outlook	2.55	1.1	1	5	0.10
Earnings in USD (PPP)	2197.28	3144.23	1.15	135387.83	0.16
Heterophilous Network	0.14				0.84
Diverse Network	0.19				0.84
Homophilous Network	0.67				0.84
Individual Controls					
Gender: Female	0.59				0.05
Tenure	142.28	127.81	1	720	0.04
Educational Years	13.54	3.65	3	21	0.05
Autonomy	3.80	0.88	1	5	0.10
Team Controls					
Female Employees in Team (Scaled Prop.)	5.20	2.32	1	9	0.30
Organizational Controls					
Number of Employees	480.69	1154.72	8	9481	0.09
Country					
GDP per capita, PPP (constant 2017 international $)					
UK	45,974.86				0.00
Germany	53,011.77				0.00
Finland	47,481.21				0.00
Sweden	52,739.01				0.00
Netherlands	55,347.91				0.00
Portugal	33,086.10				0.00
Spain	39,575.54				0.00
Hungary	29,529.31				0.00
Bulgaria	21,363.03				0.00

Note: Means and standard deviations reported using the non-imputed dataset and all available information per variable (4,435 employees in 401 teams and 113 organizations)

Analytical Strategy

We evaluate our hypotheses and research question using the Bayesian framework which builds on the below uncentered Analysis of Covariance (ANCOVA) multilevel model.

$$Job\ Outlook_{ijk} = HetMen_{100}HeterogenousMen_{ijk} + DivMen_{200}DiverseMen_{ijk} + HomMen_{300}HomophilousMen_{ijk} + HetWomen_{400}HeterogenousWomen_{ijk} + DivWomen_{500}DiverseWomen_{ijk} + HomWomen_{600}HomogenousWomen_{ijk} + \gamma' X_{ijk} + v_{0k} + u_{0jk} + e_{ijk}$$

$$Income_{ijk} = HetMen_{100}HeterogenousMen_{ijk} + DivMen_{200}DiverseMen_{ijk} + HomMen_{300}HomophilousMen_{ijk} + HetWomen_{400}HeterogenousWomen_{ijk} + DivWomen_{500}DiverseWomen_{ijk} + HomWomen_{600}HomogenousWomen_{ijk} + \gamma' X_{ijk} + v_{0k} + u_{0jk} + e_{ijk}$$

As we are interested in how the networks of men and women are associated with career success, our ANCOVA looks at the corresponding six groups: heterophilous men, diverse men, homophilous men, heterophilous women, diverse women, and homophilous women. i refers to individuals in each team j of organization k. X_{ijk} represents the matrix of covariates or the control variables outlined above, for these individuals, and γ is the matrix of their corresponding covariate slopes. We use the six group effects to evaluate our hypotheses, specifically by looking at the effects of men having heterophilous ($HetMen$), diverse ($DivMen$) and homophilous networks ($HomMen$), and the effect of women having heterophilous ($HetWomen$), diverse ($DivWomen$), and homophilous networks ($HomWomen$).

We include random intercepts in this model to account for the multilevel structure of our data, where employees are embedded within teams, which are in turn embedded within organizations. This three level structure is reflected in the model's variance partition coefficient (VPC) for both job outlook and income. In case of job outlook, 1.48% of the variation lies between organizations, 9.77% lies within organizations between teams and 88.75% lies within teams between employees. In case of income these values are 3.58%, 24.69% and 71.73% respectively, indicating even greater clustering than job outlook. Although a two level model could be argued to sufficiently account for the variation, the three level model better accounts for the multilevel dependencies in organizational structures. It facilitates including controls on the employee, team and organizational level with lesser parameter estimation.

The use of ANCOVA instead of linear regression follows from the need to use Bayesian informative hypotheses evaluation, where the use of linear regression is found to have inferior results for unequal group sizes (Hoijtink, Gu, and Mulder 2019). As the six groups we outlined above are unequally sized, we use the multilevel ANCOVA model.

Our hypotheses use inequality constraints to outline specific expectations between the parameters. For example, our competing hypothesis (CH) 1.M2 expects gender diversity to be more beneficial than having network homophily or heterophily for men, symbolically: $DivMen > (HetMen, HomMen)$. Evaluating such informative hypothesis (Hoijtink 2011) using the traditional frequentist approach can be problematic on multiple fronts (Braeken, Mulder, and Wood 2015). One common objection is that the amount of information that a null hypothesis provides is usually nil (Cohen 1994). In the case of our example, the null hypothesis for M1 would state that $DivMen = HetMen = HomMen = 0$. Although rejecting this null hypothesis would tell us that something is going on, it does not tell us anything about the magnitude or direction of the effects. A number of follow-up tests would be required that are not necessarily independent or unambiguous. Thus, the approach does not allow direct mapping of hypotheses to results, and suffers from multiple testing problems that lead to low statistical power (Kluytmans, Van De Schoot, and Hoijtink 2012). In the case of CH1.M2, first three tests would be needed for each individual effect, then multiple pairwise comparison tests of the effects would have to be conducted to see if they are significantly different from each other and in the expected direction. This would be accompanied by additional test(s) to evaluate which competing hypothesis, 1.M2 (DivMen>(HetMen, HomMen)) or 1.M1 (HomMen>(HetMen, DivMen)) receives greater support. Such inequality constraints could make it very difficult, to combine the resulting set of p values into one single answer to whether or not our hypothesis is supported by the data. The multiple testing would also require type I error corrections, such as the Bonferroni alpha-corrections, resulting in tests that are very conservative with very little power (Cohen 1992). Additionally, traditional structural equation modelling fit indices are also not considered suitable, because such ordered hypotheses are not countable in terms of degrees of freedom (Braeken et al. 2015).

The Bayesian method, however, provides a single internally consistent solution for estimation and inference (Braeken et al. 2015). While hypothesis support in the traditional frequentist approach is expressed as the probability of the observed data or data that deviates more from the null hypothesis, given that the null hypothesis is true, in the Bayesian approach it is the support for any hypothesis given the data. Specifically, hypotheses are evaluated using Bayes factors that quantify to what extent the data supports one hypothesis compared to another. A Bayes factor of 1 would indicate equal support for both hypotheses. A Bayes factor above 1 would indicate preference for the first hypothesis in the comparison and a Bayes factor below 1 would indicate preference for the other hypothesis. For example, a hypothesis with a Bayes factor of BF1u= 4 indicates that Hypothesis 1 received four times more support from the data than hypothesis u. Thus the Bayes factor is never tied to one hypothesis, rather, it is the relative support for that specific hypothesis compared to another alternative hypothesis. This also makes Bayes factor apt for evaluating the competing hypotheses we have. Thus, we analyze all our hypotheses using Bayesian Informative hypothesis evaluation (BAIN) through the bain package in R 3.8.0. The Bayes factor are computed using equal prior model probabilities, the default in bain, which means that the hypotheses are seen as equally likely before observing the data.

As a first step we translated our hypotheses into inequality constraints using the above parameters. The symbolic inequality constraints are presented along with their corresponding hypotheses in Table 2.2. Broadly, we use two types of alternative hypothesis to evaluate our hypotheses. The first type is the unconstrained hypothesis which lets us compute BF.u: the Bayes factor of a hypothesis against the unconstrained hypothesis, that is where anything can be going on. We use this in evaluating our competing hypotheses, which are presented in Table 2.2 as CH1.M1, CH1.M2, CH1.W1, CH1.W2, and CH1.W3. The hypothesis with the highest BF.u, among the competing hypotheses, would receive the greatest support from the data.

The second type of alternative hypothesis we use is the complementary hypothesis which lets us compute BF.c: the Bayes factor of a hypothesis against its complement, that is any set of restrictions between the coefficients that is not the said hypothesis. For example the complement of the constraints in H2.M: $HomMen > HetMen$ would be $HomMen \leq HetMen$. Thus, BF.c can be used to evaluate Hypotheses 2 and 3. These hypotheses are presented in Table 2.2 as H2.M, H2.W, H3.1, H3.2, and H3.3.

Additionally, we could use the BF.u to compute the Bayes factor between pairs of competing hypotheses. For example the Bayes factor between CH1.M2 and CH1.M1 would be the ratio of their BF.u (BFii' = BFiu/BFi'u). This would directly quantify the support in the data for CH1.M2 as against CH1.M1.

Table 2.2 Symbolic inequality constraints for Bayes factor estimation

Hypotheses & Research Questions	Symbolic inequality constraints
CH1.M: Which network gender composition is the best at improving men's career success?	
CH1.M1: Network homophily has the strongest association with men's career success, as compared to network gender diversity and heterophily	$HomMen > (HetMen, DivMen)$
CH1.M2: Network gender diversity has the strongest association with men's career success, as compared to network homophily and heterophily	$DivMen > (HetMen, HomMen)$
CH1.W: Which network gender composition is the best at improving women's career success?	
CH1.W1: Network homophily has the strongest association with women's career success, as compared to network gender diversity and heterophily	$HomWomen > (HetWomen, DivWomen)$
CH1.W2: Network gender diversity has the strongest association with women's career success, as compared to network homophily and heterophily	$DivWomen > (HetWomen, HomWomen)$
CH1.W3: Network heterophily has the strongest association with women's career success, as compared to network gender diversity and homophily	$HetWomen > (HomWomen, DivWomen)$
H2: In token teams, it is more beneficial to connect with men	
H2.M: In token teams, network homophily has a stronger association with men's career success, as compared to network heterophily.	$HomMen > HetMen$
H2.W: In token teams, network heterophily has a stronger association with women's career success, as compared to network homophily.	$HetWomen > HomWomen$
H3: In non-token teams, men and women benefit similarly from their networks	
H3.1: In non-token teams, men and women have a similar association between their homophilous network and career success.	$HomMen = HomWomen$
H3.2: In non-token teams, men and women have a similar association between their diverse network and career success.	$DivMen = DivWomen$
H3.3: In non-token teams, men and women have a similar association between their heterophilous network and career success.	$HetMen = HetWomen$

2.4. RESULTS

We evaluate our first question—which network gender composition is the best at improving men's and women's career success?—using the unconstrained Bayes factors that are presented in Table 2.3. For men, we found greatest support in the data for the hypothesis of gender diversity being the most effective form of network for both job outlook and income. This can be seen in Table 2.3 with the Bfu for CH1.M2 being greater than the Bfu for CH1.M1, across both job outlook (1.931>0.395) and income (1.395>0.903). This is also reflected in the posterior model probabilities (PMP), where a higher value indicates greater support for a hypothesis. As CH1.M2 has the highest PMP on both job outlook (0.581) and income (0.423), it is the best hypothesis under consideration.

Table 2.3 Unconstrained Bayes Factors (BFu) for Men's and women's Competing Hypotheses (CH_{mn}) Versus the Unconstrained Hypotheses (Hu), and their Posterior Model Probabilities (PMP)

	Model Comparison Results			
	Job Outlook		**Income**	
Competing Hypotheses for Men	Bfu	PMP	Bfu	PMP
CH1.M1: Homophily is better for men	0.395	0.119	0.903	0.274
CH1.M2: Diversity is better for men	1.931	0.581	1.395	0.423
Hu		0.301		0.303
Competing Hypotheses for Women				
CH1.W1: Homophily is better for women	0.935	0.229	0.876	0.224
CH1.W2: Diversity is better for women	1.717	0.421	0.562	0.144
CH1.W3: Heterophily is better for women	0.430	0.105	1.479	0.378
Hu		0.245		0.255

Considering the two career success measures, the support for diversity is relatively higher for job outlook as compared to income. In fact, for job outlook the evidence in favor of diversity is almost five times that in favor of homophily. This is estimated by comparing the Bayes factor between CH1.M2 and CH1.M1 (1.931/0.395). Whereas for income the evidence in favor of diversity was about one and a half times that in favor of homophily (1.395/0.903).

For women, we saw different results across the measures. Just like for men, diversity was seen to be the best form of network for women's job outlook. However, heterophily was seen to be the best form of network for their income. This is reflected in the table with CH1.W2 having the highest Bfu of 1.717 for job outlook and CH1.W3 having the highest Bfu of 1.479 for income. This hierarchy was also mirrored in the PMP for both, job outlook (0.421) and income (0.378).

Thus we generally find support for network gender diversity to be the best at improving employee careers, except for women's income. This exception, could be traced to men having a socially higher value as seen in token teams, which are dominated by men. We study this rationale further with our second question—how the team's gender composition affect men's and women's network benefits?—using the complementary Bayes factors that are presented in Table 2.4 and Table 2.5. This is studied by separately looking at the gender difference in network benefits for token and non-token teams, using hypotheses 2 and 3 respectively.

Table 2.4 Complementary Bayes Factors (BFc) for Men's and women's Hypotheses on the Relative Benefits of Homophily and Heterophily in Token Teams

	Job Outlook	Income
H2: Token teams	BFc	BFc
H2.M: For men homophily is better than heterophily	1.335	0.827
H2.W: For women heterophily is better than homophily	0.915	0.794

For token teams, we expected ties to men to be more beneficial than ties to women. This was studied separately for men and women under hypotheses 2. For men in token teams, we hypothesize H2.M, where homophilous networks have higher career success than heterophilous networks. H2.M was supported for job outlook but not income. Specifically, for job outlook the support in the data for H2.M was 1.335 times higher than its complement. For income the support in the data was slightly more for the complement with a BFc of 0.827.

For women in token teams, we hypothesize H2.W, where heterophilous networks have higher career success than homophilous networks. For job outlook, H2.W saw a BFc of 0.915, which implies that there is about equal support for both H2.W and its complement in the data. For income, like men, the support in the data was slightly more for the complement with a BFc of 0.794. Thus, our second set of hypotheses, H2, saw only partial support in the data, specifically for the job outlook of men in token teams.

Table 2.5 Complementary Bayes Factors (BFc) for the Hypotheses on network returns for men and women in Non-Token Teams

	Job Outlook	Income
H3: Non Token teams	BFc	BFc
H3.1: Homophily has similar returns for men & women	22.813	5.508
H3.2: Diversity has similar returns for men & women	7.283	2.689
H3.3: Heterophily has similar returns for men & women	20.247	8.301

For non-token teams, we expected men and women to draw similar benefits from their networks. This was studied separately for each network type: homophilous (H3.1), diverse (H3.2) and heterophilous (H3.3). All three hypotheses saw strong support in the data relative to its complement. This is reflected in the BFc for H3.1, H3.2 and H3.3 being greater than one, across both job outlook and income. In fact, the support in the data was 22.813 times larger for the hypothesis of homophily having similar returns for men's and women's job outlook (H3.1) than the hypothesis of them having unequal returns, its complement.

Robustness Checks

As a robustness check we additionally evaluated the flip side of career success—career distress. Specifically we evaluated the above three sets of hypotheses again for the outcome of job security. Information on each individual's job security was collected using the individual level questionnaire, by asking employees to indicate the extent to which they agree or disagree with the statement "I worry about keeping my job". The employees could respond on a five point scale that ranged from 'strongly agree' to 'strongly disagree'. Thus a higher value would indicate better job security.

Broadly, we found similar results to those seen under job outlook and have detailed the findings in Table A3. This could follow from job outlook and job security being seen as similar measures. The competing hypotheses for men again showed greater support in the data for diversity being the most effective form of network for their job security. The competing hypothesis for women found both diverse and homophilous networks to have similar support from the data for being the most effective form of network, with the evidence in favor of homophily (CH1.W1) being about 1.188 times (1.300/1.094) that in favor of diversity (CH1.W2). These results for the second and third hypotheses, studying token and non-token teams, have been detailed in Table A4 and Table A5, respectively. For token teams, like job outlook, job security also had equal to marginal support for our hypotheses when compared to the complement. Finally, for non-token teams, we again saw strong support in the data relative to its complement.

A second robustness check we conducted was including an additional control for the sector in which the organization operated. Given that occupations often see gender segregation (Huffman, Cohen, and Pearlman 2010), they can be argued to play an important role in shaping employee networks and determining its returns. For instance, having a network dominated by women is more likely to be beneficial in a sector that is dominated by women than in a sector that is dominated by men. However, given the limited sample size we were unable to include sectors as a control for token teams. Nevertheless, we could include it as a control when analyzing the data for all teams, under the first competing hypotheses, and for non-token teams, under the third hypotheses. Overall, the results remained consistent with negligible changes in the Bayes factors. This is detailed in Table A6 and Table A7 for the first and third hypotheses, respectively.

2.5. DISCUSSION AND CONCLUSION

Although people often prefer to form ties with others of the same gender, and many still question the value of gender diversity efforts, our findings tell a different story. We show that networks composed of both men and women are generally more beneficial for career success than those that are mostly same-gender or opposite-gender.

These findings have important theoretical and policy implications. At the individual level, both men and women benefit more when their workplace networks are gender-diverse, especially in teams where women are not severely underrepresented. This supports theories of knowledge and exchange, which argue that diverse networks improve access to information, support innovation, and enhance performance (Bantel and Jackson 1989; Herring and Henderson 2015; Van Knippenberg et al. 2004). These advantages often outweigh the traditional benefits of homophily, such as ease of communication, trust, and cooperation. At the policy level, our results emphasize the need to go beyond efforts to increase representation and instead also focus on meaningful integration (Pedro Conceição et al. 2020).

An exception to this pattern is seen in women's income, where heterophilous networks are more beneficial than gender-diverse or homophilous ones. This aligns with expectation states theory, which argues that men are often seen as higher-status contacts, providing women with greater returns in terms of salary. This differential finding for women's income follows from women needing different networks to fulfill different motives (Ibarra 1992). For financial advancement, ties to men, who often hold positions of power, are advantageous. However, for non-financial motives, such as role modeling or psychosocial support, same-gender ties can be more beneficial (Ibarra 1997; Kram 1988). Our results reflect this divide: while women with more male contacts had higher income, those with gender-diverse networks had a better job outlook.

This income effect may also stem from how people compare themselves to others in their networks. Network ties often serve as benchmarks for salary expectations and negotiations. Because men are generally paid more than women, ties to men can raise women's reference points and support stronger salary demands. By contrast, women who compare themselves to other women may expect lower pay than those engaging in more gender-mixed comparisons (Bowles and McGinn 2008; Ostroff and Atwater 2003). Even employers' assumptions about women's networks can shape salary offers. Women from same-sex colleges, assumed by employers to have more female contacts to compare with, tend to receive lower wage offers (Belliveau 2005; Kolb 2009). Since women's heterophilous networks are by definition dominated by men, these higher-wage comparison points could have helped them negotiate better pay. This has important implications, especially for programs that promote same-gender ties in mentoring and networking programs (Dobbin et al. 2007; Pini et al. 2004).

Although men often hold positions of power, women report the most positive job outlook when their networks are gender-diverse. This reinforces our broader conclusion that gender-diverse networks offer the most consistent career benefits and challenges the belief that predominantly male networks are always more effective at improving career outcomes.

Even in teams where women are in token positions and men dominate, the advantage of having mostly male contacts is limited. We found that male-dominated networks were only more beneficial in one case: men's job outlook. In all other situations, networks with more women offered benefits equal to or greater than those of male-dominated networks. For example, a man with mostly female contacts in a token team experienced similar income gains as one with mostly male contacts.

This could follow from self-selection of women with above average abilities into token teams. For instance, female analysts are found to be more skillful than male analysts as they represent a special group of competitive women who choose to pursue a career in the male-dominated sector (Kumar 2010). Additionally, our findings could be explained by occupational gender segregation. Nearly 62% of the token teams belonged to manufacturing and transport sector. Both these sectors have globally seen an increase in the proportion of men, particularly in manual and production related jobs, while crowding out women into clerical and service related jobs (Bettio and Verashchagina 2009; Fortin and Huberman 2002; Rubery and Fagan 1993). In such settings, women's rarity and their differential role could increase their value as network contacts, reducing the relative advantage of male contacts.

Our study offers a novel contribution to research on gender diversity and networks by treating gender-diverse networks as a distinct category, rather than collapsing all ties into homophily or heterophily. This helps explain why previous studies reached mixed conclusions and highlights the importance of explicitly considering gender diversity in future network research.

While we relate employee motives to career outcomes, employees could also form different networks for different motives. Prior research shows that women tend to choose women in their support and friendship networks, but choose men in their influence, advice, and communication networks (Ibarra 1992). This suggests that the effects of network gender composition may vary depending on the type of network. In this study, we focus on affective cooperative networks or the colleagues an employee prefers to work and socialize with. These could be separated into affective work-based and social networks, but our data does not allow for this level of analysis. Future research could explore how gender diversity in different types of workplace networks relates to career outcomes.

Another limitation involves causality, a common challenge when studying network effects. It could be argued that career success drives employee networks, as against employee networks driving career success. It could be that career success shapes employee networks, rather than networks shaping career success. Given the cross-sectional nature of our data, it is difficult to clearly establish the direction of influence. To address this, we use a forward-looking measure—job outlook—which captures employees' expectations about their future in the organization. This supports the idea that networks can influence careers. Still, the observed link between women's heterophilous networks and higher income could reflect either direction. Women with financial goals may actively seek out male-dominated networks to improve their earnings, since men often hold more power. Conversely, women who already hold senior positions may be more likely to have men in their network as their higher status improves their legitimacy signal. Future research could disentangle the extent to which employee networks actually drive their career outcomes and vice versa using a longitudinal set up.

Future studies could also examine how integration and inclusion shape the effects of diverse employee networks on team and organizational outcomes. While research often shows that diversity enhances performance, innovation, and information exchange, these effects are not automatic. Insights from social and organizational psychology suggest that the benefits of diversity depend on whether individuals feel included, respected, and psychologically safe within their work environments. As Ellemers (2018) emphasizes, inclusion goes beyond demographic representation to encompass the experience of being morally valued and recognized as a full member of the group. Studies by Shore et al. (2011) and Nishii (2013) further demonstrate that inclusion moderates the relationship between diversity and outcomes by shaping trust, belonging, and engagement. Thus, to better understand why diverse employee networks yield particular effects, future research should integrate structural perspectives on networks with psychological theories of inclusion, examining how perceptions of belonging and fairness condition the impact of diversity on performance and cohesion.

From a policy perspective, studying gender diversity through individual networks offers a clearer understanding of how diversity functions in organizations (Reagans et al. 2004; Van

Knippenberg and Schippers 2007). Current policies often focus on changing a group's gender composition, but such changes can have both positive and negative effects depending on how individuals network (Sabharwal 2014). Simply increasing women's representation is not enough to reduce inequality (Bellotti et al. 2022). By instead focusing on the individual level, policy makers can use integration to successfully promote gender diversity in organizations (Vedres and Vasarhelyi 2022; Waldman and Sparr 2022).

Additionally, our findings could help address the growing backlash against gender diversity measures. For instance, men have filed almost 23% of all charges alleging sex-based discrimination with the U.S. Equal Employment Opportunity Commission in the past years (Weber 2018). The European parliament has also noted an increase in the general backlash witnessed by the EU against gender equality and women's rights which has had numerous negative effects on institutional, legal and policy systems aimed at combatting gender inequities (European Commission 2018; Juhász and Pap 2018). By showing that gender-diverse networks benefit men as well as women, our findings offer a more inclusive case for promoting diversity. In particular, since men with gender-diverse networks report the highest career success, these results support calls to frame gender diversity as an opportunity rather than a threat (Dover et al. 2019).

Summarizing, our study shows that gender diversity within employee networks plays a critical role in shaping career success. By treating gender-diverse networks as distinct from homophilous and heterophilous ones, we offer a new lens to understand integration at work. This approach helps reconcile mixed findings in the literature and reveals that gender-diverse networks consistently offer the greatest benefits, especially for job outlook. These insights can support individuals in making more strategic networking choices and help organizations design more inclusive policies that move beyond representation to foster true integration.

CHAPTER 3

Are Organizational Gender Diversity Management Practices Effective?

3.1. INTRODUCTION

Organizations around the world are increasingly adopting gender diversity management practices to reduce inequality. Since Norway introduced gender quotas for corporate boards in 2008, over 39 countries have adopted similar policies (Belaounia, Tao and Zhao 2020; Deloitte 2019; Terjesen, Aguilera and Lorenz 2015). Companies have also heavily invested in such GDM. For instance, Fortune 500 companies spent over $16 billion on diversity management in 2017 alone (Staley 2017).

Despite these investments, the effectiveness of such practices remains unclear. They can increase discord and provoke backlash. For example, one online forum titled "Ever work for a woman? Roll up your sleeves and share your horror story" was created in response to GDM (Bowles 2017). Such reactions highlight a deeper issue: while GDM practices may raise the number of women in the workplace, they often fail to improve how well women are integrated into workplace relationships.

This crucial oversight stems from how gender diversity is measured. As discussed in the previous chapter, gender diversity is often measured by representation, such as the proportion of women in leadership, rather than by how individuals actually connect and collaborate. These simple counts are easy to track but only show the potential for integration. Measuring true integration requires more detailed information about who employees choose to work with or spend time with outside of work—their affective networks. While gender composition might suggest potential connections, it does not capture actual social bonds. Understanding gender diversity in teams means knowing who people choose to interact with, not just who is present.

As a result, organizations tend to design and assess GDM practices based on increasing representation, while overlooking integration. This narrow focus can unintentionally reinforce inequality and may explain why certain practices persist despite backlash and mixed evidence on their success (Dobbin, Schrage and Kalev 2015). For example, gender quotas may lead to hiring or promoting more women, which improves representation and makes GDM appear successful. At the same time, these practices can provoke resistance, induce perceptions of these women being less competent or influential, and discourage collaboration (Heilman and Welle 2006). This limits women’s access to informal networks and can widen gender gaps in pay, promotion (Cullen and Perez-Truglia 2019), job satisfaction and employee turnover (Grissom, Nicholson-Crotty and Keiser 2012). The previous chapter illustrated the crucial role integration plays in shaping both employee careers and organizational outcomes.

Despite integration’s importance, the scant studies evaluating GDM focus only on its effectiveness in improving women’s representation (e.g. Dobbin and Kalev 2013; Kalev, Dobbin and Kelly 2006; Timmers, Willemsen and Tijdens 2010). To address this gap, we introduce a new approach that evaluates gender diversity through both representation and integration. Building on the measure from the previous chapter, we use data on employees’ affective networks, i.e. whom they choose to work with or meet outside of work, to measure the integration level in a team.

We compare GDM effectiveness in improving representation and integration using both the traditional measure of representation in management and the new measure of integration. This lets us identify whether practices that appear successful on paper also support real inclusion. We expect that the traditional measure will overstate effectiveness, while the integration measure will better capture GDM's unintended consequences.

We also examine how GDM effectiveness depends on how it is implemented. Drawing from earlier research, we group GDM strategies into three approaches, each reflecting a different level within the organization: reducing isolation at the individual level, reducing bias at the managerial level, and establishing responsibility at the organizational level. We evaluate how effective each approach is in improving both representation and integration.

Analyzing such relative effects is difficult, if not impossible, using conventional null hypothesis significance testing or model comparison (Braeken, Mulder and Wood 2015). For example, null hypothesis significance testing does not allow direct mapping of our hypotheses to results and may lead to problems such as multiple testing and low statistical power. We adopt Bayesian informative hypotheses evaluation, which allows us to directly compare the effectiveness of different practices across both gender diversity evaluation measures. As in the previous chapter, we use the second wave of the European Sustainable Workforce Survey (Van der Lippe et al. 2018), as it captures employee integration, through novel information on the affective networks of 4345 employees from 401 teams across multiple countries and sectors.

This study has important policy and practical implications. Current evaluations that focus only on representation may lead organizations to adopt policies that look successful on the surface but worsen problems of inclusion, attrition, and exclusion (Castilla 2015; Rivera and Tilcsik 2019). For instance, the increased focus on gender quotas could be at the cost of poor integration and higher turnover rates for women. This calls for an urgent need to step back, evaluate and redesign GDM. The new measure proposed here is a step in that direction, offering a tool for policymakers and organizations to assess and manage gender diversity by tracking not just representation but also the integration of men and women.

3.2. GDM SIGNALS AND ITS IMPLEMENTATION SHAPES ITS EFFECTIVENESS

GDM Signals Have Different Unintended Consequences for Integration and Representation

Effective GDM depends on evidence-based practices. However, due to limited evaluations of diversity programs (Kalev, Dobbin and Kelly 2006), organizations often rely on well-intentioned but nonevidence-based practices (Davis, Frolova and Callahan 2016; Dover, Kaiser and Major 2019; Edelman et al. 2011). This is a critical issue as even the meagre presence of GDM can have negative unintended consequences. For instance, GDM initiatives such as diversity training or recruitment targets can activate stereotypes or be perceived as unfair, undermining trust and collaboration to actually worsen inequality (Dover et al. 2019).

Such unintended consequences stem from three key signals that GDM can send: fairness, inclusion, and competence. Fairness signals may create the impression that the organization is already equitable, making it harder to identify and address discrimination. They may also lead men to feel unfairly treated. Inclusion signals can make firms more attractive to women, yet may also trigger feelings of exclusion or threat among men. Competence signals suggest that women need special support to succeed, which can unintentionally portray them as less capable (Ashforth and Mael 1989; Dover et al. 2019).

While such signals may not prevent women from being hired, they can negatively influence how women are treated once in the workplace. Women may achieve numerical representation but still face subtle exclusion, limiting their ability to form meaningful professional ties. Men, in turn, may respond with stronger in-group loyalty, weakening integration. Thus, GDM may appear effective by increasing representation, while failing to foster actual integration thus limiting the benefits associated with that.

These unintended signals show why it is essential to evaluate GDM using both representation and integration. As discussed in the previous chapter, focusing only on representation can hide whether women are truly included. Some practices may increase representation while harming integration. Yet, signaling theory has largely ignored how such negative signals affect everyday workplace interactions (Connelly et al. 2011). By comparing GDM outcomes on both dimensions, we improve evaluations of gender equality and contribute to signaling theory by showing how GDM can unintentionally shape integration.

To do this, we use two distinct measures: the current gender diversity (CGD) measure, based on representation, and the new gender diversity (NGD) measure, based on integration. We refer to these as CGD and NGD, respectively.

Implementation of GDM

The level at which GDM is implemented will also determine its effectiveness. The implementation of any diversity management practice is highly subjective, giving varying signals depending on how practices are designed and executed. However organizations have taken approaches that can be categorized into three broad areas, by: (I) reducing individual isolation (II) moderating managerial bias, and (III) establishing organizational responsibility (Kalev, Dobbin and Kelly 2006). GDM's unintended consequences will vary based on which of the three approaches were used to implement it. Thus, we look at the relative effectiveness of these approaches in the context of GDM.

This categorization also aligns with the three mechanisms for remediating workplace inequality rooted in different social science literatures as outlined by Kalev, Dobbin and Kelly (2006), in their pioneering evaluation of diversity management practices. They used US federal establishment-level data on private establishments to assess changes in managerial composition after the adoption of seven diversity practices. We study all seven practices they assessed. These are categorized in Table 3.1 according to the three approaches.

Table 3.1 The three GDM approaches evaluated

I. Individual	**II. Managerial**	**III. Organizational**
Networking programs for women Mentoring programs for women	Diversity training for managers Diversity evaluations for managers	Affirmative action plans Diversity managers Diversity committees or diversity task forces

The first approach we evaluate targets individuals and draws on social network theories to counter isolation of women and to improve their career prospects. This is motivated by findings on women having limited access to or being excluded from organizational networks (Ibarra 1993). Additionally, women are also found to lag behind their male counterparts in both the quality of on-the-job contacts and the profit they are able to derive from these contacts. Both genders tend to have networks dominated by same-gender co-workers. However, as men often occupy managerial positions and have greater access to information, women profit less from their work networks. Furthermore, even when women invest in high quality work networks, they tend to receive smaller returns (Son and Lin 2012). Overall, practices with this approach focus on helping women make useful contacts and access information. Minorities have found initiatives on networking and mentoring to be useful (Wernick 1994). Thus, we evaluate these strategies through the prevalence of the following GDM practices: (1) networking programs for women and (2) mentoring programs for women.

The second approach, targets managers and draws on theories of stereotyping and bias. These theories focus on eliminating managerial bias to reduce inequality. As gender stereotypes contain status beliefs they tend to create a web of constraining expectations and interpersonal reactions that is seen as a major cause of the 'glass ceiling'. Managers may unwittingly let these beliefs shape the evaluation of employee performances, the ability attributed to them, and the likelihood of their promotion (Ridgeway 2001). Additionally, drawing from social categorization (Tajfel 1981; Turner 2010) and similarity-attraction (Byrne 1971) theories, managers may show in-group favoritism. As there are disproportionately more male managers, they are likely to favor male employees over females, trusting them more, and showing greater willingness to bond with them (Van Knippenberg and Schippers 2007). Two corporate initiatives are thought to counter this stereotyping and in-group bias among managers: education via diversity training and feedback via performance evaluations. While the former seeks to provide information, the latter evaluates managers on their diversity performance to create oversight and provide feedback. For example, manager appraisal in some companies also evaluates their success in implementing the organization's diversity practices (Kalev, Dobbin and Kelly 2006). We evaluate both these strategies through the prevalence of the following two GDM practices, respectively: (3) diversity training for managers and (4) diversity evaluations for managers.

Finally, the third approach targeting the organization as a whole, draws on the organizational institutionalists' argument of achieving goals (reduced gender inequality) by creating specialized positions. They argue that organizations that do not assign responsibility for diversity goals to a specific agent, may see their goals fall by the wayside as individuals face information overload, and stick to familiar tasks (Edelman 1990; Meyer and Rowan 1977). Thus, practices with this approach include strategies which focus on assigning responsibility within the organization structure. This is primarily done through three common approaches: responsibility and affirmative action plans, oversight via staff positions and departments, and finally, oversight and advocacy via committees. We consider all three approaches by operationalizing the prevalence of the following three GDM practices, respectively: (5) affirmative action plans, (6) diversity managers and (7) diversity committees or diversity task forces.

Thus, we evaluate the effectiveness of the three GDM approaches in improving both representation and integration using two distinct measures: CGD and NGD

3.3. HYPOTHESES ON THE EFFECTIVENESS OF GDM

Gender Diversity Measures: CGD vs. NGD

All GDM approaches, regardless of how they are implemented, can have unintended consequences through fairness, inclusion, and competence signals. While GDM aims to improve both representation and integration (Yang and Konrad 2011), these signals are more likely to undermine integration than representation. Fairness and inclusion signals may help attract women and improve representation, but they can also provoke negative reactions from male employees, leading to resistance, stereotyping, or exclusion once women are hired. Competence signals, which imply that women need extra help, may further weaken integration by discouraging collaboration and respect (Dover et al. 2019).

Although the unintended consequences from fairness and competence signals are reflected in employee integration, inclusion signals could also be argued to alter representation—by influencing how job seekers are attracted to a firm. Inclusion signals often tend to make the organization more attractive to job seekers from the underrepresented group, here women, consequently improving their representation. For instance, recruitment messages that highlight a firm's support for diversity are found to attract more diverse applicants (Avery and McKay 2006; Thomas and Wise 1999; Williamson et al. 2008).

For women, despite the intention to signal inclusion, there is also evidence of it prompting greater concerns about fairness and being liked. For instance, GDM practices might encourage women to enter organizations that have promised an inclusive and fair workplace but this may be contrary to the realities of what they actually encounter. This could set unrealistic expectations among them, which could backfire and deter their integration once they are hired (McKay and Avery 2005; Wilton et al. 2020). Thus, although women's representation could capture the signal's intended effect, their integration would capture the unintended consequences.

In case of men, the traditionally overrepresented group, these signals often do not directly influence representation as it seldom deters their attraction to the firm. Nevertheless, these signals could also negatively influence their integration, specifically through their attitudes and behavior towards the underrepresented group (Dover, Kaiser and Major 2019). For instance, Dover, Major and Kaiser (2016) found that while diversity messaging did not change white men's interest in an employer, it did increase concerns about being disadvantaged because of their race.

Thus, the unintended consequences of these signals are more likely to affect the integration, rather than representation of women employees. As the new measure explicitly accounts for integration, evaluations based on it are more likely to account for the negative effects of GDM. Therefore evaluations on the current measure, based on representation, makes GDM practices look more effective.

> *Hypothesis 1: Having any GDM practice would appear more effective on the current representation-based measure (CGD), than the new integration-based measure (NGD).*

Which GDM Approach is More Effective for Representation?

GDM practices with an organizational approach are best positioned to influence its gender representation. For instance, organizations practicing affirmative action can alter their hiring, training and promotion decisions to improve women's representation (Crosby et al. 2003). Practices with an organizational approach can thus alter gender composition across the firm. As the current measure is based on representation, evaluations based on it are likely to make practices with this approach to appear the most effective. Although practices with the managerial approach could also alter gender representation through promotion decisions, this relationship is indirect. For instance, the practice of diversity training for managers could nudge them to promote more women, thereby increasing managerial representation of women.

Practices with an individual approach are also not positioned to directly influence women's representation. Networking and mentoring programs primarily aim to provide useful contacts, support and information (Blake-Beard 2001; Castilla 2005; Thomas 2001) to improve interactions within the firm. Although these practices do not directly alter women's overall representation, mentoring can be argued to increase women's promotions and hence their representation in the upper levels. However, its success would depend on whether the organization implements homogenous or heterogenous mentoring, where they connect same and other sex mentors and mentees, respectively. Homogenous mentoring would require a high initial presence of women in the upper levels to be effective, which is seldom observed. In fact, Athey, Avery and Zemsky (2000) call this practice of homogenous mentoring an "invisible barrier" as it stalls the progress of minorities until the diversity of the upper level mirrors the lower level. Organizations are typically seen to implement homogenous mentoring (Ibarra, Carter and Silva 2010). Overall, as practices with an individual approach are specifically designed towards improving interactions among existing women rather than women's organizational representation, they are the least likely to appear effective on representation-based measures.

Hypothesis 2: When evaluating based on representation (CGD), GDM practices with an organizational approach appear most effective and those with an individual approach appear least effective.

Which GDM Approach is More Effective for Integration?

Our new measure evaluates GDM practices also based on their success in promoting interactions among men and women. As practices with a managerial approach are focused on interactions rather than just representation, we expect it to play the biggest role in promoting gender diversity. This is also backed by cognitive-dissonance and self-perception theories (Bem 1967; Festinger 1957) which suggest that by engaging managers in leading change, firms can increase managers' support of that change (Dobbin, Schrage and Kalev 2015).

Although practices with an individual approach also focus on interactions, they additionally alter resources available to men and women, and are therefore likely to increase social competition among them. For instance, mentoring programs for women would alter the pool of mentors (resources) available to men. The resource competition would threaten the social bonds between men and women, and their shared identity as a group. It would also amplify how employees perceive or even provoke greater gender differentiation than exists and disparage the other-gender group (Ashforth and Mael 1989; Brown et al. 1986; Turner 1975).

Additionally, practices with an individual approach typically promote interactions within the same gender.. Following from the social categorization theory, this is likely to foster greater in-group favoritism, increase conflict and limit within-unit behavioral and social integration (Van Knippenberg and Schippers 2007). Additionally, the exclusive focus of these initiatives on women might aggravate the negative effects of fairness and competence signals. Special networking and mentoring programs for women might be perceived as unfair by men, or alternatively could portray women as in need of help and as less competent than their male counterparts.

In the case of practices with an organizational approach, its implication for integration is unclear. Practices under this approach assign responsibility for diversity goals. As gender diversity was traditionally evaluated against a measure of representation, these goals typically focus on improving representation. For instance, Edelman and Petterson (1999) showed that equal opportunity departments expand diversity recruitment programs to improve diversity. The disregard for integration could make these practices more susceptible to the three signals.

Note that practices with organizational and individual approaches, by design, tend to be publicized throughout the organization, as its implementation typically engages most employees. For example, networking programs for women are likely to span across the organization. Therefore, the signals from practices under organizational and individual approaches will be directly perceived by all involved employees. However, the implementation of practices with a managerial approach, by design, targets just the managers. Therefore only the managers would be able to directly perceive the signals from these practices, as compared to most employees in case of practices with organizational and individual approaches. Thus, practices with a managerial approach are likely to be the most effective, as the their signals are largely limited to managers.

> *Hypothesis 3: When evaluating based on integration (NGD), GDM practices with a managerial approach appear most effective.*

Is GDM Effective?

The unintended consequences of GDM could very well overshadow its benefits to render practices ineffective. For instance, although diversity recruitment initiatives may increase minority recruitment, they could increase turnover rates among new hires if the actual workplace diversity climate is unfavorable (McKay and Avery 2005). Thus, we explore whether adopting either of the three approaches to GDM is more beneficial than not having any GDM practice, based on both, the current and new gender diversity measures.

> *Exploratory Question: Are the three approaches to GDM better than not having any GDM practice?*

3.4. Data and Method

We analyze data from 4,345 employees across 401 teams using linear regression. To evaluate the effectiveness of the three GDM approaches on both gender diversity measures, we apply Bayesian informative hypothesis evaluation. The current gender diversity (CGD) measure captures group-level gender ratios. To ensure comparability, the new gender diversity (NGD) measure, which reflects integration, is also aggregated at the team level. All analyses are therefore conducted at the team level, and Bayes factors are used to assess the relative effectiveness of each GDM approach across CGD and NGD.

Data
As introduced in the previous chapter, we use data from the second wave of the European Sustainable Workforce Survey (Van der Lippe et al. 2018). The survey was conducted in 2017–2018 across nine European countries: Bulgaria, Finland, Germany, Hungary, the Netherlands, Portugal, Spain, Sweden and the United Kingdom. We used the second wave as it included information on employees' affective network, which is needed for constructing the NGD. Organizations were invited based on their participation in the first wave (2015–2016), which used stratified random sampling by organization size and sector. This resulted in a final sample of 4,345 employees across 401 teams in 113 organizations.

Surveys were administered to HR managers, team managers, and employees, generating a multilevel dataset that links organizational, managerial, and individual responses. This structure allows us to estimate both CGD and NGD and to analyze them in parallel by aggregating information at the team level.

Since current gender diversity measures reflects the gender ratio of a group, CGD can be calculated at either the organizational or team level. However, NGD focuses on integration, which varies significantly between individuals and across teams within an organization. To accurately compare CGD and NGD, we conduct all analyses at the team level. Focusing on teams reduces variation that may arise in larger organizations spanning multiple departments or locations. It also allows us to control for how local managerial practices and immediate work environments influence both representation and integration.

Measures
Current Gender Diversity Measure (CGD): Despite there being a variety of gender diversity measures, they are all typically based on the gender ratio in a group. One of the most common measures used is the simple proportion of women in a group (Adams and Ferreira 2009; Francoeur, Labelle and Sinclair-Desgagné 2008). In fact the pioneering evaluation of the effectiveness of diversity management practices by Kalev, Dobbin and Kelly (2006) was using the proportion of women in management. The use of data on managers is driven by the need to account for integration, as managerial jobs are seen to be the hardest to integrate (Dobbin, Schrage and Kalev 2015; Stainback and Tomaskovic-Devey 2012). This could be explained by sex based ascription (Reskin and McBrier 2000) and the social closure theory; where individuals sharing an identity group, often men, tend to hoard desirable opportunities and high rewards by excluding out-groups, often women. The shared identity is said to create a mutual obligation which could be used to justify discrimination in favor of men (Reskin 1988; Ridgeway 1997; Tilly 1998). Thus, it proposes that low paying jobs would integrate earlier, and high paying jobs, such as in management, will integrate more slowly.

Therefore we use the CGD measure based on the managerial gender ratio to ensure that we account for the current measure's ability to implicitly capture integration. However, using the simple proportion of women in management to evaluate the effectiveness of GDM practices could prove inadequate on multiple fronts. For instance, the measure itself is not necessarily a measure of diversity. For instance, the highest possible value on this measure would imply that a group has 100% women and no men. As GDM often seeks to attain equal representation, it should strive for having 50% women. Additionally, this measure would also ignore those settings where men are seen as a minority, as in parts of the health and education sector.

In view of this we formulate the managerial gender ratio using the normalized Herfindahl–Hirschman index (HHI)[4], a popular formulation for diversity which provides comparable measures on a uniform scale. It is maximized when a group has equal number of men and women, and minimized if the group has only either men or women. Specifically we use information on the proportion of women (w_t) and men ($1 - w_t$) in the organizational management of team t to estimate the CGD for each team, using the normalized HHI as below:

$$\mathrm{CGD}_t = \left[1 - \frac{[w_t^2 + (1 - w_t)^2] - 0.5}{0.5}\right] \times 100$$

This measure produces a scale from 0 to 100. A score of 0 indicates that all managers are of one gender, while a score of 100 reflects an equal proportion of male and female managers. Using the Herfindahl index aligns with GDM's role of leveraging the benefits of diversity. When men and women bring different knowledge and perspectives, greater heterogeneity can promote creativity and innovation. The Herfindahl index is also one of the most common measures of diversity or concentration and takes on different forms across disciplines, such as the Simpson's index in ecology and Blau's index in in sociology (Harrison and Klein 2007).

New Gender Diversity Measure (NGD): The NGD measure was constructed by again using the normalized HHI in the same manner as outlined above. However, instead of using the proportion of women managers in an organization it is based on the proportion of women (w'_i) and men ($1 - w'_i$) in employee *i*'s affective network. We then estimate the NGD at the team level by averaging the normalized Herfindahl index for all *n* employees in team *t*, ensuring that the NGD and CGD are comparable. This was computed as below:

$$NGD_t = \frac{1}{n}\sum_{i=1}^{n}\left[1 - \frac{[w'^2_i + (1 - w'_i)^2] - 0.5}{0.5}\right] \times 100$$

Information on w'_i was collected using employee responses to the following two open questions:

1. "Whom in your department do you also see outside work?"
2. "Whom do you like to work with in your department?"

The combined responses to both these questions form the individual's affective network and reflects their integration. This follows from integration requiring voluntary choice by the employees on whom they form or want to form social bonds with, both at work and outside of work. Looking at the responses independently would provide an incomplete picture, as it would only be a subset of the employee's social bonds. Additionally, as almost 13% employees provided specific names for only one of the two questions, combining their responses also helped improve the robustness of our analysis.

We first ensured that the combined responses included only unique nominations. If an employee named the same colleague under both questions, that alter was counted only once.

[4] Blau's index or HHI ranges from 0 to (K-1)/K, where 0 reflects having no diversity and K refers to the number of groups. In the context of gender where we have two groups (men and women), K=2. Thus the range would be 0 to 0.5, where 0 implies having members of only one gender and 0.5 implies having equal members of both gender. The Normalized HHI however ranges from 0 to 1 to ensure comparability across multiple HHIs, by dropping information on the group (K). As we have a constant group K=2, we use the normalized HHI.

We then identified the gender of each named colleague to estimate w'_i or the proportion of women in each employee's network.

Gender identification followed three steps. When organizations provided a list of employee names, we were able to match named colleagues to survey respondents and use their self-reported gender. For organizations that did not provide a name list, an algorithm was developed in R 3.8.0 to match names to the World Gender-Name Dictionary: a roster of 6.2 million names from 182 different countries (Lax Martínez, Raffo and Saito 2016). Finally, for names that could not be classified by the algorithm, we manually identified gender using native speakers and online searches on social and professional networking platforms. The NGD was then estimated only for those employees whose entire network could be reliably gender-identified.

Note that both the gender diversity measures we use, CGD and NGD, are gender neutral and can account for discrimination against both men and women. This is important given the tendency for men and women to work in different types of jobs, or occupational gender segregation (Leung and Koppman 2018).

GDM Practices: To compare the effectiveness of the three approaches to GDM, four binary variables were created. The first three variables captured whether or not the organization with the said team rolled out any practice that had an individual, managerial or organizational approach, as outlined in Table 3.1. For instance, consider an organization that only implemented two practices: (i) networking programs for women, which is with an individual approach, and (ii) affirmative action plans, which is with an organizational approach. Then the GDM variables for the individual and organizational approach would take the value of one, while the managerial approach variable would take the value of zero. Note that the time and frequency of these practices can be important for its effectiveness. For instance, conducting monthly networking programs could be argued to be more effective than annual networking programs. However, the original questionnaire only captures the presence or absence of practices.

The fourth variable was constructed to indicate if the organization had no GDM practice. A value of one indicates that none of the seven practices discussed earlier were adopted, and the value of zero indicates having at least one of the seven practices. This variable is used to compare and evaluate whether GDM practices are effective, i.e. if having GDM is better than not having any GDM to begin with.

Like all self-reported measures, our GDM measure risks social desirability bias. Organizations might find it more desirable to report having a GDM practice, as against not having a GDM practice. However, this is unlikely to be the case as 50% of the organizations reported not having any GDM practice. While there is no objective benchmark we can use to evaluate whether a value of 50% indeed indicates a lack of social desirability bias, Huang et al. (2019) provide some evidence towards. They found 87% of their surveyed American companies to be committed to GDM. As this is significantly higher than the 50% we find, it is unlikely for our GDM measure to gravely suffer from social desirability bias.

Control Variables: Given the multilevel nature of the data, controls were used at the individual, managerial and organizational level. As we look at the effectiveness of GDM on representation and integration, we control for potential confounders on both relationships. To ensure comparability of the effectiveness of practices across the CGD and NGD, we aggregate and analyze all variables at the team level. Specifically, all individual controls were aggregated within a team to reflect the corresponding average for the team. In case of categorical variables, the proportion of team members under each category is used. Thus, when categorical variables at the individual level are aggregated to the team level, they are transformed to continuous variables.

At the individual level, we account for attributes likely to shape interpersonal interaction, based on similarity-attraction theory (Byrne 1971). This includes the proportion of employees in a team who live with a *partner*, have *children* and are *women*. We also controlled for the team's average frequency of *working from home* (ranging from 1-7; 1 = never or almost never, and 7 = 4 or 5 days a week), , as networking opportunities may differ by physical presence. Finally, we included factors previously reported in the literature to associate with network integration (Ibarra 1992), namely *age*, *educational years, earnings* and *occupational status*. We operationalized these controls by averaging across all employees in a team. To ensure comparability, the educational years were recorded as International Standard Classification of Education (ISCED) categories and recoded based on the PISA 2012 technical report to reflect the official number of years an individual spent receiving education (OECD 2014). The team's average net monthly earnings of the team was computed using individually reported earnings in the local currency and was standardized to reflect Purchasing power parity (PPP) in terms of the USD. This was done using the World Bank's International Comparison Program (ICP) PPP conversion factors for the reference year 2017. Each individual's occupational status was made comparable by constructing it on the International Socio-Economic Index of occupational status (ISEI-08 or new ISEI). As the ISEI is empirically constructed as a continuous and hierarchical indicator of occupational status, we can aggregate the team's occupational status by averaging the ISEI score of all team members (Ganzeboom and Treiman 2003).

At the team level we control for the type of *task* performed and the *proportion of female employees*. Given that the task performed by a team could shape both, its gender composition (Correll 2004) and the interactions (Mabry 1985; Morris 1966), we control for whether the team had a supporting or core task. The latter referred to teams that represent the organization's core activities. For example, in case of a hospital this could be the nurses, or in case of a university this could be a research department. The teams that did not perform core activity of the organization were said to have a supporting task. For example, this could include an organization's finance, communication or maintenance department.

The second control follows from the team's gender ratio providing the opportunity structure for an employee's network. This is accounted using the scale on proportion of female employees in a team (1=None, 2=1% to less than 10%, 3=10% to less than 20%, 4=20% to less than 40%, 5=40% to less than 60%, 6=60% to less than 80%, 7=80% to less than 90%, 8=90% to less than 100%, 9=All). Although the questionnaire captured this variable using an interval scale, we treat it as a continuous variable to improve the robustness of our analysis. The average proportion of female employees in a team was found to be 5.2, i.e. around 40-60%. This is understandable as the data has teams from traditionally women dominated sectors, like health care and education.

At the organizational level we controlled for the opportunity employees would have to network and the availability of managerial jobs. This was done using the *number of employees* in the surveyed establishment of the organization (Baron, Mittman and Newman 1991), the *proportion of managers* among them (Kalev, Dobbin and Kelly 2006), and looking at whether the organization offered *flexible work arrangements* (Yes = 1). As the latter is an organizational work–life policy, it not only alters networking opportunity but also reduces gender inequalities (Van der Lippe, Van Breeschoten and Van Hek 2019), making it an important control in evaluating the effectiveness of GDM. We additionally control for attributes that could influence an organization's GDM as a whole. For instance, company laws and regulations might have different directives on GDM depending on the organization's *ownership* (public, private or others), its *sector* and *country* of operation. The nine countries and six sectors surveyed provide for ample variation in policies, (gender) culture and labor market characteristics. For instance, both high (Health care and Education) and low female employment (Manufacturing and Transport) sectors are included. Further, the nine countries surveyed constitute different types of welfare regime (Esping-Andersen, 2009): Finland and Sweden belong to the socio-democratic regime; Germany, Spain and Portugal to the conservative regime; the UK to the liberal regime; Hungary and Bulgaria are post-communist; the Netherlands is a mixed case.

Table 3.2 presents the aggregated summary statistics at the team level, based on all available data for each variable. The disaggregated statistics at the individual, team, and organizational levels, along with details on missing observations, are provided in Table B1. These, along with all supplementary materials for this study, are available in Appendix B.

As detailed earlier, the NGD measure was constructed by averaging individual level information on the proportion of women and men in an employees' affective network. However, this proportion could only be estimated when the gender of all alters in an employees' network was known. This means that the network would be incomplete if the gender of even one alter was unidentified or missing. Consequently, almost 84% of the employees in our data had incomplete network information. As it was only possible to construct the NGD when complete network information was available, it saw a high proportion of missing information.

To address this we performed multiple imputation using Fully Conditional Specification through the MICE algorithm in R 3.8.0 (van Buuren and Groothuis-Oudshoorn 2010; van Buuren et al. 2015) with 100 imputations and 20 iterations on the disaggregated data. The active dataset used for the imputation was selected in line with the predictor variable selection strategy outlined by Van Buuren (2018). This included all available information on the above variables of interest along with other variables that could explain a considerable amount of variance or related to nonresponse, such as whether they were surveyed online or using paper-and-pencil. Despite including the NGD, which saw a high missing proportion, only about 12% of the cases in the active dataset were missing. This implies that our use of 100 imputations not only complies with the suggested de-facto standards but also ensures that the essential quantities in our analysis should be reproducible within some limit (Von Hippel 2009; White, Royston and Wood 2011). Given the multilevel nature of the dataset we modeled the team and organizational clustering of the data as a fixed effect during imputation and applied the predictive mean matching (PMM) method. This method has been noted to have comparable performance to dedicated methods for multilevel data (Vink, Lazendic and van Buuren 2015). Additionally, it facilitates imputation for the whole dataset which is not necessarily straightforward with multivariate multilevel models. The imputed data was then aggregated at the team level for the analysis.

Table 3.2 Descriptive statistics aggregated at the team level

	Mean/ Proportion	**SD**	**Min**	**Max**
Key Variables				
New Gender Diversity Measure	29.38	35.47	0	100
Current Gender Diversity Measure	59.71	36.09	0	100

	Mean/ Proportion	SD	Min	Max
Individual Approach Practices	0.14			
Managerial Approach Practices	0.27			
Organizational Approach Practices	0.40			
No GDM Practices	0.55			
Individual Controls				
Partner: Yes	0.75	0.22	0	1
Children: Yes	0.38	0.27	0	1
Gender: Female	0.56	0.35	0	1
Working From Home (Scaled)	1.74	0.98	1	5.8
Age	44.04	7.19	25.5	66.67
Educational Years	13.45	2.64	4	20.5
Earnings in PPP (USD)	2322.38	1842.75	399.18	28475.48
Occupational Status	51.31	12.72	16	85
Team Controls				
Female Employees in Team (Scaled Prop.)	5.20	2.32	1	9
Function: Core task	0.72			
Function: Supporting task	0.28			
Organizational Controls				
No. Employees in Establishment	574.64	1428.98	8	9481
Proportion of Managers	0.12	0.18	0	1
Flexible Work Arrangements: Yes	0.78			
Ownership: Private	0.51			
Ownership: Public	0.38			
Ownership: Others (e.g. Mixed – public & private)	0.11			
Sector				
Manufacturing	0.36			
Health Care	0.18			
Higher Education	0.20			
Transport	0.10			
Financial Services	0.06			
Telecommunication	0.09			
Country				
UK	0.06			
Germany	0.06			
Finland	0.06			
Sweden	0.08			
Netherlands	0.22			
Portugal	0.07			
Spain	0.05			
Hungary	0.12			
Bulgaria	0.28			

Note: Means and standard deviations reported using the non-imputed dataset aggregated at the team level using all available information per variable (401 teams)

Analytical Strategy

We evaluate our hypotheses using the Bayesian framework which builds on the below linear regression models for the two measures of gender diversity. To ensure comparability across the gender diversity measures all variables are estimated and analyzed at the team level.

$$TGD = \beta_{0.tgd} + \beta_{ind.tgd}Individual + \beta_{mgr.tgd}Mangerial + \beta_{org.tgd}Organisational + \beta_{no.tgd}NoGDM + \beta_{tgd}{}'Control + \varepsilon_t$$

$$NGD = \beta_{0.ngd} + \beta_{ind.ngd}Individual + \beta_{mgr.ngd}Mangerial + \beta_{org.ngd}Organisational + \beta_{no.ngd}NoGDM + \beta_{ngd}{}'Control + \varepsilon_t$$

The $\beta_{x.y}$ coefficients will be used to evaluate our hypotheses. Specifically, eight coefficients will be used which are specified using four predictor x-variables—practices with individual (ind), managerial (mgr) and organizational approach (org), and not having any GDM (no)—and two outcome y-variables—CGD (cgd) and NGD (ngd). For example, $\beta_{org.ngd}$ would indicate the effectiveness of having practices with an organizational approach on the NGD measure. The variable Control reflects the vector of controls we use. Note that the $\beta_{x.y}$ coefficients across the two models are indeed comparable on their own and do not require standardization. This follows from both models sharing the same set of predictor x-variables that are all analyzed at the same (team) level. Additionally, both the outcome y-variables, that measure gender diversity, share the same scale ranging from 0 to 100, where 100 indicates having high diversity. To analyze our hypotheses under the Bayesian framework, we first translated each hypothesis into inequality constraints using the above coefficients. The symbolic inequality constraints are presented along with their corresponding hypotheses in Table 3.3.

Table 3.3 Symbolic inequality constraints for Bayes factor estimation

<table>
<tr><th colspan="2">Hypotheses</th><th>Symbolic inequality constraints</th></tr>
<tr><td colspan="2">H1: Having any GDM practice would appear more effective on the current (representation-based) measure, than the new (representation and integration-based) measure</td><td>$\beta_{no.cgd} < \beta_{no.ngd}$
He: $\beta_{no.cgd} = \beta_{no.ngd}$</td></tr>
<tr><td colspan="2">H2: When evaluating based on current (representation-based) measure, GDM practices with an organizational approach appear most effective and those with an individual approach appear least effective</td><td>$\beta_{org.cgd} > \beta_{mgr.cgd} > \beta_{ind.cgd}$</td></tr>
<tr><td colspan="2">H3: When evaluating based on the new (representation and integration-based) measure, GDM practices with a managerial approach appear most effective</td><td>$\beta_{mgr.ngd} > (\beta_{ind.ngd}, \beta_{org.ngd})$</td></tr>
<tr><td>Exploratory Question: Are the three approaches to GDM better than not having any GDM practice?</td><td colspan="2">Compare support for each GDM approach being effective, as against being ineffective, on both the current and new measure.
<table>
<tr><th></th><th>CGD</th><th>NGD</th></tr>
<tr><td>Individual:</td><td>$\beta_{ind.cgd} > \beta_{no.cgd}$
He: $\beta_{ind.cgd} = \beta_{no.cgd}$</td><td>$\beta_{ind.ngd} > \beta_{no.ngd}$
He: $\beta_{ind.ngd} = \beta_{no.ngd}$</td></tr>
</table></td></tr>
</table>

		Managerial:	$\beta_{mgr.cgd} > \beta_{no.cgd}$ $He: \beta mgr.tgd = \beta no.tgd$	$\beta_{mgr.ngd} > \beta_{no.ngd}$ He: $\beta_{mgr.ngd} = \beta_{no.ngd}$	
		Organizational:	$\beta_{org.cgd} > \beta_{no.cgd}$ He: $\beta_{org.cgd} = \beta_{no.cgd}$	$\beta_{org.ngd} > \beta_{no.ngd}$ He: $\beta_{org.ngd} = \beta_{no.ngd}$	

Note that the hypotheses we developed are ordered hypotheses, looking at the relative ordering of predicted GDM effects both within and across the two outcome measures of gender diversity. The traditional frequentist approach is not recommended for such hypotheses as it is seen to be problematic on multiple fronts (Braeken, Mulder and Wood 2015). The null hypothesis significance testing does not allow direct mapping of hypotheses to results and suffers from multiple testing problems that lead to low statistical power. Each of our hypotheses would require numerous significance tests that are not necessarily independent or unambiguous. For example, consider H2: $\beta_{org.tgd} > \beta_{mgr.tgd} > \beta_{ind.tgd}$. First three tests would be needed for each individual effect, then multiple pairwise comparison tests of the effects would have to be conducted to see if they are significantly different from each other and in the expected direction. This would make it very difficult, if not impossible, to combine the resulting set of p values into one single answer to whether our hypothesis is supported by the data. The multiple testing would also require type I error corrections, such as the Bonferroni alpha-corrections, resulting in tests that are very conservative with very little power (Cohen 1992). Furthermore, all traditional structural equation modelling fit indices are also not considered suitable for model comparison, because such ordered hypotheses are not countable in terms of degrees of freedom. Thus these indices reduce to simple deviation measures (of the sample data from the model) and do not result in a reward for formulating a more focused informative hypothesis. The Bayesian method, however, provides a single internally consistent solution for estimation and inference. Thus, we analyze all our hypotheses using Bayesian Informative hypothesis evaluation (BAIN) through the bain package in R.

We evaluate our hypotheses using Bayes factors that quantify to what extent the data supports one hypothesis compared to another. A Bayes factor of 1 would indicate equal support for both hypotheses. A Bayes factor above 1 would indicate preference for the first hypothesis in the comparison and a Bayes factor below 1 would indicate preference for the other hypothesis. For example, a Bayes factor of BF_{12}= 4 indicates that hypothesis 1 received four times more support from the data than hypothesis 2. We evaluate all our Bayes factors using equal prior model probabilities, the default in bain, which means that the hypotheses are seen as equally likely before observing the data.

Note that the Bayes factor is never tied to one hypothesis, rather, it is the relative support for that specific hypothesis compared to another alternative hypothesis. Broadly, we use two types of alternative hypotheses to evaluate the three hypotheses we specified and our exploratory research question. The first type is the complementary hypothesis which lets us compute BF.c: the Bayes factor of a hypothesis against its complement, that is any set of restrictions between the coefficients that is not the said hypothesis. For example the complement of $H1: \beta_{no.tgd} < \beta_{no.ngd}$ would be $H1c: \beta_{no.tgd} \geq \beta_{no.ngd}$. Thus, BF.c can be used to evaluate our three hypotheses.

The second type of alternative hypothesis we use is what we call the equality hypothesis which lets us compute BF.e: the Bayes factor of a hypothesis against the equality constraint, where the coefficients in the said hypothesis are considered equal. For example the equality constraint of $H1\colon \beta_{no.tgd} < \beta_{no.ngd}$ would be $He\colon \beta_{no.tgd} = \beta_{no.ngd}$. We use BF.e to reconfirm H1 and to study our exploratory question, by comparing support for each GDM approach being effective, as against being ineffective.

3.5. RESULTS

We evaluate our three hypotheses by computing their corresponding Bayes factors against their complements. These results are presented in Table 3.4, along with information on their fit and complexity. The first hypothesis saw overwhelming support with BF_1c=222.141, i.e. the support in our data is about 222 times larger for H1 as compared to its complement. This implies that GDM practices are likely to appear more effective when evaluated on the current measure, making it misleading. This suggests that focusing only on representation, as CGD does, may significantly overstate the effectiveness of GDM by failing to capture its unintended consequences on integration.

This was reconfirmed by evaluating whether the two gender diversity measures assess effectiveness differently. Specifically, we estimated the Bayes factor between H1 and the equality hypothesis instead of the complement. This yielded BF_1e=3.998 indicating that H1 received almost four times the support as compared to the equality hypothesis. This implies that the effectiveness of practices on both measures vary such that they appear more effective on the current measure as compared to the new measure.

The second and third hypotheses did not find much support. The second hypothesis did not find any support with BF_2c=0.000, indicating that the complement was supported. This suggests that when evaluating based on CGD, GDM practices with an organizational approach are not necessarily the most effective and those with an individual approach are not necessarily the least effective. The relative effectiveness of practices would differ from what was proposed. The third hypothesis saw BF_3c=1.234, indicating marginal or about equal support in the data for H3 and its complement. This suggests that when evaluating based on NGD, practices with a managerial approach are not necessarily the most effective.

Table 3.4 Results of main hypotheses evaluation using Bayes factor of the hypothesis at hand versus its complement (BF.c)

Hypotheses	BF.c	Fit	Complexity
H1: $\beta_{no.cgd} < \beta_{no.ngd}$	222.141	0.996	0.500
H2: $\beta_{org.cgd} > \beta_{mgr.cgd} > \beta_{ind.cgd}$	0.000	0.000	0.179
H3: $\beta_{mgr.ngd} > (\beta_{ind.ngd}, \beta_{org.ngd})$	1.234	0.389	0.340

Our exploratory analysis examines whether GDM practices are associated with greater gender diversity. As shown in Appendix B, the simple average CGD and NGD scores do not differ significantly between teams with and without GDM practices. This means that teams with GDM did not have significantly higher gender diversity scores on both CGD and NGD.

For a robust evaluation of our exploratory question we use Bayes factors. Specifically, we computed the Bayes factors between the pairs of informative hypotheses indicated in Table 3.3. For each GDM approach we compare the hypothesis of it appearing 'effective'—where having the approach is better than not having any GDM, with the equality hypothesis of it appearing 'ineffective'—where having the approach is as good as not having any GDM. Thus there are six Bayes factors, indicating the effectiveness of the three GDM approaches on the CGD and the NGD. These results are presented in Table 3.5.

In case of the CGD, practices with an individual approach have a Bayes factor of 36.693. This indicates that the support in the data is almost 37 times larger for practices with an individual approach to be effective, as compared to being ineffective. The managerial approach has a Bayes factor of 1.399, indicating weak or no evidence for practices with a managerial approach to be effective, as compared to them being ineffective. This could mean two things. First, these practices might not be beneficial at bringing about equal managerial gender representation, our CGD measure. Alternatively, the unintended consequences of this approach might reduce women's managerial representation. For example, diversity training for managers can sometimes activate rather than reduce bias (Kidder et al. 2004; Steffens et al. 2017). This could hamper their willingness to promote women in higher ranks, making the practice appear less effective on the CGD. In fact Naff and Kellough (2003) found such training programs to bear no significant relationship to the promotion of women in the federal agencies examined.

For the organizational approach the Bayes factor was 0.194. The less than 1 value indicates that there is greater support in the data for practices with an organizational approach to be ineffective. This could again mean two things. First that these practices might not be beneficial at bringing about equal managerial gender representation. Alternatively again, the unintended consequences of this approach might reduce women's managerial representation. Although establishing organizational responsibility through gender quotas could be an effective way to increase women's representation, this does not have to result in more female managers. For example, organizations could set quotas for the organization as a whole, which often increases women's representation in the lower ranks while leaving the higher ranks largely unaltered. As our CGD looks at the managerial representation of women, such practices are likely to appear ineffective on it. This aligns with the findings of Naff and Kellough (2003) where having such practices was not significantly related to the promotion of women.

In case of the NGD, the Bayes factor for all three approaches to GDM is less than 1. This implies all three GDM approaches were likely to appear ineffective[5] on the NGD. As we found some evidence of effectiveness on the CGD, ineffectiveness on the NGD could be explained by it additionally accounting for integration. This means that the benefit from adopting any of the three GDM approaches is likely to be offset by its unintended consequences. This variation in the effectiveness of the measures is also reflected in their regression coefficients and credibility intervals (similar to confidence intervals in the frequentist approach). We illustrate this in Appendix Figure B2.

---

[5] Ineffectiveness for each GDM approach is hypothesized here as being 'equally effective' as not having any GDM, e.g. $\beta_{ind.tgd} = \beta_{no.tgd}$. However, ineffectiveness could also be hypothesized as being 'equally or less effective' as not having any GDM, e.g. $\beta_{ind.tgd} \leq \beta_{no.tgd}$. This formulations gives slightly different results which are presented in Appendix Table B3.

Table 3.5 Results of exploratory analysis using Bayes factors between pairs of informative hypotheses (BFne)

Hypotheses	CGD	NGD
Individual:		
$H_n: \beta_{ind.y} > \beta_{no.y}$, He; $\beta_{ind.y} = \beta_{no.y}$	36.693	0.324
Managerial:		
$H_n: \beta_{mgr.cgd} > \beta_{no.cgd}$; $H_e: \beta_{mgr.cgd} = \beta_{no.cgd}$	1.399	0.297
Organizational:		
$H_n: \beta_{org.cgd} > \beta_{no.cgd}$; $H_e: \beta_{org.cgd} = \beta_{no.cgd}$	0.194	0.205

Robustness Checks

To evaluate the robustness of our results we estimated the above Bayes factors in multiple settings. First, we restricted our analysis to the complete cases only, on the non-imputed data, and obtained similar results. Secondly, as the data included employees from multiple countries and sectors, our results could be driven by a particular country or sector. We checked for this by running the above analysis again while deleting every country and sector once from the analysis, in line with the Jackknife procedure (Quenouille 1949; Tukey 1958). Results were similar to the ones presented above with only marginal variations for the exploratory question. Overall we continue to find evidence of NGD better capturing the negative effects of GDM than CGD, across countries and sectors.

Finally, as our first hypothesis compares NGD with CGD, our results could be driven by the specific CGD we used based on the Herfindahl index. To evaluate this, we repeated the analysis using alternative CGD measures available in the data. These included the proportion of women managers, the proportion of women in the whole organization and the proportion of women in senior management. We also reformulated the latter two using the normalized Herfindahl index, consistent with our original CGD approach. We found similar results with all variants of CGD having a BF_{1c}>1, i.e. the data was more likely to support H1 than its complement. Full results from these robustness checks are presented in Appendix Table B4, Table B5 and Table B6.

3.6. Conclusion

Our findings highlight the need to evaluate GDM practices not only by their impact on women's representation but also by their ability to foster meaningful integration in workplace relationships. Current gender diversity measures focus largely on representation and overlook the unintended consequences of GDM that often undermine employee integration. Using data from 401 teams across nine European countries and multiple sectors, our analysis reveals how such representation-based measures obscure and often overstate the effectiveness of GDM.

To address this gap, we introduced a new gender diversity measure, referred to as NGD, which captures integration directly by examining employees' actual social relationships. This contrasts with current gender diversity measures that focus only on gender ratios or attempt to approximate integration indirectly by measuring the proportion of women in management. Our results show that even such a proxy, referred to as CGD, overstates the impact of GDM compared to NGD. This demonstrates the importance of directly measuring integration to gain a more accurate understanding of GDM's effects.

Our findings also suggest that current representation-focused measures can hide the unintended consequences of GDM. Prior research has shown that simply having GDM practices in place can trigger perceptions of unfairness, feelings of exclusion among men, and doubts about the competence of women. These signals may operate at the same time and weaken trust and cooperation in teams. As a result, GDM may succeed in increasing representation while failing to promote genuine inclusion. This is evident from our observation that all GDM practices performed better at improving representation rather than integration.

We also find that while some GDM practices do improve representation, they generally do not improve integration. Among the three broad approaches we studied—those targeting individuals, managers, and organizations—only the individual approach was clearly effective at improving representation. This is understandable, as practices targeting individuals through say mentoring and networking are designed to help women access leadership positions, which is exactly what the CGD captures. However, even these practices fail to enhance integration.

This has important implications. When women are promoted without being integrated into the workplace, they are more likely to leave. Research shows that women in top management roles leave their positions at nearly twice the rate of men, despite their increased representation (Krishnan 2009). At the same time, organizations benefit from having women in leadership, and their departure is costly (Krishnan and Park 2005; Smith, Smith and Verner 2006). Improving both representation and integration is therefore essential to retain talented women and realize the full benefits of gender diversity. Thus, organizations are recommended to first emphasize goals to retain minority recruits, before focusing on their recruitment (McKay and Avery 2005). Policymakers and regulators, who often promote representation-based goals such as numerical targets, should expand their focus to include integration. Even if the goal is to just improve women's representation, this can only be achieved if they can be integrated into the organization. Thus policymakers need to consider women's integration along with, if not prior to considering their representation.

One reason why we find GDM to generally be ineffective is that organizations may adopt it as a symbolic gesture, aiming to improve their public image or meet external expectations rather than to promote real change. This can prevent GDM from improving even representation. For example, when GDM is present, it may lead employees to view the organization as fair, even if outcomes remain unequal. This perception can reduce awareness of bias and limit accountability, allowing existing inequalities to persist (Brady et al. 2015).

Although we find that none of the three broad GDM approaches improve integration, this does not mean that such efforts are always ineffective. The unintended effects of GDM are likely to depend on how specific practices are designed and implemented. For instance, diversity training that is voluntary has been found to be more effective than training that is mandatory (Dobbin, Schrage and Kalev 2015; Dover, Kaiser and Major 2019). Unfortunately our data does not have such nuanced information. We discuss some additional limitations below and how future work could build on our findings.

Firstly, the three approaches to GDM we evaluated were a broad categorization to ensure robustness while accounting for comparable effects of implementation across multiple countries and sectors. Future research could look at a more narrow specification of the approaches or even evaluate each practice independently. For instance, practitioners could independently evaluate each GDM practice in an organization on its effects on both the representation and integration of women.

Secondly, the causal direction of GDM's effectiveness is difficult to ascertain. Specifically, it can be argued that organizations with poor representation and integration of women might be more likely to adopt GDM practices as a remedy which might make it seem ineffective. Nevertheless, it could also be argued that organizations with better representation and integration of women might be able to assert more agency and advocate for having GDM practices, making the practices seem more effective than they necessarily are. Future research could use longitudinal data to better understand causality. Such data could also provide more insight into the working of practices, including information on the gestation period of practices and how long its effects last. For example, positive effects of diversity training are seen to decline with time (Bezrukova et al. 2016).

Despite these limitations, the new integration-based measure developed in this study offers a valuable tool for assessing GDM. It can be adopted by any organization using digital communication with only modest effort and resources. Following data protection directives, the measure could be adapted and constructed using anonymized information on socialization. For instance, the gender composition of employee networks could be got from their anonymized exchanges on email or other internal communication platforms. Thus, organizations could leverage existing data to make their GDM practices more cost effective. Additionally, the ability to account for integration could also help organizations reap the benefits of diversity and improve organizational outcomes such as productivity and innovation.

Overall, our findings point to the need for managers and policymakers to evaluate GDM practices based on their impact on both representation and integration. This dual focus is crucial for understanding the slow progress in reducing workplace gender inequalities (Cullen and Perez-Truglia 2019; Stainback and Tomaskovic-Devey 2012) and informing course corrections for the same. By capturing integration, the NGD measure provides a more accurate assessment of GDM effectiveness by accounting for unintended consequences. Used alongside representation-based measures, it offers a practical tool to identify best practices and better equip policymakers in managing gender diversity.

CHAPTER 4

Studying Interaction of Cumulative (Dis)Advantage Factors: What Have We Done?!

Abstract

Despite cumulative advantage/disadvantage (CAD) being well known to widen inequality, making the rich richer and the poor poorer, we know little about how interactions between CAD factors like wealth and gender shape inequality across intersecting groups like rich men, rich women, poor men, and poor women. This gap is critical. Unlike single CAD mechanisms that widen inequality, CAD interactions can also reduce inequality. Yet research remains fragmented across disciplines, rooted in single CAD mechanisms, and hindered by the complexities of studying interaction effects. We conduct a general examination of how CAD interactions are studied, identify pitfalls, and propose solutions using simulation analysis. We find conventional step-wise regression to result in CAD effects being alarmingly misinterpreted, with Type-I and Type-II errors reaching 47% and 68%, respectively. To address this, we offer practical guidelines for theorizing, analyzing, and interpreting CAD interactions to better inform both inequality research and policy.

The authors jointly developed the core ideas of this chapter. **Sanjana Singh** wrote the main part of the manuscript and conducted the analyses. **Vincent Buskens** and **Eva Jaspers** substantially contributed to the manuscript.

4.1. INTRODUCTION

"Rich get richer and poor get poorer" is one of many colloquial sayings used to describe cumulative advantage/disadvantage (CAD)[6]. Such sayings show CAD's widespread use and understanding as a mechanism for inequality. Formally, CAD is the systemic tendency for *interindividual divergence* in a given resource with time. Interindividual divergence implies that CAD is not an individual but a group property. For example, "rich get richer" (cumulative advantage - CA) and "poor get poorer" (cumulative disadvantage - CD) describe wealth divergence between the initially rich and poor.

Interindividual divergence makes CAD vital for questions of fairness in the distribution of opportunities and resources (Dannefer 2003; DiPrete and Eirich 2006). Thus, CAD is used across scientific disciplines and levels of policymaking, taking several analogous forms. For instance, in sociology, psychology, and economics, the CAD process is known as the reputational effect (Gould 2002), the halo effect (Greenwald and Banaji 1995), attachment theory (Bowlby 2005), and the scarring effects of unemployment (Ellwood 1982). In organization and network research, CAD takes the form of first-mover advantage (Kerin, Varadarajan, and Peterson 1992) and Gibrat's law of proportional effect (Sutton 1997). In biology, it is the multiplier effect (Wilson 2000).

Our understanding of the CAD mechanism is based on a single CAD factor dividing actors into two—those with CA or CD—causing resource divergence between them. However, actors often have multiple CAD factors, creating more than two groups and complicating this simple division. For example, an actor from a poor low-status household (L) who succeeds at work (S) would experience two CAD mechanisms: "poor get poorer" and "success breeds success". Independently, these mechanisms suggest an actor from the SL group to have a CA compared to an unsuccessful peer (U) but a CD compared to someone from a rich high-status household (H). However, the CAD interaction mechanism remains unclear, with no information on resource divergence across the four groups: SL, UH, UL, and SH. This is crucial for inequality research and policymakers as traditional approaches cannot determine, for instance, which two groups have the largest divergence, whether SL benefits enough from S to surpass UH, or which group is the most disadvantaged? We further highlight the importance of CAD interactions in the next section.

This chapter examines binary CAD interactions through a four-group example. We first illustrate how CAD effects are typically estimated using step-wise regression. By considering all possible comparisons among the four groups, we demonstrate how there are seven CAD effects and two complementary approaches to estimate them: regression coefficients and marginal means comparison. Using this illustration as a foundation, we examine how CAD interactions are studied, identify pitfalls, and propose solutions in three phases: theorizing, analyzing, and interpreting.

We identified two unresolved pitfalls: selecting the best analysis approach and cautiously interpreting non-significant results. Both issues require understanding scenarios where CAD estimation errors are minimized, which requires rarely available true population data. We use simulation analysis to address this, generating diverse datasets with predefined

[6] A reference list of acronyms and shorthand is provided in Appendix C2.

population characteristics. This allows us to compare CAD estimates and detect Type-I errors, i.e., detecting non-existent CAD effects or false positives, and Type-II errors, i.e., missing existing CAD effects or false negatives.

We found current approaches to alarmingly neglect or misinterpret CAD effects. Type-II errors for interaction effects (IE) reached 68%, prompting reanalysis without the IE in most cases. This step-wise regression practice inflated MEs' Type-I errors from the standard 5% to a staggering 47%.

To address this, we develop guidelines for theorizing, analyzing, and interpreting CAD interactions. While we focus on binary CAD interactions, the findings have broader implications for studying both CAD and IEs. We also contribute to the literature on experimental design and the pitfall of focusing on statistical significance. Our four-group example is equivalent to the two-by-two factorial experiment, where step-wise regression drops non-significant interactions to improve power and cost-efficiency. However, as IEs are often undetected, such data-dependent model selection leads to invalid inferences, as detailed by Muralidharan et al. (2023).

4.2. CAD INTERACTIONS ARE CRITICAL FOR (IN)EQUALITY

Recent CAD and intersectionality research illustrates the frequent presence of CAD interactions and the need for understanding them, particularly for questions of equality (e.g. Bauer et al. 2021; Cho et al. 2013; Erola and Kilpi-Jakonen 2017; Ross and Mirowsky 2006; Stienstra et al. 2021). For example, despite CAD being established to create inequality, Bernardi (2014) and Pedulla (2018) show how CAD interactions can alleviate inequality for certain groups.

Bernardi (2014) describes the CAD interaction as a 'compensatory advantage' process in the educational inequality context. He builds on the idea of path-dependence in life course trajectories, which parallels the idea of CAD itself. They theorize and analyze initial success (S) as compared to being unsuccessful (U) to have a greater CA for actors from disadvantaged families (L), than for actors from advantaged families (H), i.e. $L(S - U) > H(S - U)$. They also reformulate this to show social background inequality ($H - L$) to be larger among U than S, i.e. $U(H - L) > S(H - L)$. This four-group example's compensatory advantage hypothesis can be formally represented as: $SL - UL > SH - UH$ or $UH - UL > SH - SL$.

Pedulla (2018) describes the CAD interaction of race and unemployment as a 'muted congruence' process in shaping labor market inequality. He builds on status-based theories of discrimination and impression formation to argue that the consequences of unemployment may be more limited for African Americans than for whites. As employers tend to have overlapping stereotypes on both black applicants and applicants with unemployment, e.g., low levels of competence and poor work ethics, employers may find black applicants' unemployment information to be redundant. Thus, black workers (L) would have lower CD ($U - S$) from unemployment than white workers (H), i.e. $L(U - S) > H(U - S)$. He also extends this line of thought to white workers (H), whose CA from continuous employment ($S - U$), would be higher than that of black workers' (L). Thus, our four-group example's muted congruence hypothesis is: $SH - UH > SL - UL$.

The CAD interaction hypotheses of both Bernardi (2014) and Pedulla (2018) include all four groups in a single comparison, addressing the two-group limitation of traditional CAD theories. However, their approach still does not explain how resources diverge among the four groups. For instance, these hypotheses cannot be used to learn which groups have the highest CAD, or whether all groups have significantly different outcomes. DiPrete and Eirich, (2006) highlight these challenges in their CAD review and find "the frequent lack of clarity in models, mechanisms, and tests… a continuing issue in the sociological literature on CA processes as potential generators of inequality. This lack of clarity can produce incorrect specifications, incorrect estimates, and incorrect interpretations."

Despite CAD interactions' importance for understanding inequality, our knowledge remains incomplete and fragmented across disciplines and estimation practices might misdirect conclusions. To address this, we begin by describing how CAD effects are estimated in our four-group example , which includes the two most common CAD forms—path-dependent and status-dependent. This example is broadly applicable across fields, as demonstrated with the works of Bernardi (2014) and Pedulla (2018) in education and employment.

4.3. Foundational Illustration: CAD Effect Estimation through Seven Group Differences

Like most interaction studies, CAD interactions are typically analyzed using step-wise regression or regressions with interactions. This method is widely used in intersectionality research (Bauer et al. 2021) and was applied by Bernardi (2014) and Pedulla (2018). This section demonstrates how step-wise regression applies to our four-group example, estimating CAD effects through two complementary approaches: regression coefficients and marginal means comparison. We illustrate how seven group differences capture all CAD effects and can be visualized in a single resource distribution graph. This illustration provides a benchmark for comprehensively studying CAD interactions, which informs the three-phased examination in the next section.

Step-wise Regression

Regression with interaction is the first step of a step-wise analysis, which can be symbolically expressed for our four groups through the linear regression equation:

$$Outcome_i = \beta_0 + \beta_1 PathCAD_i + \beta_2 StatusCAD_i + \beta_3 PathCAD_i \times StatusCAD_i + \varepsilon_i,$$

$$PathCAD_i = \begin{cases} 0 \text{ if } Unsuccessful\ (U) \\ 1 \text{ if } Successful\ (S) \end{cases},$$

$$StatusCAD_i = \begin{cases} 0\ if\ Low\ (L) \\ 1\ if\ High\ (H) \end{cases}, \qquad \text{(i)}$$

where β_0 is the intercept and ε_i is the error term. $PathCAD_i$ and $StatusCAD_i$ are agent i's binary CAD variables indicating their initial success (S vs U) and time-invariant status (H vs L), respectively. The $PathCAD_i{}'\ StatusCAD_i$ forms a two-by-two interaction that takes the value of 1 if an agent i has a high status and is successful, i.e. belongs to the combination group SH. For all other combination groups, it takes the value of 0, given the above binary codes for the $PathCAD_i$ and $StatusCAD_i$. This is often called the 'long' or 'full model' (Muralidharan et al. 2023).

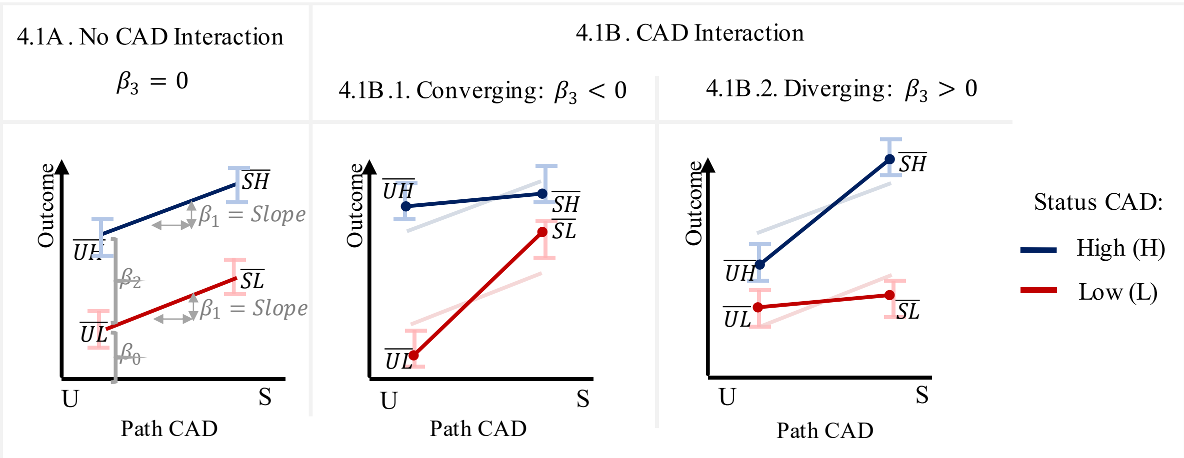
The "step-wise" aspect follows from first testing the interaction's significance. If it is non-significant, i.e. the hypothesis $\beta_3 = 0$ cannot be rejected, the model is often simplified by removing the interaction to more powerfully estimate the main variables in the 'short model':

$$Outcome_i = \beta_0 + \beta_1 PathCAD_i + \beta_2 StatusCAD_i + \varepsilon_i. \qquad \text{(ii)}$$

Thus, the step-wise analysis lets us simultaneously study both CAD variables even in the absence of an interaction. This approach also follows the natural hierarchy of modeling strategies and fits the conditional expectation function (Angrist and Pischke 2009). Moreover, this is the most commonly recommended and practiced approach to study interactions across disciplines (e.g. Engqvist 2005; Kahan 2013).

Both regression models can also be used to visualize the CAD mechanism. They allow estimation of the four group means, which can then be graphed to take one of the three forms in Figure 4.1. They all show the same relationship: outcome on the y-axis, path CAD on the x-axis, and status CAD as a grouping variable. The lines thus depict the relation between agents' CAD variables and their outcome – the CAD effect. Specifically, a positive slope would indicate path CAD, i.e. U's expected future outcomes would be lower than S, and non-overlapping grouping lines would indicate status CAD, i.e. L's expected outcomes would be lower than H's. As we simultaneously study two CAD variables, all three graphs have positively-sloped non-overlapping lines, i.e. $\beta_1 > 0$ and $\beta_2 > 0$.

Figure 4.1 Possible relations between actors' two binary CAD-inducing characteristics and their outcome. The first CAD variable is presented on the x-axis and can be categorized as Successful (S) or Unsuccessful (U). The second CAD variable is presented as a grouping variable and can be categorized as High (H) or Low (L). The outcome is presented on the y-axis. The two CAD variables can either have a non-significant (A) or a significant interaction (B) effect on agents' outcomes.



Seven Possible CAD Effect Estimates

As the CAD mechanism refers to the divergence of resources among all groups, the CAD effect can be estimated using group differences. There can be seven possible comparisons for our four groups—six direct pairwise comparisons (PC) from the four groups and one indirect difference-in-difference comparison (DiD). These can be estimated using regression equation (i) where:

1. $\overline{SL} - \overline{UL} = \beta_1$
2. $\overline{UH} - \overline{UL} = \beta_2$
3. $\overline{SH} - \overline{UL} = \beta_1 + \beta_2 + \beta_3$
4. $\overline{SL} - \overline{UH} = \beta_1 - \beta_2$
5. $\overline{SH} - \overline{UH} = \beta_1 + \beta_3$
6. $\overline{SH} - \overline{SL} = \beta_2 + \beta_3$
7. $[\overline{SH} - \overline{UH}] - [\overline{SL} - \overline{UL}] = [\overline{SH} - \overline{SL}] - [\overline{UH} - \overline{UL}] = \beta_3$.

_______________(iii)

These estimates can also be applied to the short model in equation (ii) by simply omitting β_3.

The seven group differences reflect all possible CAD estimates. This follows from CAD effects being solely determined by the four group averages, and not any other group superset like $\overline{S}, \overline{H}, \overline{U}, \overline{L}$, or even the grand mean of all actors, say $\overline{SHUL}$. When using a regression with interaction, the effect of group supersets cannot be meaningfully studied in isolation as they may change based on the values of another variable. For instance, the CAD effect of status cannot be estimated as $\overline{H} - \overline{L}$, as the difference between group H's and L's outcomes will vary according to their initial success.

The seven CAD estimates are inter-related, with all seven being a function of the three regression coefficients β_1, β_2 and β_3. This means that the seven effects should not be considered as independent estimations and a comprehensive assessment must consider all seven estimates. We next show how these seven effects can be estimated using two complementary approaches: regression coefficients and marginal means comparison.

Alternative and Complementary CAD Effect Estimations

Regression coefficients directly disentangle CAD's main and interaction effects

The regression coefficients β_1, β_2 and β_3 directly present three of the seven group differences outlined in equation-set (iii), the path CAD's main effect (ME), status CAD's ME, and the IE of both. Empirically, β_1 is the ME of success as it gives the S vs U difference in outcome when $StatusCAD_i = 0$. Similarly, β_2 gives the status CAD's ME for U, and the IE β_3 gives the additional difference in outcome between UL and SH.

The significance of the MEs and the IE can be assessed as part of the regression output, using their corresponding confidence interval (CI), p-value or t-score, typically at a 5% significance level or a Type-I error rate. Thus if β_3 is non-significant, the step-wise analysis focuses on estimating CAD's MEs with the short model.

A significant value of $\beta_1 > 0$ would suggest a path CAD, where i's outcome expectation increases by β_1 if they have S, as compared to U, i.e. S would have a CA and U would have a CD. Similarly, $\beta_2 > 0$ would suggest a status CAD effect, where $H(L)$ has a CA(CD) of β_2. Figure 4.1 separately illustrates the step-wise approach with equations (ii) and (i) being visualized in panels 1A and 1B, respectively.

Given the critical role of β_3, most CAD interaction studies focus on theorizing and testing hypotheses on it or the IE. For example the $\beta_3 < 0$ hypothesis of panel 1B.1 can be rewritten as $SL - UL > SH - UH$, while the $\beta_3 > 0$ hypothesis of panel 1B.2 can be rewritten as

$SH - UH > SL - UL$. These are indeed the hypotheses of Bernardi (2014) and Pedulla (2018) we formalized earlier.

Erola and Kilpi-Jakonen (2017) provide a detailed theoretical account of how patterns like those in Figure 4.1 can follow from different CAD interaction mechanisms. They use CAD from different familial resources to illustrate the creation of children's outcome inequality. For instance, the mechanisms in panels 1B.1 and 1B.2 are termed 'compensatory accumulation' and 'multiplicative accumulation', respectively. The former parallels Bernardi's (2014) 'compensatory advantage' mechanism and the latter parallels Pedulla's (2018) 'muted congruence' mechanism.

While this popular approach provides a straightforward way to estimate, test, and disentangle CAD MEs and the IE, it only provides three of the seven possible CAD estimates in equation-set (iii):

1. $\overline{SL} - \overline{UL} = \beta_1$
2. $\overline{UH} - \overline{UL} = \beta_2$
7. $[\overline{SH} - \overline{UH}] - [\overline{SL} - \overline{UL}] = [\overline{SH} - \overline{SL}] - [\overline{UH} - \overline{UL}] = \beta_3$

(iv)

Thus, on its own, this approach offers an incomplete assessment of the CAD mechanism.

Marginal means comparison complements and offers an alternative

Researchers interested in specific group comparisons often conduct posthoc marginal means comparisons (e.g. Herbaut 2021; Kleinert and Mochkabadi 2022). For instance, both Bernardi (2014) and Pedulla (2018) used marginal means in addition to the regression coefficient analysis of β_3 to focus on specific group comparisons.

The approach takes different forms and names across disciplines, depending on the research question. One of the most widely practiced approaches is the two-step average adjusted prediction, the default in many analytical software tools, such as the `margins` command in STATA and the `marginaleffects` package in R and Python (Arel-Bundock, Greifer, and Heiss 2024). Specifically, the approach involves two steps:

Step 1: Create a new counterfactual dataset using the original regressor values, but set specific regressors to the values of interest.

Step 2: Average the predicted values from this new dataset.

This would mean computing $\overline{SH}$ by first creating a new data where values of $PathCAD_i$ and $StatusCAD_i$ are fixed at 1 for all actors or observations in the data. This new data will be used alongside the original regression results to predict the outcome of each actor. All actors' average outcome thus estimated would be $\overline{SH}$. This method could similarly be used to estimate all seven group comparisons in equation-set (iii) through six direct PCs, and one indirect DiD comparison. This includes the MEs and IEs which are directly available through the regression coefficients, as in equation-set (iv). This approach also enables significance testing for all estimates using standard errors from the new datasets.

Thus, the marginal means approach complements and offers an alternative to the coefficient approach for estimating and testing all seven CAD effects.

4.4. THREE-PHASED CAD INTERACTION: GOALS, PITFALLS AND RESOLUTIONS

This section builds on the foundational illustration to examine CAD interaction studies in three phases: theorizing, analyzing, and interpreting findings. These studies are prone to systematic errors due to challenges in examining both the CAD mechanism and IEs. However, the nascency of CAD interaction research and its resulting lack of formalized research have often caused these errors to be overlooked. We outline each phase's goal, key pitfalls, and propose resolutions or requirements, which are addressed in subsequent sections through simulation analysis.

Phase 1. Theorize: Visualize Interaction of All Groups

Goal: Hypothesize the divergence of resources among all CAD interaction groups.
Pitfall: Focusing on specific group comparisons.
Resolution: The precise figure should be clear from the theory, i.e. all possible group differences must be hypothesized, or at least the theory should be explicit if there is no specific hypothesis on some group differences.

CAD interaction theories[7] follow from traditional CAD theorizing and the broader study of IEs. However, these fail to capture resource divergence among all groups, i.e., the seven comparisons in equation-set (iii).

Traditionally, the CAD mechanism was theorized to be kickstarted by signals. Merton (1968) observed that CAD results from ambiguity in reward systems where performance is hard to measure (Pradhan and Jena 2017). As a result, evaluations rely on signals that are more readily visible and comparable across actors, which kickstarts the CAD mechanism. As signals are inherently positive or negative, they create interindividual divergence by providing information on whether a group has an advantage over another. Thus, signals are the core of any CAD mechanism and drive inequality in outcomes. Broadly, CAD mechanisms can be kickstarted signals from either path measures or status characteristics. These signals determine the CAD form: path-dependent, status-dependent, and multi-path-status-dependent, which are detailed in the online supplement. While the first two involve single CAD mechanisms, the third introduces CAD interactions, extending single CAD theories to account for two CADs simultaneously.

CAD interaction theorizing follows the step-wise approach in beginning with the IE. Typically, researchers argue for an IE between two pre-established CAD variables. This means that the two variables can be defined, without loss of generality, in a way that they are theorized to have a positive ME. With this backdrop, a positive or negative IE, i.e. $\beta_3 > 0$ or $\beta_3 < 0$, is first hypothesized.

For instance Bernardi (2014) theorized the 'compensatory advantage' mechanism by hypothesizing $\beta_3 < 0$. Similarly Pedulla (2018) theorized the 'muted congruence' mechanism with $\beta_3 > 0$. This approach allows researchers to efficiently focus on the IE,

[7] We use 'theory' to describe general theoretical arguments used to explain a phenomenon. Theories provide a broader scope, explaining overarching principles, and are used to derive hypotheses. Hypotheses, in turn, focus on specific, testable arguments to evaluate the theory.

while implicitly forming expectations on both the IE and MEs. However, the IE and MEs only inform three of the seven group comparisons.

Thus, the pitfall would be to form an incomplete theory that does not consider all seven possible group comparisons. A comprehensive CAD theory must inform on resource divergence among the four groups and be visualized as in Figure 4.1.

This pitfall can be seen by graphing Bernardi's (2014) 'compensatory advantage' hypothesis. Following the step-wise regression norm, the mechanism is hypothesized as $\beta_3 < 0$. If this provides a comprehensive CAD assessment, we should be able to look at how resources diverge among the four groups by graphing the hypothesis. One might argue that this can be done as in Figure 4.1B.1, as β_3 = IE informs the intersection pattern of the lines.

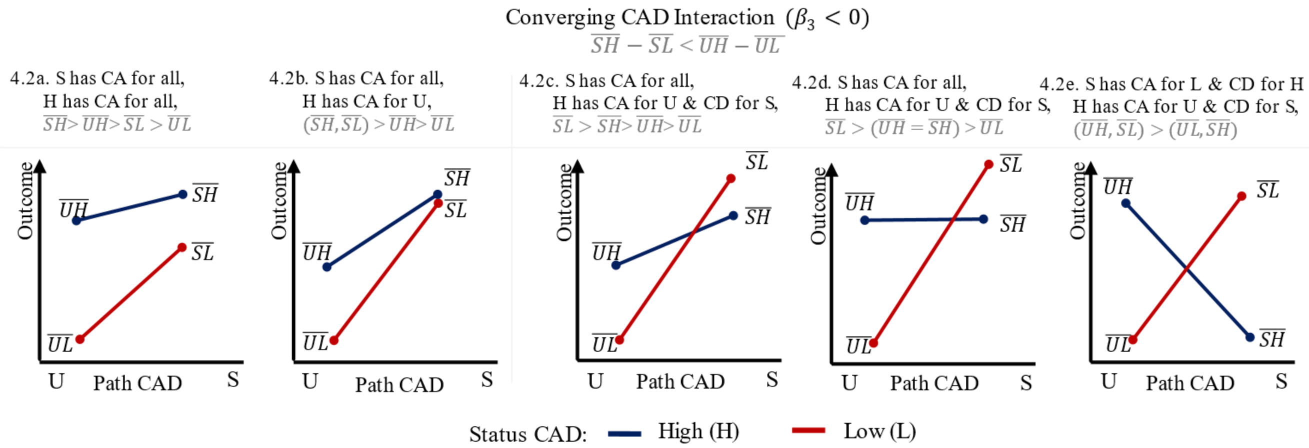
However, $\beta_3 < 0$ need not result in Figure 4.1B.1. We illustrate this using Figure 4.2 where all graphs have $\beta_3 < 0$, but are visualized differently—representing distinct CAD mechanisms.

For example, while Graphs 2a and 2b of Figure 4.2 have similar converging patterns, their varying group differences indicate different mechanisms. In Graph 2a status and path CAD can be clearly seen across all groups, with $\overline{SL}$ closest to $\overline{UH}$ showing only a small relative CD. This suggests a mechanism where status creates strong disparities, which success can help reduce but never overcome. In contrast, Graph 2b shows $\overline{SL}$ having substantial CA over $\overline{UH}$, being closest or even identical to the most advantaged group $\overline{SH}$. This suggests a mechanism where success effectively compensates disadvantages from a lower status, leading to more equitable outcomes.

Likewise, while Graphs 2c, 2d, and 2e are similar in having an intersecting pattern, they present different mechanisms. This can be seen through their varying path CAD among H, i.e. slope of their blue lines. Graph 2c has a clear path CAD where $\overline{SH} > \overline{UH}$, however, this advantage disappears in Graph 2d with $\overline{SH} = \overline{UH}$, and reverses in Graph 2e with $\overline{SH} < \overline{UH}$.

Thus, the typical step-wise interaction approach cannot provide a comprehensive understanding of the CAD mechanism. While the IE hypothesis is necessary for the graph, it is insufficient. Additional information on the MEs and other group differences is required. This is evident from Bernardi's (2014) posthoc analysis, which revealed $\overline{SH} = \overline{UH}$ and $\overline{SL} > \overline{UL}$, suggesting a pattern more like Figure 4.2d than Figure 4.1B.1.

Figure 4.2 Alternative forms of converging interaction and their different CAD mechanisms



Ideally, the theory should be transparent on how resources are expected to diverge across all interaction groups. This means that any theoretical analysis should:

1. First argue for a positive or negative IE or the DiD
2. Be transparent about what the IE argumentation implies for the MEs and other group differences. This would involve explicitly stating expectations on:
    a. Which groups significantly differ from each other;
    b. Direction of the expected group difference;
    c. Which group differences are unclear—allowing for exploratory analysis.

A simple way to check for this and avoid the theoretical pitfall is to illustrate all theorized hypotheses in a single graph, such as one of the graphs in Figure 4.2. This approach brings forth a well-established analytical practice to the theory stage. Graphing estimated group means along with their CIs is considered one of the most effective ways to study interaction effects (Finsaas and Goldstein 2021; Garofalo et al. 2022). Similarly, theories on single CAD have visualized resource divergence between two groups using fanning or K-graphs (e.g. Cole and Singer 1991), offering a useful foundation for applying this visual approach to CAD interaction research.

Pedulla's (2018) 'muted congruence' theory did explicitly consider additional hypotheses alongside $\beta_3 < 0$, however, the lack of a figure makes it unclear whether the theory was comprehensive. They had additional expectations for four PCs:

- "black workers [L] will face severe racial discrimination" i.e. $SH > SL$ and $UH > UL$;
- "White workers who maintain seamless, continuous employment trajectories [SH] are likely to benefit compared to their white counterparts who experience long-term unemployment [UH].", i.e. $SH > UH$;
- "However, for African Americans, a continuous employment trajectory [SL] may lead to similar responses from employers as having a spell of long-term unemployment [UL].", i.e. $SL = UL$.

However, there were no explicit predictions on two of the six PCs, $SH - UL$ and $SL - UH$, leaving the CAD mechanism or resource divergence among the groups unknown. Thus, theorizing with the interaction graph would ensure accounting for all seven CAD effects and their direction.

Phase 2. Analyze: ME, IE, and Pairwise Comparisons

Goal: Visualize and evaluate the theorized CAD mechanism.
Pitfall: Incomplete or incompatible analysis.
Resolution: Analyze all seven comparisons. Use unadjusted PCs for the four estimable only by the marginal means approach.
Required for Resolution: Which approach minimizes Type-I and Type-II errors in estimating the three overlapping CAD effects, i.e., MEs and IE?

The theory can only be evaluated to its full extent by testing hypotheses on all comparisons in equation-set (iii). Such an analysis would allow reconstructing and comparing the theorized graph by estimating the MEs, IE or DiD, PCs and testing their significance. If another figure emerges, it would imply that some of the hypothesized mechanisms could not be confirmed. This could occur even if the ME and IE are significant in the expected direction, as in Figure 4.2a and 4.2b.

Empirically, the graph would require estimating the four group averages, which can then be plotted. Both the coefficients and marginal means can be used to get these four estimates. The marginal means can estimate the four group averages as described earlier for $\overline{SH}$, along with their CIs. The step-wise approach can be implemented as below:

1. $\overline{SH} = \beta_0 + \beta_1 + \beta_2 + \beta_3$
2. $\overline{SL} = \beta_0 + \beta_1$
3. $\overline{UH} = \beta_0 + \beta_2$
4. $\overline{UL} = \beta_0$

The pitfall lies in overlooking the research question and following conventional IE analysis. Using coefficients from the step-wise analysis could result in considering only three of the seven CAD comparisons and/or a focus on the IE $= \beta_3$. For instance, despite muted congruence having a comprehensive theory, follow-on studies limit its analysis to just the IE or, at best, only some of the group comparisons (e.g. Cerda-Jara, Harding, and The Underground Scholars Research Cohort 2024; Veit and Thijsen 2021).

One might imagine that the marginal means approach could resolve this when all seven comparisons are considered. However, these comparisons can create the same pitfall. Conventionally, this approach evaluates the interaction using six PCs with *p*-values adjusted to account for multiple comparisons using, e.g., the Bonferroni correction. This is because performing several comparisons between all possible pairs of groups or conditions drastically increases the Type-I error or the probability of finding a statistically significant comparison just by chance. Methods such as the Bonferroni correction help control this with a more stringent threshold (α-level) for statistical significance (Garofalo et al. 2022). However, as the ideal theory has separate hypotheses for all seven marginal means comparisons, there is no need to adjust the *p*-value.

Thus, the conventional application of both approaches cannot only lead to incomplete or incompatible analysis, but can also give incorrect results depending on the *p*-value adjustments.

Ideally, researchers should attain unbiased and efficient estimates for all seven comparisons in (iii). The marginal means approach can facilitate this by not adjusting for the *p*-values. However, the regression coefficients also directly estimate three of the seven comparisons. This means that while the recommendation to use unadjusted marginal means is clear for four of the seven comparisons, we do not know which approach would be better for the three overlapping comparisons, i.e. MEs and IE. While both approaches are known to yield similar estimates, their CIs and errors differ. We would require information on which approach minimizes both Type-I and Type-II errors for the MEs and IE.

Phase 3. Interpret: Consider Data & Power when Non-Significant

Goal: Cautiously interpret estimated results.
Pitfall: High Type-II error for IE and inflated Type-I error for MEs.
Required for Resolution: How do undetected IEs and falsely detected MEs relate to the data, power, approach, and estimated graph?

IEs are notoriously difficult to detect despite strong theoretical reasons for expecting them and established methods for their identification. This failure to detect IE or Type-II error is driven by several statistical and methodological artifacts.[8] While increasing statistical power can address some issues, IEs face constraints that cannot be resolved by power alone (Rogers 2002). This issue is exacerbated for CAD interactions, which often have small effect sizes, unequal group sizes, binary variables and observational data—all requiring higher power. These factors make it critical to consider power implications when interpreting CAD interactions.

This pitfall of biased CAD effects is worsened by the widespread use of step-wise regression, especially in experimental designs (Kahan 2013). While this can increase power for detecting MEs when no IE exists, it often leads to incorrect inferences otherwise. The prevalence and cost of this trade-off were seen in a recent study where over 70% of the experiments published in top economics journals dropped the IE and used the reduced model. Including the IE considerably affected MEs: 26% of estimates changed sign, and 53% lost significance at the 5% level (Muralidharan et al. 2023). Thus, CAD researchers must carefully interpret both MEs and IE, particularly when they find non-significant results.

If a CAD interaction researcher finds a figure resembling the theory but with some non-significant differences, they must avoid dismissing the effects due to potential Type-II errors. The estimated figure should be interpreted alongside the non-significant results. For example, if a non-significant IE is found alongside a graph with non-parallel lines, as in Figure 4.2a, it could indicate a statistically undetected IE rather than no IE. Researchers must carefully reflect on the data and power before concluding on the absence of an IE, particularly given the additional challenges IE poses to power analysis (Baranger et al. 2023).

Unfortunately, the data and power characteristics that increase the likelihood of undetected CAD effects remain unclear. While estimated graphs can help with interpretation, any concrete inference on a (un)detected effect would require true population information, which is typically unavailable.

Addressing Unresolved Pitfalls

The unresolved pitfalls of Phases 2 and 3 require Type-I and Type-II error information, which relies on true population data—rarely available to researchers. To address this, we use simulation analysis with our generalizable four-group example to provide evidence-based guidance on how to choose the correct analysis approach and interpret the results.

Simulation analysis allows us to study errors across scenarios by constructing synthetic populations with predefined characteristics. We created synthetic datasets with 441 unique combinations of ME, IE, and group size characteristics, each simulated 1,000 times. This enables us to study how these characteristics influence errors in CAD effect detection and find out which approach minimizes errors and how these errors affect interpretation based

[8] For example: small sample size (Alexander & DeShon, 1994), measurement error (Busemeyer & Jones, 1983), small population effect sizes (Stone-Romero & Anderson, 1994), range restriction (Aguinis & Stone-Romero, 1997), heterogeneity of error variance (Aguinis & Pierce, 1998), field studies (McClelland & Judd, 1993), and unequal sample sizes across IE-based subgroups (Stone-Romero, Alliger, & Aguinis, 1994)

on their relation with different characteristics in situations in which MEs and the IE do and do not exist (cf. Smith et al. 2002).

4.5. THE SIMULATION: WHAT IS THE BEST WAY TO STUDY CAD EFFECTS AND WHEN DO THEY GO UNDETECTED?

The IE literature has used simulations as a tool to identify systematic errors plaguing IE as they offer a convenient yet robust way to evaluate Type-I and Type-II errors (e.g. McClelland and Judd 1993; Yzerbyt, Muller, and Judd 2004). We use this established method to identify when CAD effects go undetected by varying the size of the four CAD interaction groups and the main regression parameters in equation (i). This follows from the critical role of group size in power analysis, and the foundational illustration's equation-set (iii), which showed how all CAD effects can be estimated as a function of the three regression coefficients.

To simplify nomenclature, the simulation study uses the following standardized form of the regression equation (i):

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + \varepsilon_i,$$

$$x_{1i} = PathCAD_i = \begin{cases} 0 \text{ if } Unsuccessful\ (U) \\ 1 \text{ if } Successful\ (S) \end{cases},$$

$$x_{2i} = StatusCAD_i = \begin{cases} 0\ if\ Low\ (L) \\ 1\ if\ High\ (H) \end{cases}, \tag{v}$$

where y is the dependent variable, x_1 and x_2 are the two independent variables, i.e. our two CAD variables such that they have a ME of β_1 and β_2, respectively, and an IE of β_3. The error term ε_i is normally distributed with a standard deviation (SD) of 1.

Simulation Characteristics: Group Size and Regression Coefficients

Below, we explain the 441 unique combinations of characteristics we use that follow from 9 relative group size combinations, 7 ME, and 7 IE variations. These are detailed in the online supplement, along with the true population CAD effects used to identify estimation errors using both regression coefficients and marginal means approaches. We briefly specify these variations in group sizes and the regression coefficient characteristics below.

i. Relative group sizes: nine size combinations or three levels of homogeneity

We considered nine combinations of relative group sizes that resulted in three different levels of homogeneity, as shown in the table below. These combinations are based on the 'CAD Main Group Sizes' and use three proportions of $x_1 = 0$ and $x_2 = 0$: 20%, 50%, and 80%. So, for example, we distinguish between the successful being 20%, 50%, and 80% of the population and the same for the status dimension. Note that we only vary these dimensions independently in the simulations, which does not mean we think these dimensions vary independently, but it gives us nine types of populations with different levels of heterogeneity and different sizes of advantaged groups that we feel are important to cover.

Table 4.1 Nine variations of x_1x_2 group sizes constructed based on the proportion of $x_1 = 0$ and $x_2 = 0$, presented alongside their three homogeneity levels estimated using NHHI.

Combi. No.	Relative group Size								Homogeneity	
	CAD Main Group Sizes				CAD Interaction Group Size for x_1x_2					
	x_1=0 U	x_2=0 L	x_1=1 S	x_2=1 H	00 UL	01 UH	10 SL	11 SH	NHHI	Level
1	20%	20%	80%	80%	4%	16%	16%	64%	0.283	High
2	50%	20%	50%	80%	10%	40%	10%	40%	0.120	Low
3	80%	20%	20%	80%	16%	64%	4%	16%	0.283	High
4	20%	50%	80%	50%	10%	10%	40%	40%	0.120	Low
5	50%	50%	50%	50%	25%	25%	25%	25%	0.000	Heterogeneous
6	80%	50%	20%	50%	40%	40%	10%	10%	0.120	Low
7	20%	80%	80%	20%	16%	4%	64%	16%	0.283	High
8	50%	80%	50%	20%	40%	10%	40%	10%	0.120	Low
9	80%	80%	20%	20%	64%	16%	16%	4%	0.283	High

The resulting homogeneity levels are quantified using the Normalized Herfindahl–Hirschman index (NHHI) (Cracau and Lima 2016):

$$NHHI = \frac{\sum_{i=1}^{4} p_i^2 - \frac{1}{4}}{1 - \frac{1}{4}}$$

where p_i is the proportion of people in group i. NHHI ranges from 0 (heterogeneous) to 1 (homogenous). For instance, Combi. No. 5 has a 0 score as each group has an equal 25% share, which means that the populations could not be better spread over the groups. The HHI is a common measure for concentration or homogeneity. Given symmetry in the x_1x_2 size combinations, there are only three unique homogeneity values, which we categorize as high homogeneity, low homogeneity, and heterogeneous. Additionally, we vary inequality by randomly assigning outcomes to the four groups to vary the association between groups' minority(majority) size and their (dis)advantage.

ii. Regression coefficients: 49 combinations from size, valence, and presence of MEs and IE.

Given the role of power three characteristics were considered in deciding the MEs and IE variations:

1. **Effect's presence:** Whether the CAD effects are actually present in the data.
2. **Effect size**: We use three SD-based sizes that align with the effect size indices of Cohen (1988), Sawilowsky (2009), Gignac and Szodorai (2016), and Lovakov and Agadullina (2021) (Ben-Shachar et al. 2023):
   - Large effect size = 1 SD,
   - Medium effect size = 0.5 SD,
   - Small effect size = 0.25 SD.
3. **Effect's valence**: This is only used for IEs, addressing the traditional focus on the $\beta_3 > 0$ and $\beta_3 < 0$ hypotheses as illustrated in Figure 4.1.

We considered seven variations for MEs and seven for IEs, detailed in Table 4.2 and Table 4.3, respectively. We include both MEs for symmetry and additional characteristics for IE variations, given our interest in IE errors of different sizes and valences. Together the 49 unique combinations of MEs and IE provide enough context to systematically study CAD estimation errors.

Table 4.2 Seven ME variations: two sizes, positive valence, four effect combinations

ME Variation	MEs that have effects	Size of ME	Valence
1.	No ME	No ME	No ME
2.	Both x_1 and x_2	Large	+
3.	Both x_1 and x_2	Small	+
4.	Only x_1	Large	+
5.	Only x_1	Small	+
6.	Only x_2	Large	+
7.	Only x_2	Small	+

Table 4.3 Seven IE variations: three sizes, two valence, two effect combinations

IE Variation	Size of IE	Valence
1.	No IE	No IE
2.	Large	+
3.	Medium	+
4.	Small	+
5.	Large	-
6.	Medium	-
7.	Small	-

Simulation Steps and Data Creation

The 441 unique characteristic combinations represent the true or population CAD effects. These effects are compared to estimates obtained from regression coefficients and marginal means comparison to calculate Type-I and Type-II errors for CAD effects. Since regression coefficients used to construct the true effects only provide three of the seven CAD effects, we recalculated all seven effects for the population using the marginal means approach and present them alongside simulation characteristics in the online supplement.

For each of the 441 combinations, we simulated 1,000 datasets, each with 1,000 observations. The outcome or y variable for each observation was generated using equation (v), independently adding the random variable ε_i for each case.

The estimated CAD effects were then obtained through regression analysis of all datasets, using both coefficients and marginal means comparisons. This resulted in a dataset of 441,000 rows of estimates, each containing the estimated CAD effects for a given simulation. By comparing these estimates to the true population effects, we calculated the errors of interest.

Simulation Results and Pitfall Resolution

This section has three parts. The first two parts detail results of the simulation analysis to address the pitfalls of the analysis and interpretation phase, respectively. The third part discusses what implications these results have for CAD research.

The simulation analysis focuses on the IE's and MEs' errors, using both marginal means and regression coefficient analysis, across different data groups based on three parameters:

1. CAD homogeneity: high, low, heterogeneous.
2. Own effect size:
   IE: small, medium, large.
   ME: small, large.
3. Interdependent effects:
   IE's dependence on MEs' size
   MEs' dependence on IE's size and valence

To address analysis approach choices, we first examine errors associated with the popular step-wise analysis approach, as compared to using just the full model. Specifically, we look at how often an IE can go undetected and its implication on the MEs' errors. Secondly, we examine how the IE and MEs' errors vary with the coefficient and marginal means approach:

- IE: $\hat{\beta}_3$ vs DiD
- ME x_1: $\hat{\beta}_1$ vs PC:00-10
- ME x_2: $\hat{\beta}_2$ vs PC:00-01

This is evaluated when addressing the interpretation phase's pitfall where we separately look at variations in the MEs' and IE's errors across different data groups.

I. Addressing Analysis Pitfall: Approach Choice

IE is generally more prone to being undetected than MEs

The IE was more likely to be undetected than either of the MEs, given the same large or small effect size. Table 4.4 shows this with the IE's Type-II error exceeding those of the MEs, regardless of the evaluation approach. When considering 'All' the data, the IE was about 15 percentage points (pp) more likely to be undetected compared to a ME, regardless of the approach.

The table also shows how this trend generally holds across different group homogeneities, effect sizes, and valence. ME and IE sizes are identical because we only included datasets where both matched, so the error estimated for a small IE is the same as for a small ME. We also found similar results when repeating the analysis for the entire data, including cases with unequal ME and IE sizes, and present the result in Table C1. This is available in Appendix C along with other supplementary results.

Table 4.4 Comparison of Type-II error rates between the IE and the MEs of x_1 and x_2 where they were constructed to have the same large or small effect size, using both marginal means and the regression coefficient approach

Data Group	Type-II Error					
	Marginal Means Approach			Regression Coefficient Approach		
	IE DiD	ME x_1 PC:00-10	ME x_2 PC:00-01	IE $\hat{\beta}_3$	ME x_1 $\hat{\beta}_1$	ME x_2 $\hat{\beta}_2$
All	34.02%	19.09%	18.97%	34.06%	19.13%	19.01%
Homogeneity						
Heterogeneous	25.08%	10.51%	10.68%	25.14%	10.56%	10.71%
Low	32.35%	17.94%	17.66%	32.39%	17.97%	17.70%
High	37.92%	22.39%	22.36%	37.95%	22.43%	22.38%
IE Size						
Large IE	0.03%	0.00%	0.00%	0.03%	0.00%	0.00%
Small IE	68.01%	38.18%	37.94%	68.08%	38.26%	38.01%
IE Valence						
Positive IE	33.96%	19.11%	18.88%	33.99%	19.14%	18.92%
Negative IE	34.09%	19.08%	19.06%	34.12%	19.12%	19.09%
MEs Present						
One	33.93%	19.05%	18.88%	33.95%	19.09%	18.92%
Both	-	19.13%	19.06%	-	19.17%	19.09%
ME Size						
Large ME	0.03%	0.00%	0.00%	0.03%	0.00%	0.00%
Small ME	68.01%	38.18%	37.94%	68.08%	38.26%	38.01%

False detection of MEs spikes with step-wise analysis

When an IE is non-significant, the convention is to use step-wise regression and remove the IE in a reduced model—the short model. However, IEs are highly prone to Type-II errors, with about 68% of small IEs going undetected. As a result, MEs are often estimated using the short model where the excluded IE adds to unobserved covariates, increasing omitted variable bias (Brambor, Clark, and Golder 2006). To assess the impact of step-wise analysis on ME estimation, we reanalyzed the datasets with undetected IEs using both short and long models. The resulting ME errors for x_1 and x_2, corresponding to regression coefficients β_1 and β_2, are shown in Table 4.5.

Although the step-wise analysis aims to improve MEs' power with the short model, the gains were minimal, if at all. Table 4.5 shows this with 'All' datasets having a ~5% power increase, i.e. decrease in Type-II error, when moving to the short model. However, this gain came at a considerable cost: a nine-fold increase in Type-I error, jumping from the standard ~5% in the long model to ~47% in the short model. This spike in Type-I error alongside nominal power gains was seen across most data groups. This trade-off pattern advocates using the long or full model over the step-wise analysis, especially when there is no prior information about the IE.

Table 4.5 Type-I and Type-II error for the MEs of x_1 and x_2, estimated using short and long model with coefficients β_1 and β_2, for those datasets where IE had a Type-II error in the initial long or full model.

Datasets: Type-II error in IE	**Short Model: Without IE**				**Long Model: With IE**			
	x_1 Error Type		**x_2 Error Type**		**x_1 Error Type**		**x_2 Error Type**	
Data Group	**I**	**II**	**I**	**II**	**I**	**II**	**I**	**II**
All	47.43%	14.61%	47.85%	14.57%	5.42%	20.08%	5.35%	19.91%
Homogeneity								
Heterogeneous	51.69%	13.29%	52.57%	13.18%	6.86%	11.69%	6.78%	11.80%
Low	47.75%	14.69%	48.36%	14.81%	5.81%	18.14%	5.95%	17.89%
High	46.66%	14.71%	46.91%	14.56%	4.95%	22.54%	4.74%	22.40%
IE Size								
Large IE	83.33%	19.23%	92.00%	10.53%	72.22%	23.08%	64.00%	42.11%
Medium IE	66.89%	14.10%	67.36%	14.65%	10.44%	19.98%	10.13%	19.95%
Small IE	42.26%	14.74%	42.63%	14.54%	4.05%	20.11%	4.04%	19.88%
IE Valence								
Positive IE	47.45%	0.43%	47.89%	0.38%	5.29%	12.20%	5.36%	11.97%
Negative IE	47.41%	28.81%	47.82%	28.80%	5.55%	27.98%	5.35%	27.86%
MEs Present								
None	47.77%	-	48.50%	-	5.28%	-	5.50%	-
One	47.26%	14.65%	47.53%	14.55%	5.49%	20.07%	5.28%	19.71%
Both	-	14.57%	-	14.58%	-	20.09%	-	20.10%
ME Size								
Large ME	46.78%	0.01%	47.89%	0.00%	5.27%	0.00%	5.27%	0.00%
Small ME	47.74%	29.21%	47.17%	29.10%	5.70%	40.16%	5.28%	39.77%

Note: Since we only analyzed datasets with undetected IEs, MEs' are more likely to appear significant, especially with larger undetected IEs. The long model's Type-I error reflects this with significantly higher errors for large and medium IEs compared to the standard 5%. Additionally, these errors are slightly higher than those in Table C2 and Table C3, which includes all simulated datasets, not just those with undetected IEs.

Three robustness checks for this trade-off pattern were conducted and the pattern persisted in all cases. First, we re-estimated both short and long models, as in Table 4.5, for all 441000 simulated datasets and present the results in Table C2 including now also the simulations with IE = 0. Second, we looked at how the errors would look if all datasets underwent step-wise analysis, just as researchers conventionally analyze in the absence of information on the true effects. The long model was applied when an IE was estimated as significant and the short model was applied when it was not, regardless of the true IE. Both models together covered all 441,000 simulated datasets, and their outputs are presented in Table C3.

Across these two robustness checks, the step-wise approach was beneficial in only one exceptional case: when the 'No IE' group is correctly estimated as non-significant. In this scenario, where there is no true IE and the IE is found non-significant, the short model increased power while maintaining the standard Type-I error of ~5%. However, the trade-off pattern reappears when a significant IE is wrongly estimated in the absence of a true IE. This is shown in the long model results for the 'No IE' group in Table A3.3, where MEs' Type-I error rose to ~35%, far exceeding the standard 5%. In contrast, the short model results for the same group in the same table showed substantial power gains with Type-II errors reduced to ~4%, while preserving Type-I errors at ~5%.

Without true IE information, the 'No IE' group is also better evaluated using the full model rather than the step-wise approach. This conclusion follows from both the above-discussed risks of the step-wise analysis and the full model's performance for all datasets in the 'No IE' group. This can be seen by comparing long model results for the 'No IE' group in Table C2 (full model) and Table C3 (step-wise analysis). While both had similar Type-II errors, the Type-I errors with step-wise analysis were ~35%, as against the standard ~5% with the full model.

For the third robustness check, we focused on datasets where the IE was estimated as non-significant, as these are most likely to benefit from the step-wise approach of dropping the IE. To evaluate the step-wise approach, we compared the results from the short model (Table C3) to their counterfactual full model analysis (Table C4). The error differences are presented in Table 4.6. Positive values, highlighted in red, indicate higher error with the short model, while negative values, highlighted in green, indicate higher error with the full model. Uncolored cells represent negligible differences between the models.

Type-I errors were consistently higher with the short model. Even the No IE group favored the full model with a 1 to 2pp error difference. For all other groups, this difference ranged from 11.11pp to 57.23pp. Type-II error differences were negligible or negative, with a maximum power gain of 31.58pp for only 44 datasets for the short model. Across all datasets, the step-wise short model increased MEs' power by about 9pp but raised Type-I error by 28pp. This modest power gain, coupled with a threefold increase in Type-I error, reaffirms the trade-off pattern identified earlier.

Table 4.6 Type-I and Type-II error differences for MEs of x_1 and x_2 between the short and full models for the 168,441 datasets estimated to have non-significant IEs. Cells with changes are highlighted: red for increases (positive) and green for decreases (negative) when using the short model.

Datasets: IE Non-Sig.	**Difference in Error Type: Short Model – Full Model (pp)**				
Data Group	x_1		x_2		**No. Datasets**
	I	**II**	**I**	**II**	
All	**27.68**	**-8.72**	**27.91**	**-8.69**	168441
Homogeneity					
Heterogeneous	24.35	-3.31	24.57	-3.31	13947
Low	26.42	-7.52	26.59	-7.36	69310
High	29.25	-10.58	29.52	-10.66	85184
IE Size					
Large IE	**11.11**	-3.85	28.00	**-31.58**	44
Medium IE	56.45	-5.88	**57.23**	-5.30	22808
Small IE	38.20	-5.37	38.59	-5.34	85722
IE Valence					
Positive IE	42.16	-11.77	42.53	-11.60	54386
Negative IE	41.86	0.83	42.47	0.93	54188
No IE	1.75	-14.62	**1.48**	-14.77	59867
MEs Present					
None	27.99	-	28.11	-	24088
One	27.52	-8.78	27.81	-8.56	95888

Datasets: IE Non-Sig.	**Difference in Error Type: Short Model – Full Model (pp)**				
Data Group	x_1 **I**	x_1 **II**	x_2 **I**	x_2 **II**	**No. Datasets**
Both	-	-8.66	-	-8.82	48465
ME Size					
Large ME	27.32	0.00	28.15	0.00	72205
Small ME	27.72	-17.44	27.45	-17.37	72148

Similar trade-off patterns were observed in economic field experiments (Muralidharan et al. 2023) and medicinal randomized factorial trials (Kahan 2013), (Kahan 2013), both recommending the full model over the step-wise approach. The step-wise analysis often led to biased and misleading conclusions by ignoring potential interactions, while the full model consistently outperformed it, even in the absence of a true interaction.

Thus, when no prior information on the IE is available, the full model is preferred over the step-wise approach. It considerably reduces Type-I, ensuring that it is at the 5% that it should be. Type-II errors slightly increase, but these costs do not weigh against the Type I error problem. This recommendation also aligns with previous research discouraging data-dependent model selection due to its tendency to produce invalid inferences (e.g. Leeb and Pötscher 2005).

II. Addressing Interpretation Pitfall: Parameter-Error Association

IE: Small effect size and high homogeneity reduce detection, independent of method and MEs

Table 4.7 summarizes variations in the IE's Type-I and II errors across our simulation parameters. The errors estimated using both the marginal means and the regression coefficients approach are presented under columns 'DiD' and '$\hat{\beta}_3$', respectively. Both methods showed similar Type-I and II errors, with differences shown in 'DiD – $\hat{\beta}_3$' columns which was at most a meagre 0.09pp. Certain parameters cannot have certain errors by design; for instance, when considering the IE's size and valence, we limit data to cases with an IE, eliminating Type-I errors.

Table 4.7. Type-I and II errors in estimating the IE using the marginal means difference-in-difference (DiD) and the regression coefficients ($\hat{\beta}$) approach, alongside the corresponding difference in errors, thus estimated. Estimates are presented for the entire data and separately for subsets varying according to homogeneity, IE's size & valence, and ME's presence & size.

Data Group	**Type-I Error** **DiD**	$\hat{\boldsymbol{\beta}}_3$	**DiD - $\hat{\boldsymbol{\beta}}_3$**	**Type-II Error** **DiD**	$\hat{\boldsymbol{\beta}}_3$	**DiD - $\hat{\boldsymbol{\beta}}_3$**
All	5.00%	4.97%	0.03pp	28.67%	28.72%	-0.05pp
Homogeneity						
Heterogeneity	5.27%	5.26%	0.01pp	17.37%	17.42%	-0.05pp
Low	5.07%	5.04%	0.03pp	25.39%	25.43%	-0.04pp
High	4.86%	4.84%	0.03pp	34.79%	34.84%	-0.06pp
IE Size						
Large IE	-	-	-	0.03%	0.03%	0.00pp
Medium IE	-	-	-	18.04%	18.10%	-0.06pp
Small IE	-	-	-	67.95%	68.03%	**-0.09pp**

Data Group	Type-I Error DiD	Type-I Error $\hat{\beta}_3$	Type-I Error DiD - $\hat{\beta}_3$	Type-II Error DiD	Type-II Error $\hat{\beta}_3$	Type-II Error DiD - $\hat{\beta}_3$
IE Valence						
Positive IE	-	-	-	28.73%	28.78%	-0.05pp
Negative IE	-	-	-	28.62%	28.67%	-0.05pp
No IE	5.00%	4.97%	0.03pp	-	-	-
MEs Present						
None	4.91%	4.86%	0.06pp	28.69%	28.75%	-0.06pp
One	5.18%	5.15%	0.03pp	28.54%	28.58%	-0.04pp
Both	4.69%	4.68%	0.01pp	28.93%	28.99%	-0.06pp
ME Size						
Large ME	4.97%	4.96%	0.01pp	28.68%	28.73%	-0.05pp
Small ME	5.05%	5.03%	0.03pp	28.66%	28.71%	-0.04pp

The Type-I error remained consistently around 5%, showing no significant variation with any of the parameters. However, the Type-II errors varied considerably with homogeneity and IE sizes—underlying the need for researchers to carefully examine non-significant IE and realize that with smaller effect sizes and in populations in which some groups are small, power reduces considerably to pick up IEs effects. This is crucial as CAD studies typically involve observational or field research, which are known to generally have small IEs (Baranger et al. 2023; Murphy and Russell 2017). Realize that all these results are based on a sample size of 1000 units, which is not small for many social science studies.

IE's detection does not depend on MEs' presence or size

Given the interdependence of IE and MEs, we examined how IE's Type-II error varies with MEs' presence and size. Building on our findings on IE's Type-II error variation by size, we analyzed all three IE sizes across ME parameters and found no variation. Results for MEs' presence are shown in Table 4.8. Large IEs continued to be almost always detected with negligible error, regardless of the ME. Medium and small IEs again had higher error rates but showed no variation with the ME. The results remained consistent when extended to all combinations of ME size and presence, as shown in Table C5.

Table 4.8 Association between the estimated MEs present and the Type-II Error for the IE, from both marginal means (DiD) and the regression coefficients approach, subset by the three IE sizes constructed

IE Size	MEs Present	Type-II Error of IE DiD	Type-II Error of IE $\hat{\beta}_3$
Small IE	Both	68.29%	68.41%
Small IE	One	67.87%	67.93%
Small IE	None	67.58%	67.69%
Medium IE	Both	18.48%	18.53%
Medium IE	One	17.72%	17.78%
Medium IE	None	18.44%	18.51%
Large IE	Both	0.03%	0.03%
Large IE	One	0.03%	0.03%
Large IE	None	0.05%	0.05%

ME: Small effect size & high homogeneity reduce detection, independent of method & IE

Table 4.9 summarizes variations in both MEs' Type-I and II errors across our simulation parameters. The errors estimated using both the marginal means and the regression coefficients approach are presented for x_1 under 'PC:00-10' and '$\hat{\beta}_1$', and for x_2 under 'PC:00-01' and '$\hat{\beta}_2$'. Both methods showed similar Type-I and II errors, such that the difference in errors estimated by the two methods never exceeded 0.07pp.

These differences are detailed in Table C6 under 'PC – $\hat{\beta}$', along with differences estimated using adjusted PC, reflecting typical *p*-value adjustments in PC studies. Such adjustments can impact error estimation and approach choice, as noted alongside Table C6.

Table 4.9 Type-I and II errors in estimating the ME of x_1 and x_2, using the marginal means pairwise comparison (PC) and the regression coefficients ($\hat{\beta}$) approach. Estimates are presented for the entire data and separately for subsets varying according to homogeneity, IE's size & valence, and ME's presence & size.

Data Group	x_1				x_2			
	Type-I Error		Type-II Error		Type-I Error		Type-II Error	
	PC:00-10	$\hat{\beta}_1$	PC:00-10	$\hat{\beta}_1$	PC:00-01	$\hat{\beta}_2$	PC:00-01	$\hat{\beta}_2$
All	5.05%	5.02%	18.98%	19.02%	5.01%	4.98%	18.97%	19.00%
Homogeneity								
Heterogeneous	5.30%	5.26%	10.37%	10.43%	5.13%	5.09%	10.33%	10.36%
Low	5.06%	5.03%	17.76%	17.79%	4.98%	4.95%	17.78%	17.83%
High	4.98%	4.95%	22.35%	22.38%	5.00%	4.98%	22.31%	22.34%
IE Size								
Large IE	5.01%	4.98%	18.92%	18.96%	5.10%	5.09%	19.11%	19.14%
Medium IE	5.11%	5.06%	18.98%	19.01%	4.94%	4.92%	18.85%	18.89%
Small IE	4.94%	4.92%	19.09%	19.13%	5.01%	4.97%	18.97%	19.01%
IE Valence								
Positive IE	4.96%	4.93%	18.96%	18.99%	5.00%	4.97%	18.83%	18.86%
Negative IE	5.08%	5.04%	19.03%	19.07%	5.03%	5.01%	19.13%	19.16%
No IE	5.25%	5.22%	18.88%	18.92%	4.96%	4.92%	18.91%	18.96%
MEs Present								
None	4.99%	4.97%	-	-	5.01%	4.99%	-	-
One	5.08%	5.05%	19.14%	19.17%	5.01%	4.97%	18.97%	19.01%
Both	-	-	18.82%	18.86%	-	-	18.97%	19.00%
ME Size								
Large ME	5.06%	5.02%	0.00%	0.00%	5.02%	4.98%	0.00%	0.00%
Small ME	5.11%	5.07%	37.96%	38.03%	5.00%	4.97%	37.93%	38.01%

As with the IE, Type-I errors stayed around 5% with no notable variation across parameters, while Type-II errors varied with homogeneity and IE sizes—underlying the need for researchers to carefully examine non-significant MEs.

The increase in homogeneity paralleled the increase in ME's Type-II error for both x_1 and x_2. The ME was about 12pp more likely to go undetected with high homogeneity, as compared to heterogeneous groups. This pattern remained consistent regardless of the MEs' presence, as shown in Table C7.

Large MEs were almost always detected, with negligible Type-II error for x_1 and x_2, across both methods. However, this error strongly increases for small MEs, jumping to about 38% for x_1 and x_2, across both methods. This pattern remained consistent regardless of the number of MEs present, as shown in Table C8.

MEs detection does not depend on the IE's size or valence

Building on our above findings on the ME's error variation by size, we analyzed the Type-II error separately for small and large ME sizes across IE parameters and found no variation. Results across the IE's valence and size are shown in Table 4.10 and Table 4.11, respectively. The large MEs' continued to be almost always detected, regardless of the IE's valence or size, with negligible error. While small MEs again had a higher error rate, they did not vary with the IE's valence or size. The same results were found when extending the analysis to all combinations of IE valence and size, as shown in Table C9.

Table 4.10 Association between MEs' Type-II error and the IE's valence, using both marginal means (PC) and the regression coefficients ($\hat{\beta}$) approach, subset by the two ME sizes

ME Size Constructed	**IE Valence**	**Type-II Error for ME of**			
		x_1		**x_2**	
		PC:00-10	$\hat{\beta}_1$	PC:00-01	$\hat{\beta}_2$
Small ME	No IE	37.75%	37.84%	37.82%	37.92%
Small ME	Negative IE	38.05%	38.14%	38.25%	38.32%
Small ME	Positive IE	37.93%	37.99%	37.65%	37.72%
Large ME	No IE	0.00%	0.00%	0.00%	0.00%
Large ME	Negative IE	0.00%	0.00%	0.00%	0.00%
Large ME	Positive IE	0.00%	0.00%	0.00%	0.00%

Table 4.11 Association between MEs' Type-II Error and the IE's size, using both marginal means (PC) and the regression coefficients ($\hat{\beta}$) approach, subset by the two ME sizes

ME Size Constructed	**IE Size**	**Type-II Error for ME of**			
		x_1		**x_2**	
		PC:00-10	$\hat{\beta}_1$	PC:00-01	$\hat{\beta}_2$
Small ME	No IE	37.75%	37.84%	37.82%	37.92%
Small ME	Large IE	37.83%	37.92%	38.22%	38.28%
Small ME	Medium IE	37.96%	38.01%	37.69%	37.77%
Small ME	Small IE	38.18%	38.26%	37.94%	38.01%
Large ME	No IE	0.00%	0.00%	0.00%	0.00%
Large ME	Large IE	0.00%	0.00%	0.00%	0.00%
Large ME	Medium IE	0.00%	0.00%	0.00%	0.00%
Large ME	Small IE	0.00%	0.00%	0.00%	0.00%

Implications for CAD Research

For analyzing unbiased and efficient group difference estimates, one could either use the marginal means approach to estimate all group differences from the full model, or complement it with three direct group difference estimates from the full model's regression coefficients. Both methods provided comparable estimates for the three common group differences. The step-wise approach, however, resulted in biased and invalid conclusions, with elevated Type-I errors even in the absence of an interaction. Hence, the step-wise approach is strongly advised against.

To resolve interpretation challenges, we explored how parameter information could help interpret non-significant effects. Both MEs and IE were more likely to go undetected when effect sizes are small and homogeneity is high. Small effect sizes make CAD interactions particularly difficult to detect. Since IEs are generally harder to detect than MEs and often have small effect sizes in field and observational studies, CAD interactions are highly likely to go undetected. Thus, it is crucial to consider power implications when interpreting non-significant results. For example, to detect an IE half the size of an ME, an adequately powered study would need eight times the sample size (Muralidharan et al. 2023).

Homogeneity plays a crucial role in identifying inequality in resource distribution. If the CAD interaction creates the same ME and IE across homogeneity combinations, as in Table 4.1, CAD effects are easiest to detect in a heterogeneous population where all groups have the same size. In contrast, they are far less likely to be observed in highly homogeneous populations with skewed group sizes ($UL{:}\,UH{:}\,SL{:}\,SH\ =\ 4\%{:}\,16\%{:}\,16\%{:}\,64\%$).

Detecting resource inequality becomes systematically harder in highly homogeneous populations, a critical challenge for inequality research which often focuses on such populations. For instance, gender inequality in pay becomes harder to detect at higher career levels as the proportion of women drops, increasing homogeneity. Similarly, SES inequality is harder to identify with academic progression or wealth accumulation as the share of low SES individuals declines, increasing homogeneity. This risk of undetected CAD applies to any scenario with skewed representation, regardless of status. For example, highly homogenous gender is found with both women's underrepresentation in STEM and men's underrepresentation in care roles.

4.6. FUTURE RESEARCH GUIDELINES AND LIMITATIONS

We present a guideline for CAD interaction research that synthesizes the above insights and addresses the pitfalls identified in the theorizing, analyzing, and interpreting of CAD interaction research. These guidelines are outlined along the three phases below.

Phase 1. Theorize: Visualize Interaction of All Groups

1. **Develop comprehensive theories:** Include hypotheses for all group comparisons (MEs or six PCs) and for differences between comparisons (IE or DiD). Verify comprehensiveness by graphing these hypotheses to create line graphs as in Figure 4.1.
2. **Focus on interaction groups:** Limit comparisons to the four interaction groups using PC or DiD. Do not compare with parent group or grand mean, because each subgroup is part of the parent group and its impact on the grand mean depends on the size of the group.

Phase 2. Analyze: ME, IE, and Pairwise Comparisons
3. **Do not use step-wise analysis:** If an IE is hypothesized always include the interaction term in the regression analysis, regardless of its significance. The estimates of the MEs are sound in this case in contrast to the model without the interaction. 4. **Unadjusted *p*-values**: Do not adjust *p*-values when evaluating a specific hypothesis for each PC. This leads to an unnecessary reduction in power given that there is a hypothesis about a specific comparison. 5. **Equivalent methods:** Regression coefficients and marginal means give equivalent results when a hypothesis exists for each PC.
Phase 3. Interpret: Consider Data & Power when Non-Significant
6. **Cautiously interpret non-significant results:** Examine estimated graphs, data, and power, particularly when effect sizes are small or there is high group homogeneity and realize the potentially large Type-II errors. This also implies that really large sample sizes are necessary if one is interested in IEs for which some of the subgroups are small.

While these guidelines provide a strong starting point, they focus on the static study of two binary CAD interactions. However, CAD variables can take many other forms. For instance, while gender is often studied as a time-invariant binary variable, race might be a categorical CAD, education a ordinal CAD, and age a continuous one. Future research should explore how the guideline could be extended to such variations and whether they require different considerations. Most likely, if more complex variables or higher-order interactions become relevant, similar guidelines will be relevant and especially power issues will blow up.

Additionally, we used continuous outcomes to study resource divergence, but outcomes can also be categorical. For instance, Bernardi's (2014) compensatory advantage looked at whether students were promoted or not. We expect our findings to extend to categorical outcome variables, given the work of Kahan (2013). They similarly assessed the effects of step-wise regression on bias and Type-I error using two sets of simulations, one using continuous outcomes and the other using binary outcomes, and found similar results. Nevertheless, future research could validate whether this holds in the context of CAD interactions.

Finally, CAD is a dynamic mechanism with effects that change over time. For instance, gender-based CAD drives inequality among scientists during the early stages of their careers but diminishes later (DiPrete and Eirich 2006; Long and Fox 1995). While we do not delve into the time aspect here our guidelines can be extended to a longitudinal study of CAD, where one can observe how group differences change over time and whether any systematic patterns could be added to the guidelines.

Future research could also explore Bayesian informative hypotheses evaluation for CAD interactions. Bayesian methods are better suited at directly and powerfully evaluating specific and complex questions[9], that traditional significance testing struggles with. For

[9] A Bayesian approach is often recommended over traditional approaches when studying interactions with specific research hypothesis as it can more powerfully evaluate precise research hypotheses by defining all relevant hypotheses in terms of inequality constraints among parameters and the comparison of their associated probability within a Bayesian inferential framework (Garofalo et al. 2022; Gu, Mulder, and Hoijtink 2018; Hoijtink 2011).

instance, how CAD effects vary over time, whether status CAD differs across sectors, or if UL consistently has the lowest outcomes.

Nevertheless, some guidelines apply broadly to any two-by-two interaction analysis and studies on resource distribution. In particular, avoiding stepwise regression is good practice for all two-by-two interaction analyses. This approach is also recommended in factorial experimental design and supported by evidence from other contexts (e.g. Kahan 2013; Muralidharan et al. 2023). Likewise, the Phase 1 guidelines on theorization apply broadly to CAD studies and any research concerned with how finite resources are distributed across groups. When resources are limited, distribution becomes a zero-sum game, which requires considering all possible group comparisons for a comprehensive theory. While this guideline applies to studies of how group interactions influence resource distribution, it does not extend to all interaction studies.

We hope these guidelines serve as a foundation for CAD interaction research. By addressing current limitations and expanding to dynamic CAD mechanisms and diverse variable types, future studies can build on this framework to deepen our understanding and address inequality.

CHAPTER 5

Cumulative Advantage as a Mechanism for Career (In)Equality: Can Some Women Benefit?

Abstract

Despite extensive research, gender inequality in career outcomes like pay, productivity and innovation remain poorly understood. We study whether this could be explained by cumulative advantage/disadvantage (CAD), the tendency for initial advantages or disadvantages to accumulate over time. Using quasi-random evaluations of over 2.8 million patent applications from the U.S. Patent and Trademark Office, we examine how CAD arising from inventors' gender and their first patent application outcome causally affects their innovation output. Challenging traditional views, we find that CAD can both widen and reduce inequality, depending on gender and early outcomes. Over seven years, initially successful women and initially unsuccessful men overcame early disadvantages, matching the innovation output of initially successful men, the most advantaged group. In stark contrast, output for initially unsuccessful women was only about half as much, making them the only group to incur large and irreversible disadvantages. These findings show that gender inequality is structurally embedded early in careers, highlighting the need for early-career interventions as against later-stage diversity efforts that address symptoms rather than causes.

The authors jointly developed the core ideas of this chapter. **Sanjana Singh** wrote the main part of the manuscript and conducted the analyses. **Eva Jaspers** and **Tanja Van der Lippe** substantially contributed to the manuscript. An earlier version of this chapter has been presented at *Dag van de Sociologie (2021)*, and *Networks 2021: A Joint Sunbelt and NetSci Conference (2021)*.

5.1. INTRODUCTION

Women have a harder time climbing the career ladder. Men and women start out equally represented in jobs that require no prior experience. But, women's presence steadily declines along the career ladder, with only 29% of American C-suite executives being women (Krivkovich et al. 2024). This widening gender gap is not limited to representation but also appears in other career outcomes across industries and countries (e.g. Albrecht et al. 2018; Arulampalam, Booth, and Bryan 2007; Blau and Duncan 1967; Costa Dias, Joyce, and Parodi 2018; Huang et al. 2020). For instance, the British gender pay gap is negligible at labor market entry, but unexplainedly widened to 8% in ten years (Manning and Swaffield 2008).

While factors like education, childcare, job experience, mobility, flexibility and psychological characteristics like competitiveness contribute to the gap (e.g. Albrecht et al. 2018; Arulampalam et al. 2007; Averett, Argys, and Hoffman 2018; Blau and Kahn 2012, 2017; Manning and Swaffield 2008; Mincer and Polachek 1978), a significant portion remains unexplained (e.g. Pelletier, Patterson, and Moyser 2019).

Career-family conflict, from marital and parental status, has been considered a key driver of the gap (e.g., Blau and Kahn 2017; Perry-Jenkins and Gerstel 2020; Toczek, Bosma, and Peter 2021) and also the last major barrier to gender equality, especially for highly educated workers (Goldin 2021). However, recent studies find the conflict mechanism to have a declining role in explaining the gap (England et al. 2016; Kelley, Galbraith, and Strong 2020; Perry-Jenkins and Gerstel 2020), with its impact being overstated (Cha, Weeden, and Schnabel 2023) and fleeting (Morgan et al. 2021). For example, women inventors earn about 14% less than men, even after accounting for parenthood, job selection, and sources of heterogeneity (Hoisl and Mariani 2017). This is crucial as scientific careers have high career-family conflict due to their long work-hours, strict performance standards, and strong identification with work (Fox, Fonseca, and Bao 2011).

So, how can we effectively study gender inequality in career progression? First, the widening gender gap suggests the need for a *cumulative* approach, examining an employee's entire career trajectory. For instance if the British gender pay gap was only evaluated at the start of a career, it would appear non-existent. Second, the consistently widening inequality observed across outcomes, industries and countries suggests a *tendency* for men to accumulate a relative advantage and women a disadvantage, increasing their career inequality over time. Finally, the persistence of the unexplained gender gap suggests a *systematic* mechanism beyond what standard controls capture.

Social science literature describes such *cumulative systematic tendencies* as cumulative advantage/ disadvantage (CAD), where a relative advantage becomes a resource for further gains. Gender is a CAD factor that drives inequality, with men's (M) higher social status giving them an initial advantage over women (W). Over time this advantage accumulates, widening the gender gap in resource accumulation with men having a relative cumulative advantage (CA), and women a disadvantage (CD).

In this contribution, we examine CAD as a mechanism for gender inequality in inventor careers. The high career-family conflict observed in scientific careers and their comparable outcomes make them ideal for understanding inequality in career progression. However, gender is not the only CAD factor triggering career inequality—career success itself can reinforce inequality. Early success (S) allows individuals to accumulate more advantages than those who were initially unsuccessful (U) (e.g. Bol et al. 2018), a pattern reflected in sayings like "success begets success." For our inventors, S or U corresponds to their first innovation being successful or not. Since patents validate successful innovations, patent application outcomes naturally measure success, and this is why it is worthwhile to focus on inventors. Inventors with a successful first application, S, gain a CA over those with an unsuccessful application, U, widening the resource gap between them over time.

Thus, inequality in inventor careers follow from two CAD factors: their gender (M vs W) and the outcome of their first patent application (S vs U), dividing them into four mutually exclusive CAD groups: successful men (SM), successful women (SW), unsuccessful men (UM), and unsuccessful women (UW). Unlike the traditional CAD mechanism, where we study the relative (dis)advantage between two groups, we must examine how all four groups differ from each other to know if and how CAD drives inequality. However, unlike the well-established independent CAD mechanisms of gender and success, their interaction remains unknown so far in the literature.

To address this, we conduct the first systematic study on if and how CAD can explain gender inequality in career progression. Following the guidelines developed in Chapter 4, for studying CAD interactions, we examine whether CAD always drives inequality or, in some cases, fosters equality. Our findings highlight CAD's potential in examining contemporary gender inequality in the workplace, alongside providing novel insights for policymakers and researchers.

Why has CAD Remained Untapped as a Mechanism for Career Inequality?

Although CAD evolved from studies on career inequality, such as Merton's (1968) Matthew effect in scientific careers, its use in organizational contexts has declined. Methodological constraints have pushed most theoretical and empirical advances in CAD research outside organizations. In 2006 DiPrete and Eirich emphasized these challenges, noting that most sociologists studying careers focused on other areas, with CAD often studied only implicitly. This situation has changed little since, with three key challenges hindering the study of CAD as a mechanism for career inequality.

First, external validity or generalizability of findings would require observed rather than experimental data, which ideally tracks employee career outcomes across time, organizations, and sectors. Second, studying career progression across time, organizations, and sectors would need quantifiable and comparable career outcome measures.

We address both these challenges using large-scale population data from the United States Patent and Trademark Office (USPTO), the federal agency for granting U.S. patents. Drawing on detailed records of all patent applications, their examination, and outcomes, we compiled a dataset linking 486,144 unique inventors to over 2.8 million applications filed between 2000 and 2020, and to their examination across 680 USPTO technology art units. As our data includes all applications, regardless of their patenting success, we can identify the start of each inventor's career and track their progression even if they changes firms or become self-employed.

The second challenge is resolved by the streamlined patent evaluation and reward process. All inventors share the same initial career outcome – their first patent application being successful or not. This is a comparable measure, as all inventors must submit their application to the USPTO and are subjected to the same standardized evaluation process which results in either a successful or unsuccessful patent grant. The first patent application's outcome is also particularly salient. For instance a start-up's first patent application outcome is found to drive its long-term growth (Farre-Mensa, Hegde, and Ljungqvist 2020). Thus, the outcome of an inventor's first application can trigger CAD effects, shaping their career progression.

The third and most critical challenge is ensuring internal validity or establishing a causal link between CAD and employees' future outcomes. This arises from CAD's endogeneity problem—separating its effects from unmeasured heterogeneity. For instance, in the earlier gender pay gap that widened from 0 to 8%, it would be difficult if not impossible to know how much was caused by CAD versus unmeasured factors like intelligence or work quality. While endogeneity could be resolved with experimental randomization, this is rarely feasible in organizations and raises the external validity challenge, i.e. that findings from controlled experiments may not generalize to real-world settings. As organizations are shaped by complex norms, incentives, and power structures, even well-designed experiments may fail to capture how CAD operates over the course of actual careers.

In addition to the endogeneity problem, ensuring internal validity requires addressing the interplay between gender, initial outcomes, and future outcomes. This stems from gender's influence on initial outcomes, given gender-based discrimination in hiring, job assignments, and evaluations (Bohnet et al. 2016). We address these issues by exploiting USPTO's quasi-random patent evaluation process through a two-stage least squares (2SLS) instrumental variable (IV) regression, as proposed by Farre-Mensa et al., (2020) and Sampat and Williams (2019). The random assignment of examiners to applications allows us to use the strictness of an inventor's first application examiner as an IV to causally identify how first application outcomes affect men's and women's career progression using population-level real-world data.

5.2. Theoretical Overview: CAD Interactions and (In)Equality in Inventor Career Progression

CAD is the systemic tendency for interindividual divergence in a given characteristic, e.g. money, health or status, over time (Dannefer 2020). While most CAD research focuses on a single factor, interactions between CAD factors are found to explain much of inequality and even reduce it for some groups (Bernardi and Triventi 2020; Pedulla 2016). Chapter 4 outlined how the study of CAD interactions has been limited by its many pitfalls and offered guidelines to address them. Following these guidelines, we build a comprehensive theory that captures and visualizes all possible group comparisons.

Consider a group of identical inventors differing only in their gender (M or W). These inventors submit their first patent application to the USPTO, which would be of identical quality and have the same probability for being successful (S) or unsuccessful (U). Say, half our inventors are randomly successful in attaining a patent. This patent evaluation outcome is publicly released by the USPTO, allowing differentiating our identical inventors into four mutually exclusive groups: successful men (SM), successful women (SW), unsuccessful men (UM), and unsuccessful women (UW).

In an equal world, with no CAD from inventors' gender or initial success, all inventors would innovate similarly over time. Since they have identical capabilities, neither gender nor initial success would influence future innovation. As a result, we would expect all four groups to have similar opportunities, filing the same number of patent applications over time: $\mathrm{E}(UM) = \mathrm{E}(UW) = \mathrm{E}(SM) = \mathrm{E}(SW)$. However, if CAD exists, it will create inequalities among these groups, leading to differences in subsequent applications despite their equal abilities.

The CAD mechanism is triggered when inventors' true capabilities are unknown, and perceptions of their capability is shaped by their initial outcome and gender – creating inequality. Both gender (M vs. W) and initial outcome (S vs U) are independently known to signal actors' relative capability, diverging resources between two groups: One that has a CA and the other a CD.

Inventors from group S would signal to be more capable than those from U, accumulating more advantages. Initial success boosts motivation, appraisals, collaboration, and networks, leading to CA for S and CD for U. This is the path-dependent CAD or the Matthew effect, where past success positively affects future success in a self-reinforcing dynamic to create career inequality (Bol et al. 2018).

Similarly M would accumulate more advantages than W due to status beliefs associating men with greater competence (Ridgeway 2001; Ridgeway, Johnson, and Diekema 1993). This status-dependent CAD follows the same mechanism as the path-dependent CAD, where status is a resource.

Both path-dependent and status-dependent mechanisms consider only two groups of inventors where positive and negative signals diverge future outcomes. However we have four groups as we simultaneously consider inventors' initial outcomes and gender: SM, SW, UM and UW. This means that unlike previous scenarios the CAD mechanism cannot be theorized as a single comparison between two groups.

Chapter 4 showed how such CAD interactions can be comprehensively theorized using a line graph that captures seven comparisons: one indirect difference-in-difference comparison (DiD) or the commonly referred interaction effect, and six direct pairwise comparisons (PC). These seven comparisons are listed in the first two columns of Table 5.1 for our four inventor groups.

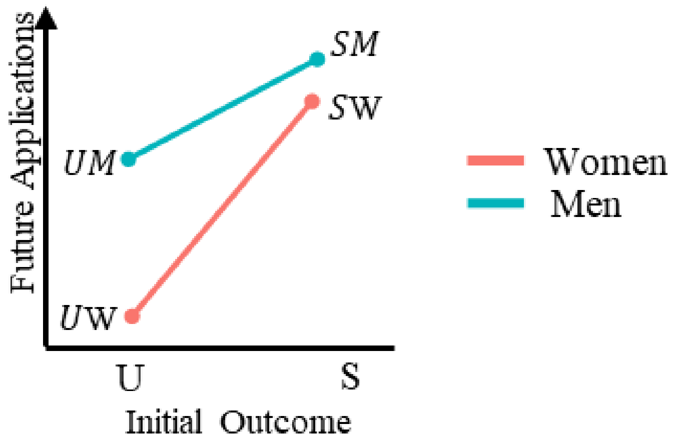
In this section we develop separate hypotheses, H_1-H_7, for each of the seven comparisons in Table 5.1. These hypotheses quantify how we expect CAD to create group inequalities over seven years: "> 0" indicates a relative advantage, "< 0" indicates a relative disadvantage, and "$= 0$" indicates no CAD effect. For example, H_2: $\mathrm{E}(SW) - \mathrm{E}(UW) > 0$ indicates that inventors in group SW are expected to accumulate more patents than those in UW.

Table 5.1 Seven possible group comparisons or CAD effects considered alongside our hypotheses on them, and how it changes over time

#	Group Comparisons	CAD Effect		
		Equal World	Our Hypotheses	Our Hypotheses Over Time
1	$[E(SM) - E(UM)] - [E(SW) - E(UW)] = [E(SM) - E(SW)] - [E(UM) - E(UW)]$	0	< 0 _(H_1)	Strengthens _(H_{1t})
2	$E(SW) - E(UW)$	0	> 0 _(H_2)	Strengthens _(H_{2t})
3	$E(UM) - E(UW)$	0	> 0 _(H_3)	Strengthens _(H_{3t})
4	$E(SM) - E(UW)$	0	> 0 _(H_4)	Strengthens _(H_{4t})
5	$E(SM) - E(UM)$	0	≥ 0 _(H_5)	Weakens but never negative _(H_{5t})
6	$E(SM) - E(SW)$	0	≥ 0 _(H_6)	Unclear: Weakens / Strengthens _(H_{6t})
7	$E(SW) - E(UM)$	0	$= 0$ _(H_7)	Unclear: Stable/ Positive _(H_{7t})

The seven hypotheses, H_1-H_7, are visually summarized in Figure 5.1. The future outcome of subsequent patents applications is denoted on the y-axis, initial outcome is on the x-axis, and gender is a grouping variable. Thus, the lines depict the CAD effect through the relation between inventors' CAD variables and their future outcomes, i.e. how future outcome diverges among the groups[10].

Figure 5.1 The theorized CAD effects



Building on Chapter 4's approach we also examine how the seven CAD effects evolve over time. We hypothesize whether the effects remain stable, strengthen or weaken, as outlined by H_{1t}-H_{7t} in Table 5.1. This follows evidence that career gender differences not only widen over time but can stabilize (Long 1992), diminish or reverse in the long term (Petersen and Saporta 2004; Stokke 2021). While H_1-H_7 hypothesize CAD effects at a single time point, H_{1t}-

[10] Appendix Figure D1 details how such graphs capture the relation between our inventors' future outcome and the CADs from their gender and initial outcome.

H_{7t} examine their trend over time using multiple time points. We develop these hypotheses below starting with H_1, the interaction effect hypothesis and examining its implication on other group comparisons.

Compensatory advantage of men's gender creates a converging interaction effect (H_1)
Path-dependent CAD suggests that initially unsuccessful inventors, UM and UW, accumulate disadvantages as compared to their successful counterparts, SM and SW. However, men can compensate for failure through their high-status gender. For instance, men are found to generally receive better appraisals (Bowen et al. 2000), often have bigger networks with a higher proportion of men, and have higher returns from these networks (Woehler et al. 2020), which can be particularly advantageous in male-dominated sectors like innovation.

Women, however, receive harsher punishments upon failure, a phenomenon known as the gender punishment gap (e.g. Egan, Matvos, and Seru 2022; Sarsons 2017). For instance, after financial advisory misconduct, women were 20% more likely to lose their jobs and 30% less likely to find new jobs relative to men (Egan et al. 2022).

Thus, men's future outcome will be less impacted by failure $[\mathrm{E}(SM) - \mathrm{E}(UM)]$, than women's $[\mathrm{E}(SW) - \mathrm{E}(UW)]$. This CAD interaction is known as compensatory advantage: a general mechanism for inequality where trajectories of high status individuals are less dependent on prior negative outcomes when compared to low status individuals (Bernardi and Triventi 2020). This expectation on the DiD or interaction effect is our first hypothesis, H_1.

$$[\mathrm{E}(SM) - \mathrm{E}(UM)] < [\mathrm{E}(SW) - \mathrm{E}(UW)]$$
$$\Rightarrow [\mathrm{E}(SM) - \mathrm{E}(UM)] - [\mathrm{E}(SW) - \mathrm{E}(UW)] < 0 \qquad (H_1)$$

This creates the converging pattern in Figure 5.1. Women's relative disadvantage from their initial failure is reflected through the red line having a steeper slope than the blue line, i.e. $[\mathrm{E}(SW) - \mathrm{E}(UW)] > [\mathrm{E}(SM) - \mathrm{E}(UM)]$.

Over time, we expect the compensatory advantage to strengthen, increasing the convergence in Figure 5.1, with the red line becoming steeper and the blue line flatter. As men dominate higher career levels, their high-status gender advantage is unlikely to fade and may even grow. Men are seen as more suitable for leadership (Paustian-Underdahl, Walker, and Woehr 2014; Ridgeway 2001), making their career progression less affected by early setbacks. In contrast, women face shorter scientific careers and higher dropout rates despite similar productivity (Huang et al. 2020) indicating greater sensitivity to initial outcomes.

Hence, we hypothesize H_1 to strengthen over time, with men's output becoming less influenced by initial setbacks than women's as their careers progress.

H_{1t}: H_1 Strengthens over time.

Unsuccessful women have the greatest disadvantage (H_2-H_4)
Among all groups, UW faces the most significant disadvantage due to both their low-status gender and initial failure. Unlike men, if women's first patent application is unsuccessful, it raises questions on both her competence and commitment. The ideal employee is expected to work long hours, be flexible, and limit aspects of their personal lives that conflict with

their paid work (Fox et al. 2011). As women are disproportionately burdened by demands outside the workplace, such as childcare and household work, organizations may attribute women's poor performance to career-family conflict (Kelly et al. 2010). This is reflected in the gender punishment gap. For example, venture-capital investors faced with two previously failed entrepreneurs, identical in all respects but gender, viewed the signal of failure more negatively for the woman (Pistilli et al. 2023).

In contrast, other groups benefit from at least one positive signal. SM, with positive gender and outcome signals, has a clear CA over UW. Groups with mixed signals—SW and UM—can reduce their disadvantages through positive attributes, also creating a CA relative to UW. SW's success helps overcome women's legitimacy deficit, which is higher in male-dominated fields like invention (Burt 1998; Kanter 1977c). SW's initial success legitimizes their position, increasing their access to resources, enabling them to accumulate more advantages than UW. UM can leverage their higher-status gender to gain more opportunities. As employee networks are often same-gendered (Woehler et al. 2020) and workplaces become more male-dominated higher up the hierarchy, UM can leverage their gender to build senior connections and access more resources than UW. This is in fact the gender punishment gap, referenced alongside H_3 below.

$$\mathrm{E}(SW) - \mathrm{E}(UW) > 0 \qquad \_(H_2)$$
$$\mathrm{E}(UM) - \mathrm{E}(UW) > 0 \qquad \underline{\text{ Gender Punishment Gap }}(H_3)$$
$$\mathrm{E}(SM) - \mathrm{E}(UW) > 0 \qquad \_(H_4)$$

Over time, UW's CD is expected to grow, strengthening the above hypothesized differences. Consistent with H_{1t}, these UW comparisons also align with evidence of women's shorter careers and higher dropout rates being key drivers of the gender gap in scientific productivity (Huang et al. 2020).

H_{2t}: H_2 Strengthens over time.
H_{3t}: H_3 Strengthens over time.
H_{4t}: H_4 Strengthens over time.

Men can reduce disadvantages from initial setbacks (H_5)
As outlined in H_1, men's future outcome is less impacted by failure. UM's high-status gender mitigates the impact of initial failure, reducing their disadvantage relative to SM. However, UM is unlikely to outperform SM, given their similar capabilities. Thus, UM's initial failure may remain a negative signal of capability, with SM having no or little advantage over UM.

$$\mathrm{E}(SM) - \mathrm{E}(UM) \geq 0 \qquad \_(H_5)$$

As with H_{1t}, we expect UM to continue leveraging their networks over time to reduce CD, but never outperform SM.

H_{5t}: H_5 Weakens but never reverses or becomes negative.

Initial success can reduce women's disadvantages (H_6-H_7)
Historically, men received higher career rewards (e.g. Goldin 2014; Neumark, Bank, and Van Nort 1996), while women were penalized for success (e.g. Heilman et al. 2004; Rudman et al. 2012). This suggests a gender reward gap similar to H_3's punishment gap, where men

have a relative advantage, i.e. $\mathrm{E}(SM) - \mathrm{E}(SW) > 0$. However, SW can leverage their success to reduce or even erase[11] gender disparities, just as UM leveraged their gender above.

$$\mathrm{E}(SM) - \mathrm{E}(SW) \geq 0 \qquad \text{Gender Reward Gap } (H_6)$$

Over time, H_6 could strengthen, weaken or disappear depending on how gender is valued at senior levels.

H_6 strengthens if men's gender advantage grows with career progression. This follows from innovation being a male-dominated field, with women accounting for a meagre 12.8% of all U.S. inventor-patentees (USPTO 2020). As the dominant group, men would hold more resources and power, bolstering perceptions of men being "better" inventors, regardless of their ability (Eagly and Carli 2007; Kanter 1977c; Ridgeway 2014). This is reflected in men typically being perceived as more effective leaders (Paustian-Underdahl et al. 2014) and women's equally impactful contributions leading to less successful careers (Hofstra et al. 2020). Thus as men would have more representation at senior levels, it is likely to reinforce their relative advantages, strengthening H_6.

H_6 weakens or even reverses if instead women's gender advantage grows with career progression. Unlike UM, who would not outperform SM due to shared gender (H_{5t}), SW could outperform SM if being a woman is valued sufficiently more at senior levels. This could happen when there is a push for gender diversity in leadership, giving a premium to the few women who remain in a male-dominated field. As UW are more likely to drop out, SW could benefit more from these practices, gaining greater access to resources than even SM (e.g. Green, Jegadeesh, and Tang 2009; Hill, Upadhyay, and Beekun 2015).

This is supported by findings that women can earn more than men when they are seen as high-potential (Leslie et al. 2017), or already hold senior roles (Hill et al. 2015). So while SM might initially have a relative advantage over SW, over time, this advantage could strengthen, weaken, disappear or even reverse depending on how gender is valued over their careers.

H_{6t}: H_6 can strengthen, weaken or disappear over time.

For the SW-UM comparison, we do not expect either group to have a significant relative advantage over the other, following both their mixed signals. Both can reduce their relative CD against SM. While UM's high-status gender would mitigate their initial setback, SW's initial success would reduce gender disparities, bringing their outcome close to or on par with SM. Consequently, SW and UM would have similar outcomes.

$$\mathrm{E}(SM) - \mathrm{E}(SW) \geq 0 \text{ and } \mathrm{E}(SM) - \mathrm{E}(UM) \geq 0 \Rightarrow \mathrm{E}(SW) - \mathrm{E}(UM) = 0 \qquad (H_7)$$

Similar to H_{6t}, H_{7t} will depend on how gender is valued. If there are no significant gender differences at senior levels, SW and UM will continue to have similar outcomes. If women are increasingly valued, SW will gain a relative advantage. If men are increasingly valued,

[11] The possibility of $\overline{SM} = \overline{SW}$ aligns with H_1's interaction effect convergence and can graphically be seen with $\overline{SM}$ and $\overline{SW}$ coinciding in Figure 5.1 while $\overline{UM}$ continues to be above $\overline{UW}$. Symbolically this would translate H_1 to H_3: $[\overline{SM} - \overline{SW}] - [\overline{UM} - \overline{UW}] < 0 \Rightarrow [\overline{UM} - \overline{UW}] > 0$.

UM could leverage their gender to overcome their initial setback, but this need not provide them an advantage over SW. UM's initial failure may remain a strong negative signal of capability, due to incongruence with their high status gender. For instance men faced harsher penalties for nonstandard employment histories compared to those with standard employment (Pedulla 2016). This hypothesis is consistent with H_{2t}, H_{3t}, and H_{4t}.

H_{7t}: H_7 can remain stable or become positive over time.

Following Chapter 4, Figure 5.1 visualizes our seven comparison hypotheses, H_1-H_7, ensuring the comprehensiveness of our theory. The corresponding hypotheses on CAD trends, H_{1t}-H_{7t}, are internally consistent. The next sections describe how these hypotheses are studied using USPTO data.

5.3. INSTITUTIONAL SETTING AND DATA

Patent Examination Process

When an inventor or a team of inventors apply for a patent at the USPTO, the examination process starts at the Office of Patent Application Processing. First, the application is assigned an application number and an "art unit", based on its technological class and subclass. The art unit is a department of patent examiners specializing in a narrowly defined technology field. Once the application is at the art unit, the supervisory patent examiner assigns it to one of the unit's examiners. The assignment is typically based on the number of cases pending with the examiners, the application number or the filing date of the application (Shu, Tian, and Zhan 2021). Over our sample period, there were 10,119 examiners in 667 art units. In a year, the median art unit had 9 examiners and the largest had 61.

While inventors and firms can tailor applications to be sent to a given art unit, they cannot affect the choice of the specific patent examiner who will review their application. Thus, the application assignment to examiners is functionally random (Sampat and Williams 2019), and unrelated to applicant or firm characteristics (Farre-Mensa et al. 2020; Lemley and Sampat 2012). This random assignment of applications to examiners, by art unit and year, has been validated as an effective identification strategy in many recent studies (e.g. Farre-Mensa et al. 2020; Feng and Jaravel 2020; Hegde, Ljungqvist, and Raj 2022; Sampat and Williams 2019) and is also our central identification strategy – letting us causally study CAD through identical inventor groups.

Once the application is assigned to an examiner, the examiner conducts a thorough review of prior "art" or literature to evaluate whether or not the claims meet the legal standards for novelty, usefulness, and non-obviousness. The examiner then sends a "first-action letter", the first communication between the applicant and the examiner, where a nonfinal rejection is issued. As the first indication of success or failure, this letter kickstarts the CAD mechanism. As the final decision typically aligns with the first-action letter (Carley, Hedge, and Marco 2015), it would not signal any additional information.

Therefore, we use the "first-action date" to compare future outcomes, a pivotal information point in the patenting literature where applicants first learn the examiner's identity and can thereon revise their claims. In contrast, the final-decision date is endogenous, as inventors can influence the date by choosing how they respond to the first-action letter, blurring the causal effect of their initial outcome on their future outcomes.

In our sample, examiners on average took almost 1.5 years to send the first-action decision and for applications that are eventually granted, the final decision to took about 2.8 years, i.e. about 1.3 years after the first-action decision.

Data Source and Construction

We have created a detailed dataset tracking individual inventors throughout their careers, regardless of the organizations they work for. To study the impact of an inventor's gender and initial success on future innovation, we required a chronological record of all their patent applications, both successful and unsuccessful. While successful patents are public, unsuccessful patent applications were traditionally not disclosed. The American Inventors Protection Act (AIPA), November 2000, changed this by mandating publication of all patent applications, except those important for national security, or classified as non-provisional or design patent applications. The different types of patent applications are detailed in Appendix D, along with all supplementary materials for this study.

We focus on first-time utility applications filed between November 2000, and April 2020 to ensure data completeness. Other application types—such as provisional, PCT, design, and plant patents—are excluded for reasons detailed in the Appendix. Utility patents are the most common type and are regarded as representative of the patenting process in general (Lemley and Sampat 2012), as they make up most of USPTO's published application data.

Our final sample builds on the PatentsView's inventor disambiguation and gender attribution algorithms to identify 486,144 inventors, who submitted over 2.8 million non-provisional utility applications over their careers. To ensure comprehensive data coverage and validity, we identified application outcomes using both PatentsView and PatEx, managed by the USPTO. The Appendix section on data construction further details these datasets. Robustness checks were conducted when variable-related information was available on both datasets. A detailed description of variables and sources are outlined in Appendix Table D1.

Time horizon & future career outcome

As we study inventor careers we limit our sample of inventors to those who file at least two patent applications at least a year apart. This helps exclude one-time inventors while still including those who entered innovation but exited early. On average, sampled inventors have about 6 years between their first and last application—referred to as career length—with values ranging from 1 to about 20 years. Over 75% of inventors have careers lasting around 7 years, as shown by the 75th percentile at 7.6 years. Thus, longer career horizons reduce the number of inventors available for analysis. For example, while 486,144 inventors have observable outcomes after one year, only 1,363 can be observed after twenty years.

Based on these patterns, our main analysis focuses on inventors' outcomes 7 years after receiving their first-action letter. Additionally, to study the trend hypotheses, we consider future outcomes after 3, 5, 7, and 9 years, i.e. how CAD effects evolve with time. As a robustness check, we also measure future outcomes using the number of patents successfully granted to the inventor over the same time periods.

Summary Statistics

Table 5.2 presents descriptive statistics for our first-time inventors, broken down by gender, initial patent application outcome, and their interaction. Table 5.3 provides additional statistics on examiner and art-unit characteristics pertaining to the inventors' first application evaluation.

Table 5.2 shows how many inventors collaborated on their first application: 486,144 inventors filed 375,932 first-time applications, with over 78% resulting in at least one granted patent. With men making over 84% of inventors, SM formed the majority group (~67%) and UW formed the smallest group (~4%). Men also had a slightly higher rate of initial success on their applications than women (79% vs 75.8%), see Table D2.

Panel-A presents unadjusted future outcomes seven years after an inventor's first patent decision, while Panel-B shows its evolution over time. Gender inequality can already be seen in Panel-A with men on average submitting more applications, securing more patents and subsequently enjoying higher approval rates. The panels also show signs of CAD interactions benefitting some groups. For example, SW outperforms the overall group of women, suggesting that initial success may help offset gender disadvantages. A similar pattern is seen when comparing UM with U. However, as SW and UM are sub-groups of W and U, such comparisons can be misleading.

Figure 5.2 presents unadjusted group differences from Table 5.2, plotting average differences in patent applications across the four inventor sub-groups (SM, SW, UM, UW), i.e. through their seven possible comparisons from Table 5.1. While we formally test our hypotheses later, Figure 5.2 shows how even simple averages align with most of our expectations.

The negative interaction, H_1, appears as the red line below zero, declining over time, consistent with H_{1t}. The remaining hypotheses, H_2-H_7, predict zero or positive estimates by year seven, supported by all other lines hovering around or above the gray zero line.

Support for trend hypotheses comes from how comparisons evolve. Strengthening is indicated by positive values growing or negative ones decreasing, as with comparisons 2, 4, 5, and 7, while weakening shows the opposite pattern, as in comparisons 3 and 6. These patterns generally align with expectations, except for comparison 5, which unexpectedly strengthened instead of weakening, and comparison 3, which initially strengthened as expected but weakened later.

While Figure 5.2 summarizes group differences based only on descriptive statistics, we will use a similar visualization of regression-based estimates later when formally testing our hypotheses.

Table 5.2 **Summary Statistics for Main Variables**: The table reports summary statistics for first-time patent inventors, broken down in three ways: (i) whether they are men (M) or women (W), (ii) whether their first application was successful (S) or unsuccessful (U), and (iii) the combination of the two, i.e. whether they are successful men (SM), successful women (SW), unsuccessful men (UM), or unsuccessful women (UW). The variable definitions and construction is detailed in the Appendix.

			Inventors who are		**Inventors whose 1st patent application was**					
	Units	**All (N)**	Women (W)	Men (M)	Unsuccessful (U)	Successful (S)	**Successful Men (SM)**	**Successful Women (SW)**	**Unsuccessful Men (UM)**	**Unsuccessful Women (UW)**
No. of inventors	N(%)	486144	15.550%	84.450%	20.870%	79.130%	67.332%	11.798%	17.118%	3.752%
No. of applications	N(%)	375932	17.257%	82.743%	21.595%	78.405%	65.346%	13.073%	17.397%	4.184%
No. of patents	N(%)	385532	14.909%	85.091%	0.000%	100.000%	85.091%	14.909%	0.000%	0.000%
Avg. Career Length	Years	5.81	5.357	5.894	5.843	5.801	5.882	5.340	5.938	5.410
Panel-A : Future Outcomes: Seven Years after Inventor's First Application's First-Action Decision										
No. active inventors	N(%)	167753	13.608%	86.392%	21.042%	78.958%	68.658%	10.300%	17.734%	3.308%
No. subsequent patent applications	Mean	9.587	9.395	9.617	8.689	9.826	9.849	9.672	8.717	8.534
	SD	*14.548*	*14.941*	*14.485*	*15.433*	*14.294*	*14.134*	*15.313*	*15.738*	*13.682*
No. subsequent approved patents	Mean	5.986	5.792	6.016	4.043	6.476	6.491	6.378	4.084	3.819
	SD	*8.019*	*8.039*	*8.015*	*7.026*	*8.177*	*8.144*	*8.392*	*7.146*	*6.330*
Approval rate of subsequent applications	%	77.035%	76.178%	77.167%	54.943%	82.241%	82.131%	82.991%	55.457%	52.132%
Panel-B : Future Outcome Trend: Patent Applications Filed Said Years after N Inventor's First Application's First-Action Decision										
No. subsequent patent applications over the next ...										
... Three years	N(%)	430999	66143	364856	90605	340394	290277	50117	74579	16026
	Mean	3.014	2.883	3.038	2.654	3.11	3.13	2.992	2.677	2.543
	SD	*4.923*	*4.956*	*4.916*	*4.95*	*4.911*	*4.891*	*5.021*	*4.996*	*4.73*
... Five years	N(%)	268592	38846	229746	56993	211599	182203	29396	47543	9450
	Mean	5.813	5.585	5.851	5.149	5.992	6.025	5.788	5.188	4.952
	SD	*9.189*	*9.069*	*9.209*	*9.543*	*9.083*	*9.049*	*9.291*	*9.77*	*8.307*
... Nine years	N(%)	112222	14507	97715	23671	88551	77645	10906	20070	3601
	Mean	13.834	13.8	13.839	12.667	14.146	14.14	14.188	12.674	12.627
	SD	*20.931*	*21.988*	*20.77*	*23.163*	*20.282*	*19.963*	*22.423*	*23.598*	*20.573*

Figure 5.2 Seven possible comparisons of the four CAD interaction groups—successful men (SM), successful women (SW), unsuccessful men (UM), and unsuccessful women (UW)—quantifying the mean difference in patent applications submitted seven years after the inventor's first application decision. The trends visualize these comparisons also at the third, fifth, and ninth years.

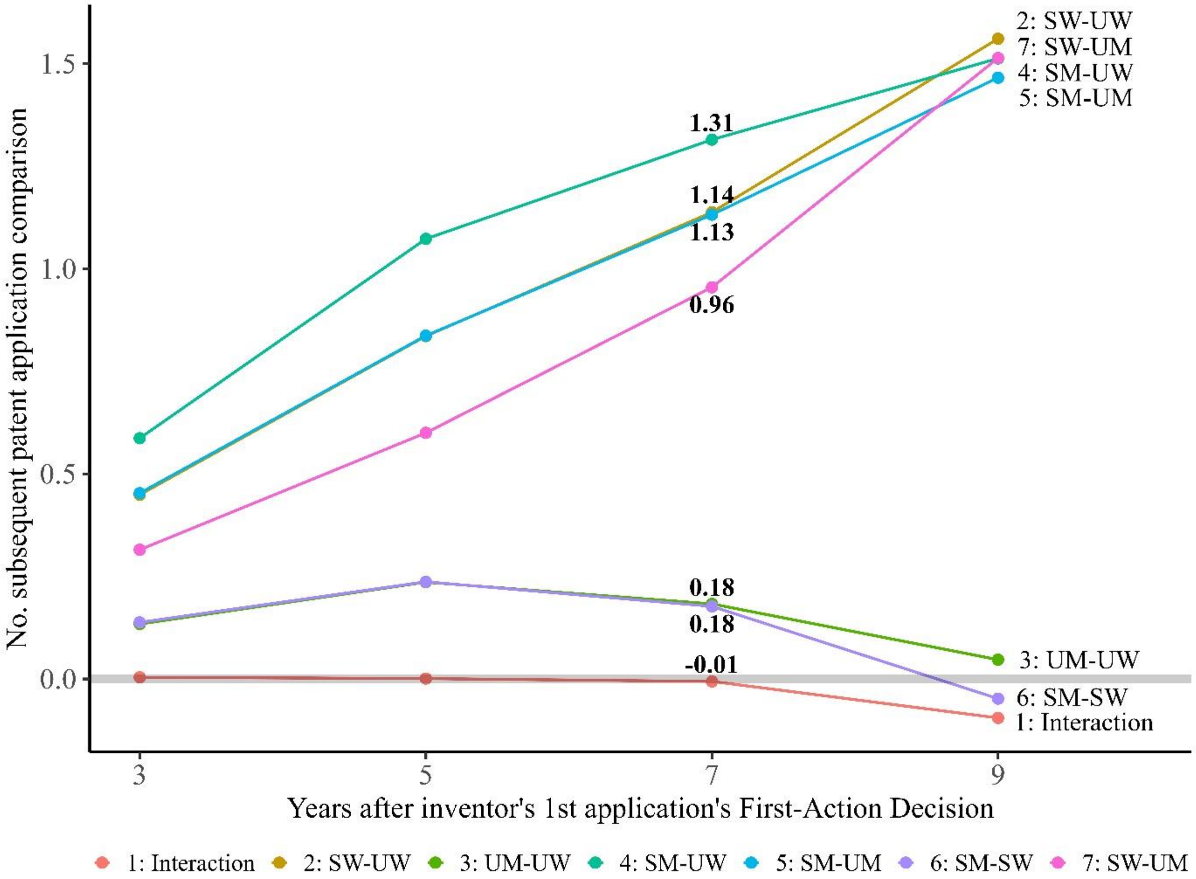


Table 5.3 **Summary Statistics for Other Variables**: The table reports summary statistics for other variables associated with the examination of the first-time patent inventors patent applications.

	N	Mean	SD	Min	Max
Panel-A : Sample Statistics					
No. of examiners	10119				
No. of art units	667				
Examiner experience when application is evaluated	486144	629.532	680.866	10	5949
Examiner leniency when application is evaluated	486144	72.807	21.499	0	100
Panel-B : Art unit-year statistics					
Examiners	104045	9.376	6.98	1	62
Applications evaluated	401728	36.201	41.781	1	390

5.4. ESTIMATION STRATEGY

The ideal experiment to identify the causal effect of CAD from inventors' initial outcome and gender on their future outcome would randomize initial patent approvals. This would ensure that successful applicants and their inventions do not differ systematically from unsuccessful ones ex ante. Thus, our key empirical challenge is disentangling an application's success, from the quality and other unobserved patent characteristics which might influence a patent examiner's grant decision. We tackle this by exploiting the quasi-random assignment of patent examiners to applicants within each art unit. Specifically, we use the instrumental variables (IV) approach proposed by Farre-Mensa et al. (2020), where the instrument measures the "leniency" or the propensity of the examiner to grant a patent.

Imagine two nearly identical inventors, A and B, both applying for patents on "automobile parts." Their applications go to the same "vehicles" art unit but are randomly assigned to different examiners. Examiner 1 is more lenient and more likely to approve patents, while Examiner 2 is stricter. So, A has a higher chance of success just by getting Examiner 1. This random assignment makes it possible to study how early success or failure—caused by chance—can lead to long-term career inequality through CAD.

We use a 2SLS instrumental variable regression strategy - the recommended strategy in the methodology literature to obtain consistent model parameter estimates when dealing with endogeneity (Baltagi 2021). Examiners' quasi-random assignment to patent applications and variation in their interpretation of rules make examiner leniency a viable IV. Previous studies also find Examiner leniency to be a valid instrument for patent application approval (e.g. Gaulé 2018), allowing us to consistently estimate coefficients in the 2SLS equation:

$$\begin{aligned}\mathrm{FutureOutcome}_{\mathrm{ijat+k}} &= \beta_1 \mathrm{E}\left[{}^{PatentApproval}/_{FirstAction\ decision}\right]_{ijat} + \beta_2 \mathrm{Gender}_i \\ &+ \beta_3 \mathrm{E}\left[{}^{PatentApproval}/_{FirstAction}\, decision\right]_{ijat} \times G\mathrm{ender}_i + \Phi \mathrm{X}_{ijat} \\ &+ \nu_{a\tau} + \varepsilon_{ijat+k}\end{aligned}$$

____(Equation 1)

where i refers to the first-time inventor, j the examiner who is from art unit a, who evaluates the application filed at year τ, and t is the year of the first-action date. Outcomes are measured over course of inventor careers and reported here for k = 3,5, 7, and 9 years following the first-action decision that the employee receives in year t.

Examiner leniency is measured as the approval rate of examiner j belonging to art unit a and assigned to review inventor i's patent application submitted at time τ . This enters at the first stage of the above 2SLS equation:

$$\mathrm{PatentApprova}l_{\mathrm{ijat}} = \theta \mathrm{Examiner\ Lenienc}y_{ija\tau} + \Pi \mathrm{X}_{ijat} + \nu_{a\tau} + u_{ijat}$$

____(Equation 2)

X_{ijat} is the vector of controls detailed alongside all variables in Appendix Table D2. As inventors could influence the patent prosecution process through the technology of the patent and the time they file their application, we always include $\nu_{a\tau}$, examiner art unit × application year fixed effects. Gender's influence on initial outcomes is also accounted at the first stage of our 2SLS regression strategy, alongside its subsequent influence with initial outcomes in the second stage. The importance of this step is reflected in the summary statistics, where we could already see a gender difference in inventors' initial success rates, with the proportion of successful applications being higher for men.

PatentApproval is measured as successful or unsuccessful based on the final decision made for the application, which could occur after the first-action date. However, as discussed above, we base the time on the first-action letter which is highly predictive of final decision.

The above equations are used to predict and test the seven comparison estimates we theorized. The first, interaction effect, is directly given by the regression coefficient β_3 from *Equation 1*. The remaining six pairwise comparisons are estimated as average adjusted predictions using the `marginaleffects` package in R, the recommended approach when focusing on predicted values (Arel-Bundock et al. 2024). The package also allows testing whether all seven comparison estimates are significantly different from zero, using the conventional null hypothesis significance testing (NHST). However, it is impossible to use this approach to test the absence of a difference, as in H_5, H_6, and H_7. A p-value above the 5% level implies the lack of evidence on a difference, it is not evidence on the absence of a difference. Such zero effect hypotheses can instead be tested by looking how the CIs of the two groups overlap (Garofalo et al. 2022). When there is considerable overlap, over 25% of the full CI length, there would be reasonable evidence of no difference between the population means.

We use the NHST wherever possible as it directly tests the comparison estimate, as against the CI overlap approach which looks at the distribution of the two estimated group means. Thus, H_5, H_6, and H_7 are first tested using NHST. If the comparison estimate is found to have a p-values<0.05, it is taken as evidence of a significant difference between the groups. If not, i.e. p-value>0.05, we visually inspect the CI overlap. If there is considerable overlap it would be evidence of a zero comparison estimate or similar future outcome for the two groups in consideration.

5.5. RESULTS

CAD Effects After Seven Years

Figure 5.3 and Table 5.4 summarize the results of our seven main hypotheses, showing how CAD from inventors' gender and initial success influence their future outcomes. Figure 5.3 shows estimated group means and their 95% CI, while Table 5.4's Panel-A presents group comparison estimates alongside their significance at the 0.1%, 1% and 5% level. The resemblance between Figure 5.1 and Figure 5.3 already reflects general support for our theory.

The converging lines of Figure 5.3 support the negative interaction hypothesis (H_1). This implies that a mechanism of compensatory advantage is at play, where men are less affected by their prior negative outcomes than women, i.e. $SM - UM < SW - UW$. The consequent steeper slope theorized for women in Figure 5.1 is confirmed in Figure 5.3 and quantified by the significant interaction coefficient of -4.486 in Table 5.4's Panel-B.

The interaction is reflected by two equivalent DiD comparisons. The first is the SM-UM and SW-UW comparisons (0.982-5.468=-4.486), corresponding to comparison-5 and 2. While initial failure drastically reduced the number of future applications for women (H_2)—by 5.468 compared to SW—it had no significant impact on men (H_5), as shown by their non-significant difference of 0.982 and considerable CI overlap in Figure 5.3. This is especially striking given that the average inventor in our data participates in only about nine applications.

The second DiD comparison refers to the SM-SW and UM-UW (-0.852-3.633≈-4.486), corresponding to comparison-6 and 3. These comparisons reflect the gender gaps in reward and punishment, which appear to be eliminated and widened, respectively.

These findings on the interaction effect have two key theoretical implications. First, it supports our theory on UW having the greatest disadvantage, as specified with H_2-H4. Figure 5.3 visually illustrates this with UW's outcome being significantly lower than those of all other groups, with no overlapping CI. Table 5.4 reconfirms this, showing all other groups to have a significant advantage over UW. SW had the greatest advantage—participating in more than twice as many applications on average as UW—highlighting the criticality of initial success for women.

Note that UW has the narrowest CI in Figure 5.3, despite being smallest group per Table 5.2. This implies that UW experienced the least variation in future outcomes, consistently accumulating disadvantages across contexts. In contrast, other groups, including SW, showed greater variability. Thus, while initial success can improve women's outcomes, its effects are uneven. This variation may reflect the mixed responses women face to success, as they are often both rewarded and penalized (e.g. Rudman et al. 2012). Meanwhile initial unsuccessful women consistently face clear disadvantages, indicating a strong gender punishment gap.

Second, our results show how CAD interactions can reduce inequality by creating compensatory advantages for certain groups, specifically UM and SW. Men can compensate disadvantages from initial failure: UM can leverage the advantages of their gender status to improve outcomes. This is evident from the SM-UM comparison (H_5) where no difference was found. In other words, men can recover from initial setbacks, matching the outcomes of their successful male counterparts in seven years. Figure 5.3 illustrates this through the flattened blue line, where SM and UM have nearly identical outcomes with considerable overlap in their CIs.

For women, the negative interaction effect indicates how initial success can drastically reduce their disadvantage from gender inequality: SW can compensate the disadvantage of being a woman with the advantages of being initially successful to improve their future outcomes. This was further examined through SM-SW(H_6) and SW-UM(H_7) comparisons, with both suggesting reduction in gender inequality. The first comparison shows women's initial success to not just close the gender gap but also slightly reverse it, with SW on average submitting 0.852 applications more than SM. The second comparison found a non-significant difference between SW and UM, with considerable overlap seen among their CI in Figure 5.3. This aligns with our earlier observation of UW being the most and perhaps only disadvantaged group, as both UM's and SW's compensatory advantage enable them to reduce inequalities in comparison to SM.

Overall, all hypotheses were supported with the exception of H_6. This is reflected by the similarity of Figure 5.3 with the theorized Figure 5.1, differing only in $SW > SM$ as against our Figure 5.1 where H_6: $E(SM) \geq E(SW)$. While we expected SW to reduce or eliminate their initial disadvantages in comparison SM, SW appears to completely overcome their initial disadvantage and even gain a slight relative advantage—eliminating the gender reward gap.

Figure 5.3 CAD interaction: Inventors' predicted future outcomes after 7 years, based on the 2SLS IV regression from equation 1.

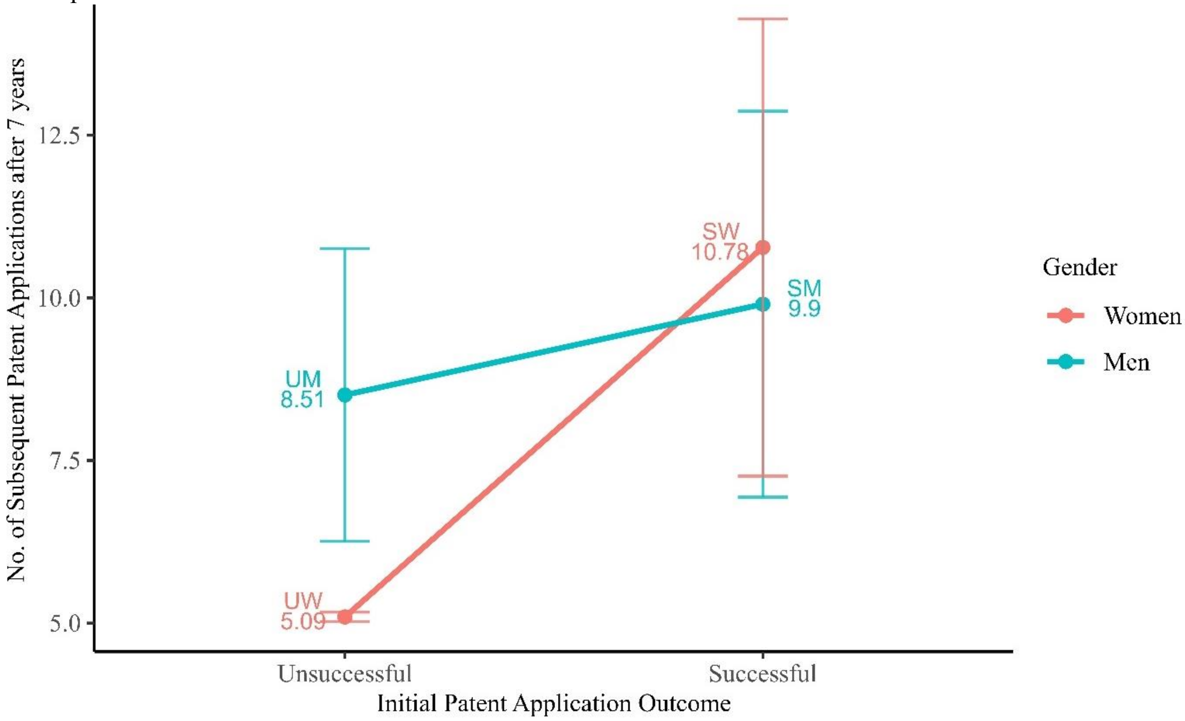


Table 5.4 Seven comparison and the main regression coefficient estimates from equation 1 using the number of subsequent applications submitted by an inventor 7 years after receiving their first-action decision. Panel-A presents the seven comparison estimates based on the corresponding regression coefficients from Panel-B which presents the main second stage results of equation 1. All variables are defined in Table D1. Standard errors reported in parentheses are clustered at the art unit-year level.

Hypothesis		**No. subsequent patent applications filed Seven Years after Inventor's First-Action Decision**
Panel-A : Comparison Estimates Predicted		
1: Interaction Effect	< 0 (H_1)	-4.486** (1.511)
2: SW-UW	> 0 (H_2)	5.468** (1.815)
3: UM-UW	> 0 (H_3)	3.633** (1.150)
4: SM-UW	> 0 (H_4)	4.616** (1.540)
5: SM-UM	≥ 0 (H_5)	0.982 (1.112)
6: SM-SW	≥ 0 (H_6)	-0.852* (0.384)
7: SW-UM	$= 0$ (H_7)	1.835 (1.161)
Panel-B : Regression Output		
Initial outcome: Successful		5.468** (1.815)
Gender: Man		3.633** (1.150)
Successful x Man		-4.486** (1.511)
Control: Examiner experience		Yes
Fixed-Effects: Examiner art unit x application year		Yes
Observations		167,753

R2	0.074
Within R2	-0.003
F-test (IV only)	5.911
Wald (IV only), p-value	0.002

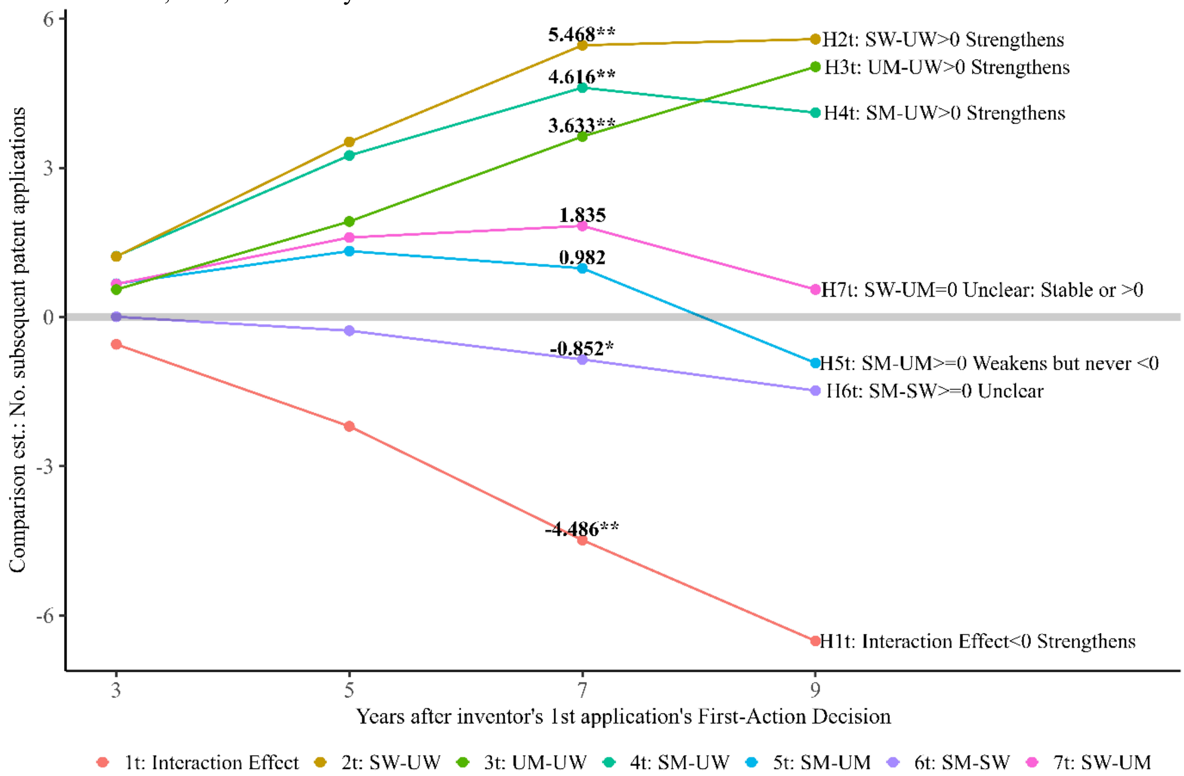
Signifiance codes: p < 0.05 (*), p < 0.01 (**), p < 0.001 (***)

As a robustness check, in addition to the above future outcome measure based on applications we also used additional patent based measures. Specifically, we used the number of subsequent patents granted to an inventor 7 years after receiving their first-action decision, along with the natural log of both the patent and application measures. These results continued to align with our theory and were largely similar to the results above. See Table D3 for details.

CAD Trend: Evolution Over Three, Five, Seven, and Nine Years

Figure 5.4 and Table 5.5 present the trend hypotheses results for our seven comparisons. Figure 5.4 illustrates how the seven comparison estimates evolve over time with their corresponding hypotheses labeled at the end of each line, for ready reference. Generally, points away from the gray zero line indicate high inequality comparisons and points hovering around the gray zero line indicate low inequality comparisons. Thus lines moving away from 0 imply that a comparison is strengthened, and conversely lines moving towards 0 imply weakening. Note that the labeled estimates for year 7 are the same as the above results `-A, which are extended to other time points in Table 5.5. Just as above, Panel-A lists the comparison estimates of Figure 5.4, alongside their significance and standard errors, and Panel-B reports the underlying regression coefficients.

Figure 5.4 Seven possible comparisons of the four CAD interaction groups—successful men (SM), successful women (SW), unsuccessful men (UM), and unsuccessful women (UW)—quantifying the predicted difference in patent applications submitted seven years after the inventor's first application decision. The plot visualize these estimates also at the third, fifth, and ninth years.



All our hypotheses were generally supported. As expected, H_{1t}, H_{2t}, H_{3t}, and H_{4t} all generally strengthen, with their corresponding lines generally moving away from the gray zero-line. Similarly, H_{7t}, H_{5t}, and H_{6t} hover around 0, indicating weakening, stability or equality over time. The falling red line reflects strengthening of the negative interaction effect, H_{1t}, or the compensatory advantage of gender with time. This implies that over time the impact of initial unsuccess on future outcome will reduce for men, but increase for women.

This trend is further understood by looking at how the gender gaps in reward and punishment evolve. The reward gap disappears, as indicated by the hovering of H_{6t} around 0, with gender inequality reducing among those who had initial success, narrowing the gap between SW and SM. However the punishment gap grows, indicated by a rising H_{3t} where over nine years UM are seen to submit about 5 or 1.5 times more applications than UW.

If this punishment gap followed from UW's subsequent lack of motivation or poor performance, over time they would either quit or be laid off. This implies that they would have a considerably higher dropout rate over time. However, a shorthand estimation of dropout rates indicated otherwise, see Table D5. SW and UW continued to have similar dropout rates— suggesting that UW remained in the race despite being part of fewer patent applications than any other group. In fact the only differential pattern found was that women in general had higher dropout rates than men, suggesting women's accumulation of disadvantages to be systematic rather than due to relatively poor performance.

We also find support for our theory on UW being the most disadvantaged group, with its relative disadvantages increasing over time. This can be seen in Figure 5.4 and Table 5.5, where all comparisons with UW, i.e. 2, 3, and 4, widen significantly over time. The stark disadvantage of UW relative to the other three groups was also seen with the evolution of their estimated group means and their overlapping CI in Appendix Figure D2. All time points reflected a similar pattern as Figure 5.3, with UW having the lowest estimate with CIs that do not overlap with any other group. While the estimated group means of all other groups come closer together over time, UW's remained considerably lower.

While we expected H_{5t} to never become negative, i.e. SM would always have more or the same outcomes as UM, we see the blue line becoming negative in year nine. However this difference is not significantly different from 0, as shown in Table 5.5's Panel-A. Further examination of the trend in SM and UM's CI overlap, provided evidence of the difference in their outcomes to disappear in seven years, as discussed above for H_5. This absence of a difference continued even in nine years and is illustrated in Figure D2. This strongly supports our theory on how men can reduce disadvantages from their initial failure over time such that they can completely overcome it.

One might argue that SM's convergence with UM and SW is due to a ceiling effect, where SM's innovation growth slows. However, this is unlikely. First, most patents are produced by teams (Wuchty, Jones, and Uzzi 2007), allowing senior inventors to join more projects as their careers progress, not less. Second, our above robustness checks, using logged outcomes to account for potential ceiling effects, still showed convergence. Hence the convergence of SM, UM and SW's future outcomes is more likely to follow from the compensatory advantage of UM's gender and SW's initial success, rather than decline in SM's innovation.

Table 5.5 Seven comparison and the main regression coefficient estimate trend based on the main second stage results of equation 1. The dependent variable is the number of subsequent applications submitted by an inventor k years after receiving their first-action, where k reflects the different columns of: 3, 5, 7 and 9 years. All variables are defined in Table D1. Standard errors reported in parentheses are clustered at the art unit-year level.

	Future Outcome: Number of subsequent applications submitted after			
Estimate	**3 Years**	**5 Years**	**7 Years**	**9 Years**
Panel-A : Comparison Estimates Predicted				
1: Interaction Effect	-0.547* (0.256)	-2.197*** (0.662)	-4.486** (1.511)	-6.514* (2.982)
2: SW-UW	1.219*** (0.303)	3.525*** (0.790)	5.468** (1.815)	5.591 (3.617)
3: UM-UW	0.554** (0.194)	1.924*** (0.502)	3.633** (1.150)	5.036* (2.273)
4: SM-UW	1.226*** (0.261)	3.252*** (0.677)	4.616** (1.540)	4.113 (3.032)
5: SM-UM	0.672** (0.213)	1.328* (0.532)	0.982 (1.112)	-0.923 (2.004)
6: SM-SW	0.007 (0.067)	-0.273 (0.172)	-0.852* (0.384)	-1.479* (0.747)
7: SW-UM	0.665** (0.213)	1.601** (0.540)	1.835 (1.161)	0.556 (2.178)
Panel-B : Regression Output				
Initial outcome: Successful	1.219*** (0.3033)	3.525*** (0.7898)	5.468** (1.815)	5.591 (3.617)
Gender: Man	0.5541** (0.1938)	1.924*** (0.5020)	3.633** (1.150)	5.036* (2.273)
Successful x Man	-0.5474* (0.2561)	-2.197*** (0.6623)	-4.486** (1.511)	-6.514* (2.982)
Control: Examiner experience	Yes	Yes	Yes	Yes
Fixed-Effects: Examiner art unit x application year	Yes	Yes	Yes	Yes
Observations	430,999	268,592	167,753	112,222
R2	0.06021	0.06017	0.07386	0.08399
Within R2	-0.00109	-0.00351	-0.00264	-0.00195
F-test (IV only)	22.875	16.713	5.9110	2.4898
Wald (IV only), p-value	3.5e-13	6.68e-10	0.00211	0.16099

Signifiance codes: $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***)

As a robustness check, we re-estimated Figure 5.4 and Table 5.5 using patents granted instead of applications submitted and found similar results, see Appendix Figure D3 and Table D4. Thus, we find robust evidence for all our hypotheses.

5.6. Conclusion and Discussion

With workplace gender inequality being largely unexplained and unresolved we studied if and how CAD could explain gender (in)equality in inventor careers. Causally examining population data from the USPTO, we find CAD both widens and reduces inequalities, depending on inventors' gender and their first patent application's success.

An unsuccessful application would typically create cumulative disadvantage that widens inventor inequality over time. However, we find that men can reduce and even eliminate this disadvantage through the compensatory advantage of their gender. In a male-dominated field like innovation, men's dominant position enables them to offset disadvantages with their gender status (Kanter 1977c), for instance through greater access to same-gender mentorship and networking opportunities (Ibarra 1992; Ibarra et al. 2010). This allows initially unsuccessful men to build stronger networks and secure better work opportunities than women. Within seven years, these men had fully overcome their initial setback, with their innovation output matching that of initially successful men.

In stark contrast, women's careers remained closely tied to their initial setback. Initially unsuccessful women had less than half as many applications as successful women, with a gap of over five applications between them. They also lagged behind every other group, with innovation gaps widening over time. Even their smallest gap, nearly four applications compared to initially unsuccessful men, was substantial, given that the average inventor in our data has only about nine applications.

This gender comparison among the initially unsuccessful reflects a growing gender punishment gap. Although initially unsuccessful men and women innovated similarly in their early years, by year nine, men were part of 1.5 times more applications than women. This widening gap cannot be explained by lower effort or motivation among women, as their dropout rates were similar to those of successful women. Thus, the gender punishment gap is indicative of structural inequality rather than differences in resilience or commitment.

We found CAD to reduce inequality for not just initially unsuccessful men, but also initially successful women. Over seven years, the innovation outcomes of initially successful women and unsuccessful men converged with successful men. We even found successful women participating in slightly more applications than both groups of men, at different time points.

This gender comparison among the initially successful reflects the gender reward gap, which, contrary to the punishment gap, reduces and even slightly reverses. This is consistent with the literature on gender pay gap reversals for certain groups, where women are not uniformly disadvantaged with the female penalty reversing and becoming a female premium for High Potential Women (Leslie et al. 2017). This premium is often attributed to selection, with the beneficiaries being women who are exceptionally capable, resilient or committed as they manage to enter and remain in male-dominated professions (Huang et al. 2024; e.g. Kumar 2010). However, such selection effects are unlikely to explain our findings, given that our quasi random estimation strategy controls for inventors capabilities and other unobserved characteristics. One could argue that this argument against selection might not hold over time if women's capabilities fundamentally change after their initial outcome. For instance, over time more women who are initially unsuccessful would leave, as compared to those who were successful. However, this is unlikely, as we find women to have similar dropout rates and career lengths regardless of their initial outcome. Thus, the gender gaps in both punishment and reward reflect structural inequality driven by the CAD interaction mechanism.

Overall, only one group faced widening and irreversible inequality: initially unsuccessful women. This contrasts with Pedulla's (2016) findings, where men in the broader U.S. labor market were penalized more severely for negative outcomes like part-time work due to their rarity, signaling greater incompetence. For women, such outcomes were seen as more common and thus judged less harshly (Pedulla 2016; Spence 1973). In other words, men's larger penalties stem from the rarity or low visibility of their negative outcomes in the public domain. In patenting, however, negative outcomes are publicly visible, making men's unsuccessful applications common enough to weaken their negative signaling power and therefore their penalties. Innovation contexts also tend to frame early setbacks as part of experimentation and learning, especially for men, whose dominant position provides access to stronger support networks that help them recover. Women, lacking such advantages, remain structurally disadvantaged.

These disadvantages reflect a lost opportunity. Despite initially unsuccessful women having similar potential, they were the only group to experience persistent and widening disadvantages. This is crucial, as more women on patent teams boosts economic value, raises invention rates, and steers innovation toward unmet needs, especially for women (Bell et al. 2019; Koning, Samila, and Ferguson 2021; Pairolero et al. 2025). For instance, an AI patent with an equal number of male and female inventors has a value approximately $1.038 million higher than a patent with all-male inventors (Giczy, Pairolero, and Toole 2024). More broadly, increasing the participation of women and minorities in innovation could raise GDP per capita by between 0.6 and 4.4 percent (Hunt et al. 2012; Subramani 2021). Therefore, addressing structural gender inequality is not just socially desirable but also economically beneficial.

Although the gender gap in innovation is narrowing, at the current rate it will take over 110 years to reach parity (Bell et al. 2019). Our findings offer three key policy implications to help accelerate progress toward workplace gender equality.

First, our findings offer new insights by examining CAD under a fair external reward system: the USPTO. Traditionally, policy recommendations from CAD research emphasize establishing fairer external reward systems. This follows from CAD originally being described as a problem in the workings of reward systems, turning up when individuals or organizations take on the job of gauging and suitably rewarding lofty performance on behalf of a large community (Merton 1968). For instance, CAD in science funding largely followed from candidates who won prior awards being evaluated more positively than nonwinners (Bol et al. 2018). To tackle such biases, the Dutch Research Council (NWO) implemented reforms like narrative CVs, which emphasize the work's quality, rather than the individual's quantity of past awards (Kasper Gossink-Melenhorst and NWO 2019). Similar global efforts include the Declaration on Research Assessment, aiming to improve evaluation of scholarly research across disciplines (DORA 2025).

However, we find considerable structural inequality persisting despite USPTO offering a fair external reward system, which evaluates patents strictly based on innovation quality without reference to inventors' past successes or identities. This suggests the critical role played by other factors in driving inequality, such as employees' psychosocial conditions and the subsequent distribution of resources and opportunities (Biegert et al. 2023; Bol et al. 2018). As these other factors operate at a more micro level than external reward systems, understanding them can provide for more accessible and tailored policies.

Second, our findings suggest that gender inequality would be best addressed by offering targeted support to women facing initial setbacks. This stems from the unequal long-term impact of initial failure, with initially unsuccessful women being the only group to accumulate irreversible disadvantages. Such support could be offered at different levels. For example, through fairer appraisal policies within firms or national policies like tax incentives for minority inventor applicants, which are shown to boost corporate innovation (Dechezleprêtre et al. 2023). By showing how inequality grows from early-career outcomes, our findings highlight an opportunity to nip structural gender inequality in the bud—well before it escalates at higher organizational levels.

This leads to our final policy implication on the need to prioritize early-career interventions over later-stage diversity targets. In recent years, there has been a sharp increase in the use of board gender quotas and practices to increase women's managerial presence (OECD 2022). However, these measures are often inefficient as they occur too late, i.e. after early disadvantages have either ended careers or widened inequalities, making subsequent redressals more challenging and costly. Furthermore, such measures may inadvertently worsen inequality by disproportionately benefiting women who have already successfully navigated structural barriers, creating a female premium at the top (Leslie et al. 2017). In our context, such policies would mainly benefit successful women, widening their existing advantage and increase overall inequality. Moreover, this would further harm our most disadvantaged group, widening the gap between unsuccessful and successful women's future outcomes.

The potential effectiveness of early interventions in gendered careers is evident in the success of the USPTO's randomized control trial, which provided additional help to applicants without legal representation. While both men and women benefited, women's likelihood of obtaining a patent increased by over 12 percentage points. This disproportionate benefit reflects women's higher likelihood of lacking legal representation, as they often apply independently, or with small businesses or non-profits. These groups typically lack experience, legal representation and face undue resource constraints in drafting and prosecuting their applications (Pairolero et al. 2025). Thus, the intervention supported not just women but also other groups. such as small businesses and non-profits, that are relatively disadvantaged compared to large, experienced multinational corporations (Hall and Lerner 2010). With increasing backlash against diversity initiatives, such early-career, gender-neutral interventions can be framed as merit-based efforts that benefit all under-resourced groups, not just women.

In addition to the societal contribution, we also make a scientific contribution to the literatures on gender, careers, and stratification by expanding our understanding of CAD. While traditionally seen as a source of inequality, we find that CAD can also reduce inequality for certain groups. To our knowledge, we conduct the first systematic study of CAD interactions. Specifically, we theorize and empirically evaluate the interaction of two CAD factors that are typically studied independently, or at best together, without fully accounting for their interplay due to methodological challenges.

Additionally, we extend career inequality research into the domain of innovation—a valuable context for both career and gender inequality studies. This approach uniquely broadens the scope and improves the external validity of our findings, given its population-level coverage across diverse industries and organizational types, from independent inventors to multinational corporations. Furthermore, inventors face both highly competitive career environments and significant career-family conflicts, making them particularly informative for understanding CAD's role in gender inequality. Since career-family conflict can potentially contribute of workplace gender inequality, finding structural gender inequality among inventors underscores its broader relevance, even for occupations with lower levels of such conflict.

While career-family conflict and inequality has often been studied among academics, studying inventors provides additional methodological benefits. For example, the standardized, quasi-random patent evaluation process enables rigorous causal analysis of CAD, overcoming a key methodological challenge in previous career research. Moreover, standardized criteria for patent quality reduce uncontrolled variance, increasing the reliability of our causal estimates. All issued patents meet the common quality criteria of the invention being new to the world, practically useful, and unobvious to persons reasonably skilled in the art. In contrast, Academic publications have substantial variance from differential quality standards across journals, both within and across fields (Huber 1998; de Solla Price 1963).

However, using patent data also has limitations, especially in defining the start and end of inventor careers. We use a simplified yet consistent measure based the inventors' first and last application dates. While this allows for group comparisons, it may overlook earlier informal or preparatory work. For example, Percy Spencer, a Navy radio operator who later worked at Raytheon, discovered the microwave effect after a chocolate bar in his pocket melting near a radar magnetron. Though he filed a patent later, much of the innovative work began beforehand. Thus his career could be argued to have begun when he noticed the melted chocolate, was reassigned to magnetron research, joined Raytheon, or even during his military service. These varied entry points, while potentially more accurate, are difficult if not impossible to identify for all inventors. Thus, our approach offers a conservative yet consistent estimate suitable for large-scale analysis.

Another limitation is our focus on inventors' patenting careers. This excludes prior experiences which could influence future outcomes. For instance, Spencer's navy experience as a radio operator helped him with his microwave patent. Nevertheless, our 2SLS-IV regression effectively addresses these missing or unobserved variables, obviating the need for extensive controls (Angrist and Pischke 2009).

While we study relative disadvantage through differences in innovative output, other forms of disadvantage, such as exits from innovation careers, remain outside its scope. Nonetheless, our findings suggest the presence of such disadvantages, with women showing slightly but consistently shorter careers and higher dropout rates. Future research could explore CAD's role in employee punishment, instead of reward, by examining negative outcomes like layoffs across different employee groups. This would offer a more nuanced view of the gender punishment gap. This could prove crucial in furthering our understand and in addressing workplace gender inequality, especially given the irreversible and sharp growth of disadvantages we find for initially unsuccessful women.

Further, we consider only the quantity of innovation, not its quality or broader career outcomes like pay or job roles. Future research could integrate these dimensions to provide a fuller understanding of gender inequality in innovation careers. For instance, investigating patent quality and broader career outcomes might clarify why initially unsuccessful women continue in innovation despite filing fewer patents. This may follow from women inventors taking on more administrative tasks, producing higher-quality patents, and receiving less organizational support (Colyvas et al. 2012; Whittington and Smith-Doerr 2005). Future studies could leverage established quality metrics from the patent literature, such as forward citations, generality, originality, and economic value (Farre-Mensa et al. 2020; Sampat 2005), to explore how gender influences career trajectories as well as the broader societal, economic, and individual value of innovations.

Overall, our findings demonstrate how seemingly minor early-career outcomes can trigger deep and lasting gender inequalities, disproportionately penalizing women. Effectively addressing gender inequality thus requires proactive, targeted interventions early in individuals' careers—before initial setbacks become permanent disadvantages. By doing so, policies can prevent structural gender inequalities from taking root, fostering a more equitable and productive innovation environment.

APPENDICES

Appendix A

Supplementary material for Chapter 2: Should Birds of a Feather Flock Together? How Networking with Men and Women Relates to Employees' Career Success

Autonomy scale

The questions below concern the amount of freedom you have when carrying out your work in this organisation.

How often are you free to decide ...

	All the time	Most of the time	Sometimes	Mostly not	Never
The tasks you do in your job	O	O	O	O	O
How you do your work	O	O	O	O	O
The order in which you carry out tasks	O	O	O	O	O
When you do your work	O	O	O	O	O

Table A1 Detailed ANCOVA multilevel model

Level 1 equation:	$Income_{ijk} = \beta_{0jk} + \beta_{1jk}HeterogenousMen_{ijk} + \beta_{2jk}DiverseMen_{ijk} + \beta_{31jk}HomophilousMen_{ijk} + \beta_{4jk}HeterogenousWomen_{ijk} + \beta_{5jk}DiverseWomen_{ijk} + \beta_{6jk}HomogenousWomen_{ijk} + \beta_{7jk}Tenure + \beta_{8jk}EucationalYears + \beta_{9jk}Autonomy + e_{ijk}$ $e_{ijk} \sim N(0, \sigma_e^2)$
Level 2 equation:	$\beta_{0jk} = \beta_{00k} + \beta_{01k}FemaleEmployees + u_{0jk}$ $\beta_{1jk} = HetMen_{100}$ $\beta_{2jk} = DivMen_{200}$ $\beta_{3jk} = HomMen_{300}$ $\beta_{4jk} = HetWomen_{400}$ $\beta_{5jk} = DivWomen_{500}$ $\beta_{6jk} = HomWomen_{600}$ $\beta_{7jk} = \gamma_{700}$ $\beta_{8jk} = \gamma_{800}$ $\beta_{9jk} = \gamma_{900}$ $u_{0jk} \sim N(0, \sigma_{u_0}^2)$
Level 3 equation:	$\beta_{00k} = \gamma_{001}NumberEmployees + \gamma_{002}GDP + v_{0k}$ $\beta_{01k} = \gamma_{010}$ $v \sim Multivariate\ normal\ distributed\ with\ \sigma_{v_0}^2, \sigma_{v_1}^2, and\ \sigma_{v_{01}}$

Combined:	$$\begin{aligned} Income_{ijk} = {} & HetMen_{100}HeterogenousMen_{ijk} \\ & + DivMen_{200}DiverseMen_{ijk} \\ & + HomMen_{300}HomophilousMen_{ijk} \\ & + HetWomen_{400}HeterogenousWomen_{ijk} \\ & + DivWomen_{500}DiverseWomen_{ijk} \\ & + HomWomen_{600}HomogenousWomen_{ijk} \\ & + \gamma_{700}Tenure + \gamma_{800}EucationalYears \\ & + \gamma_{900}Autonomy + \gamma_{010}FemaleEmployee \\ & + \gamma_{001}NumberEmployees + \gamma_{002}GDP + v_{0k} + u_{0jk} \\ & + e_{ijk} \end{aligned}$$

Table A2 Pearson correlation of the dependent variables for the multiply imputed datasets

	Job Outlook	**Income**
Job Outlook	1	
Income	-0.053	1
Job Security	-0.079	0.081

Table A3 Unconstrained Bayes Factors (BFu) for the Competing Hypotheses (CH_{mn}) Versus the Unconstrained Hypotheses (Hu), and their Posterior Model Probabilities (PMP) when using Job Security as the dependent variable

	Model Comparison Results	
	Job Security	
Competing Hypothesis for Men	Bfu	PMP
CH1.M1: Homophily is better for men	0.594	0.208
CH1.M2: Diversity is better for men	1.261	0.442
Hu		0.350
Competing Hypothesis for Women		
CH1.W1: Homophily is better for women	1.300	0.326
CH1.W2: Diversity is better for women	1.094	0.274
CH1.W3: Heterophily is better for women	0.593	0.149
Hu		0.251

Table A4 Complementary Bayes Factors (BFc) for Men's and women's Hypotheses on the Relative Benefits of Homophily and Heterophily in Token Teams when using Job Security as the dependent variable

	Job Security
H2: Token teams	BFc
H2.M: For men homophily is better than Heterophily	1.072
H2.W: For women heterophily is better than homophily	1.182

Table A5 Complementary Bayes Factors (BFc) for the Hypotheses on network returns for men and women in Token Teams when using Job Security as the dependent variable

	Job Security
H3: Non Token teams	BFc
H3.1: Homophily has similar returns for men & women	25.970
H3.2: Diversity has similar returns for men & women	5.100
H3.3: Heterophily has similar returns for men & women	7.638

Table A6 Unconstrained Bayes Factors (BFu) for Men's and women's Competing Hypotheses (CH_{mn}) Versus the Unconstrained Hypotheses (Hu), and their Posterior Model Probabilities (PMP) when additionally controlling for sectors

	Model Comparison Results			
	Job Outlook		Income	
Competing Hypotheses for Men	Bfu	PMP	Bfu	PMP
CH1.M1: Homophily is better for men	0.402	0.121	0.884	0.268
CH1.M2: Diversity is better for men	1.923	0.578	1.419	0.430
Hu		0.301		0.303
Competing Hypotheses for Women				
CH1.W1: Homophily is better for women	0.949	0.233	0.863	0.220
CH1.W2: Diversity is better for women	1.664	0.409	0.549	0.140
CH1.W3: Heterophily is better for women	0.460	0.113	1.505	0.384
Hu		0.245		0.255

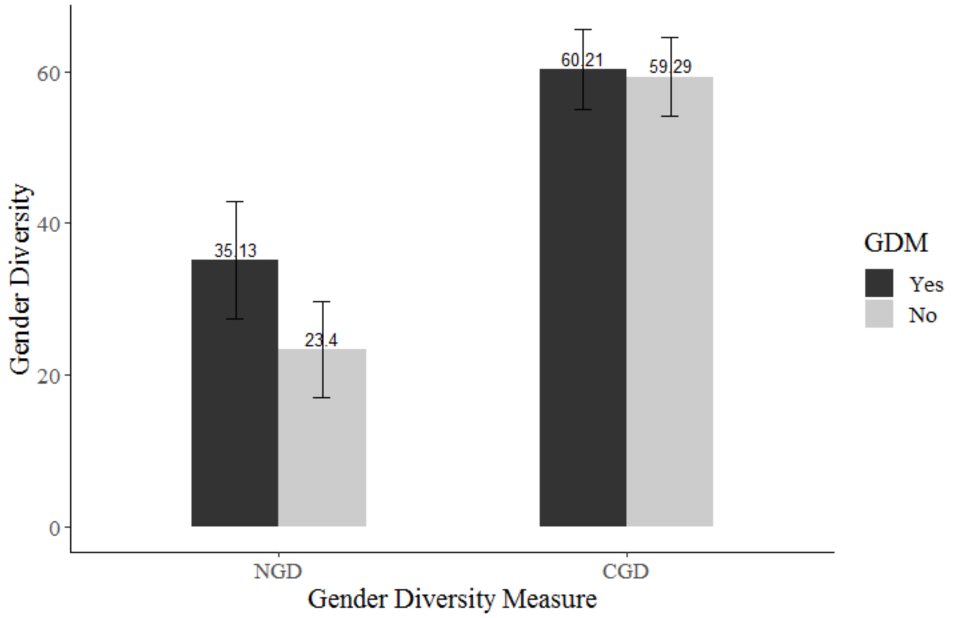
Table A7 Complementary Bayes Factors (BFc) for the Hypotheses on network returns for men and women in Non-Token Teams when additionally controlling for sectors

	Job Outlook	Income
H3: Non Token teams	BFc	BFc
H3.1: Homophily has similar returns for men & women	23.172	5.416
H3.2: Diversity has similar returns for men & women	7.071	2.672
H3.3: Heterophily has similar returns for men & women	20.421	9.145

Appendix B

Supplementary material for Chapter 3: Are Organizational Gender Diversity Management Practices Effective?

Figure B1: Average gender diversity scores on the CGD and NGD for teams with and without GDM along with their 95% confidence interval



Note: Computed using the non-imputed dataset aggregated at the team level and all available information per variable.

Figure B2:Regression coefficients of the effects of having no gender diversity management (GDM) and the three GDM approaches along with their 95% credibility intervals (CI), when separately evaluated on the outcome variables – the current gender diversity (CGD) and the new gender diversity (NGD) measures

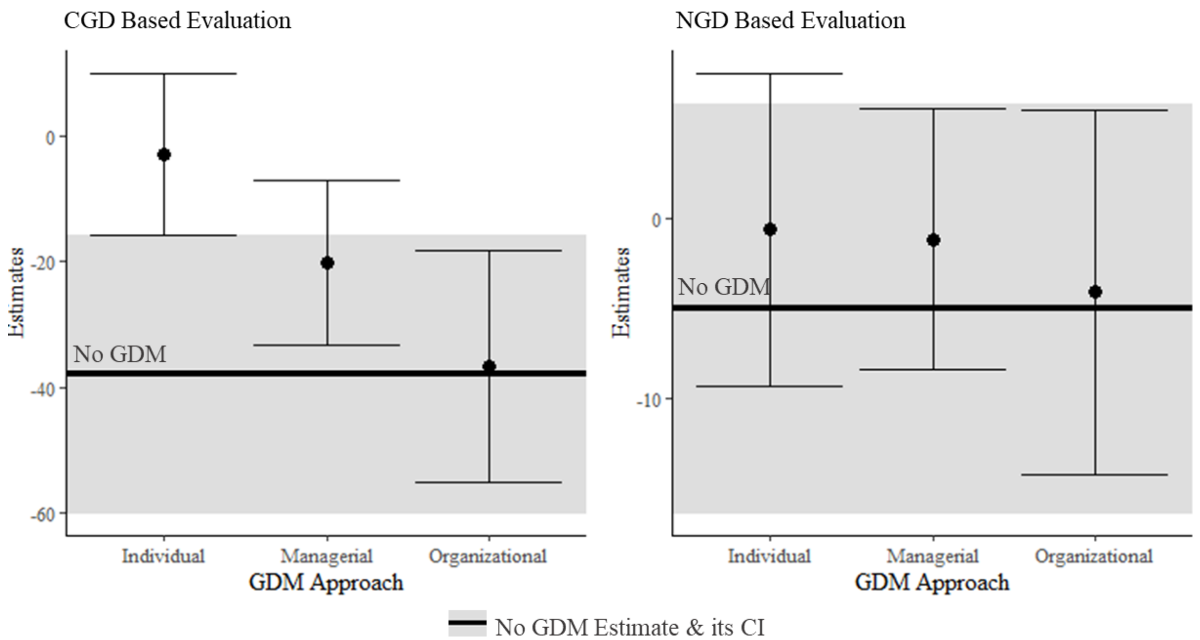


Table B1. Disaggregated descriptive statistics

	Mean	SD	Min	Max	Missing
Key Variables					
New Gender Diversity Measure	27.27	41.12	0	100	0.84
Current Gender Diversity Measure	58.69	37.23	0	100	0.11
Individual Approach Practices	0.18				0.10
Managerial Approach Practices	0.28				0.10
Organizational Approach Practices	0.44				0.10
No GDM Practices	0.50				0.10
Individual Controls					
Partner: Yes	0.73				0.09
Children: Yes	0.37				0.11
Gender: Female	0.59				0.05
Working From Home	1.74	1.43	1	7	0.06
Age	44.34	11.59	19	79	0.06
Educational Years	13.54	3.65	3	21	0.05
Earnings in USD (PPP)	2197.28	3144.23	1.15	135387.83	0.16
Occupational Status	51.69	16.95	16	85	0.18
Team Controls					
Female Employees in Team (Scaled Prop.)	5.20	2.32	1	9	0.30
Function: Core task	0.72				0.00
Function: Supporting task	0.28				0.00
Organizational Controls					
Number of Employees in Establishment	480.69	1154.72	8	9481	0.09
Proportion of Managers	0.11	0.15	0	1	0.10
Flexible Work Arrangements: Yes	0.81				0.11
Ownership: Public	0.36				0.08
Ownership: Private	0.53				0.08
Ownership: Others	0.12				0.08
Sector					
Manufacturing	0.31				0.00
Health Care	0.24				0.00
Higher Education	0.25				0.00
Transport	0.09				0.00
Financial Services	0.06				0.00
Telecommunication	0.05				0.00
Country					
UK	0.04				0.00
Germany	0.06				0.00
Finland	0.03				0.00
Sweden	0.08				0.00
Netherlands	0.20				0.00
Portugal	0.05				0.00
Spain	0.05				0.00
Hungary	0.18				0.00
Bulgaria	0.32				0.00

Note: Means and standard deviations reported using the non-imputed non-aggregated dataset and all available information per variable (4,435 employees in 401 teams and 113 organizations)

Table B2: Average gender diversity scores on the CGD and NGD for teams with No GDM along with those that adopted individual, managerial and organizational approaches to GDM

GDM Approach Adoption	Average CGD Score			
	No GDM	Individual Approach	Managerial Approach	Organizational Approach
No	60.21	60.66	62.82	61.84
Yes	59.28	53.78	51.27	56.50
	Average NGD Score			
No	35.14	28.20	27.93	24.56
Yes	23.40	29.88	29.69	34.40

Note: Computed using the non-imputed dataset aggregated at the team level and all available information per variable.

Table B3. Results of exploratory analysis using Bayes factors between pairs of informative hypotheses (BFne) using alternate framing for ineffectiveness (He)

Hypotheses	CGD	NGD
Individual:		
$H_n: \beta_{ind.y} > \beta_{no.y}$, He; $\beta_{ind.y} \leq \beta_{no.y}$	954.019	3.300
Managerial:		
$H_n: \beta_{mgr.cgd} > \beta_{no.cgd}$; $H_e: \beta_{mgr.cgd} \leq \beta_{no.cgd}$	23.228	2.883
Organizational:		
$H_n: \beta_{org.cgd} > \beta_{no.cgd}$; $H_e: \beta_{org.cgd} \leq \beta_{no.cgd}$	1.420	1.562

Table B4. Results of Bayesian evaluation of informative hypotheses for the complete case analysis (CCA) and alternative CGD measures

Hypotheses	**CCA**	**CGD**	**Alternative CGD Measures** %Women Managers	Normalized HHI: %Women in whole organization	%Women in whole organization	Normalized HHI: %Women in the most senior management levels	%Women in the most senior management levels
Results of main hypotheses evaluation using Bayes factor of the hypothesis at hand versus its complement (BF.c)							
H1: $\beta_{no.cgd} < \beta_{no.ngd}$	3.858	222.141	8.817	2.834	1.569	13.146	3.212
H2: $\beta_{org.cgd} > \beta_{mgr.cgd} > \beta_{ind.cgd}$	0.004	0.000	0.005	0.032	0.002	0.074	0.068
H3: $\beta_{mgr.ngd} > (\beta_{ind.ngd}, \beta_{org.ngd})$	0.107	1.234	1.234	1.234	1.234	1.234	1.234
Results of exploratory analysis using Bayes factors between pairs of informative hypotheses (BFne)							
CGD							
Individual: H_n: $\beta_{ind.cgd} > \beta_{no.cgd}$, He: $\beta_{ind.cgd} = \beta_{no.cgd}$	3.287	36.693	1.291	1.529	0.209	1.421	1.106
Managerial: H_n: $\beta_{mgr.cgd} > \beta_{no.cgd}$; H_e: $\beta_{mgr.cgd} = \beta_{no.cgd}$	0.219	1.399	1.019	0.380	1.976	3.482	0.390
Organizational: H_n: $\beta_{org.cgd} > \beta_{no.cgd}$; H_e: $\beta_{org.cgd} = \beta_{no.cgd}$	0.628	0.194	0.083	0.216	0.060	0.671	0.332
NGD							
Individual:H_n: $\beta_{ind.ngd} > \beta_{no.ngd}$, He: $\beta_{ind.ngd} = \beta_{no.ngd}$	0.298	0.324	0.341	0.358	0.345	0.349	0.346
Managerial: H_n: $\beta_{mgr.ngd} > \beta_{no.ngd}$; H_e: $\beta_{mgr.ngd} = \beta_{no.ngd}$	0.127	0.297	0.312	0.328	0.316	0.320	0.317
Organizational: H_n: $\beta_{org.ngd} > \beta_{no.ngd}$; H_e: $\beta_{org.ngd} = \beta_{no.ngd}$	0.547	0.205	0.216	0.226	0.218	0.221	0.219

Table B5. Results of Bayesian evaluation of informative hypotheses for country following the Jackknife procedure (upon excluding said country)

Hypotheses	**UK**	**Germany**	**Finland**	**Sweden**	**Netherlands**	**Portugal**	**Spain**	**Hungary**	**Bulgaria**
Results of main hypotheses evaluation using Bayes factor of the hypothesis at hand versus its complement (BF.c)									
H1: $\beta_{no.cgd} < \beta_{no.ngd}$	98.310	140.147	148.200	40.295	413.448	136.527	218.936	336.842	56.181
H2: $\beta_{org.cgd} > \beta_{mgr.cgd} > \beta_{ind.cgd}$	0.001	0.001	0.001	0.001	0.001	0.001	0.008	0.000	0.178
H3: $\beta_{mgr.ngd} > (\beta_{ind.ngd}, \beta_{org.ngd})$	1.218	1.028	1.169	0.976	0.983	1.285	0.970	1.196	2.500
Results of exploratory analysis using Bayes factors between pairs of informative hypotheses (BFne)									
CGD									
Individual: $H_n: \beta_{ind.cgd} > \beta_{no.cgd}$, He: $\beta_{ind.cgd} = \beta_{no.cgd}$	27.401	35.471	26.180	10.529	26.387	20.278	7.489	82.145	4.894
Managerial: $H_n: \beta_{mgr.cgd} > \beta_{no.cgd}$; $H_e: \beta_{mgr.cgd} = \beta_{no.cgd}$	0.954	0.910	1.257	0.917	2.799	1.181	0.741	20.256	0.710
Organizational: $H_n: \beta_{org.cgd} > \beta_{no.cgd}$; $H_e: \beta_{org.cgd} = \beta_{no.cgd}$	0.191	0.248	0.199	0.124	0.265	0.204	0.282	0.072	13.448
NGD									
Individual: $H_n: \beta_{ind.ngd} > \beta_{no.ngd}$, He: $\beta_{ind.ngd} = \beta_{no.ngd}$	0.302	0.347	0.377	0.482	0.356	0.423	0.151	0.343	0.377
Managerial: $H_n: \beta_{mgr.ngd} > \beta_{no.ngd}$; $H_e: \beta_{mgr.ngd} = \beta_{no.ngd}$	0.284	0.280	0.324	0.370	0.301	0.363	0.181	0.304	0.530
Organizational: $H_n: \beta_{org.ngd} > \beta_{no.ngd}$; $H_e: \beta_{org.ngd} = \beta_{no.ngd}$	0.208	0.202	0.229	0.215	0.294	0.197	0.274	0.176	0.274

Table B6. Results of Bayesian evaluation of informative hypotheses for sector following the Jackknife procedure (upon excluding said sector)

Hypotheses	**Manufacturing**	**Health**	**Higher Education**	**Transport**	**Financial Services**	**Telecommunication**
Results of main hypotheses evaluation using Bayes factor of the hypothesis at hand versus its complement (BF.c)						
H1: $\beta_{no.cgd} < \beta_{no.ngd}$	45.311	3587.779	7.365	77.984	31.232	21.719
H2: $\beta_{org.cgd} > \beta_{mgr.cgd} > \beta_{ind.cgd}$	0.010	0.000	0.002	0.001	0.064	0.000
H3: $\beta_{mgr.ngd} > (\beta_{ind.ngd}, \beta_{org.ngd})$	0.957	1.776	0.708	1.295	0.810	1.658
Results of exploratory analysis using Bayes factors between pairs of informative hypotheses (BFne)						
CGD						
Individual: $H_n: \beta_{ind.cgd} > \beta_{no.cgd}$, He: $\beta_{ind.cgd} = \beta_{no.cgd}$	2.446	360.926	1.213	40.331	0.918	23.322
Managerial: $H_n: \beta_{mgr.cgd} > \beta_{no.cgd}$; $H_e: \beta_{mgr.cgd} = \beta_{no.cgd}$	3.432	18.235	0.383	0.933	0.230	0.268
Organizational: $H_n: \beta_{org.cgd} > \beta_{no.cgd}$; $H_e: \beta_{org.cgd} = \beta_{no.cgd}$	0.195	0.186	0.078	0.344	0.525	0.344
NGD						
Individual: $H_n: \beta_{ind.ngd} > \beta_{no.ngd}$, He: $\beta_{ind.ngd} = \beta_{no.ngd}$	0.311	0.283	0.440	0.389	0.154	0.354
Managerial: $H_n: \beta_{mgr.ngd} > \beta_{no.ngd}$; $H_e: \beta_{mgr.ngd} = \beta_{no.ngd}$	0.267	0.360	0.271	0.335	0.173	0.361
Organizational: $H_n: \beta_{org.ngd} > \beta_{no.ngd}$; $H_e: \beta_{org.ngd} = \beta_{no.ngd}$	0.229	0.207	0.193	0.240	0.349	0.173

Appendix C

Supplementary material for Chapter 4: Studying Interaction of Cumulative (Dis)Advantage Factors: What Have We Done?!

Appendix C1. Supplementary Simulation Results

IE is generally more prone to being undetected than MEs

Table C1 Comparison of Type-II error rates between the IE and the MEs for all datasets, including those where they were constructed to have different effect sizes

Data Group	Type-II Error					
	Marginal Means Approach			Regression Coefficient Approach		
	IE	ME x1	ME x2	IE	ME x1	ME x2
	DiD	PC:00-10	PC:00-01	$\hat{\beta}_3$	$\hat{\beta}_1$	$\hat{\beta}_2$
All	28.72%	19.02%	19.00%	28.67%	18.98%	18.97%
Homogeneity						
Heterogeneous	17.42%	10.43%	10.36%	17.37%	10.37%	10.33%
Low	25.43%	17.79%	17.83%	25.39%	17.76%	17.78%
High	34.84%	22.38%	22.34%	34.79%	22.35%	22.31%
IE Size						
Large IE	0.03%	18.96%	19.14%	0.03%	18.92%	19.11%
Medium IE	18.10%	19.01%	18.89%	18.04%	18.98%	18.85%
Small IE	68.03%	19.13%	19.01%	67.95%	19.09%	18.97%
IE Valence						
Positive IE	28.78%	18.99%	18.86%	28.73%	18.96%	18.83%
Negative IE	28.67%	19.07%	19.16%	28.62%	19.03%	19.13%
No IE	-	18.92%	18.96%	-	18.88%	18.91%
MEs Present						
None	28.75%	-	-	28.69%	-	-
One	28.58%	19.17%	19.01%	28.54%	19.14%	18.97%
Both	28.99%	18.86%	19.00%	28.93%	18.82%	18.97%
ME Size						
Large ME	28.73%	0.00%	0.00%	28.68%	0.00%	0.00%
Small ME	28.71%	38.03%	38.01%	28.66%	37.96%	37.93%

Note: Most data groups continue to have higher Type-II error for the IE. Exceptions follow from the effect size and have been highlighted. Data with large and medium IEs, and small MEs had MEs with a higher Type-II error rate than IEs. This highlights the importance of considering effect size in interaction studies.

False detection of MEs spikes with step-wise analysis

Table C2 Type-I and Type-II error for MEs of x_1 and x_2, estimated using short and long model with coefficients β_1 and β_2, for all 441000 datasets

Data Group	Short Model: Without Interaction					Long Model: With Interaction [Full Model]				
	x_1 Error Type		x_2 Error Type		No. Datasets	x_1 Error Type		x_2 Error Type		No. Datasets
	I	II	I	II		I	II	I	II	
All	60.03%	12.13%	60.08%	12.20%	441000	5.02%	19.02%	4.98%	19.00%	441000
Homogeneity										
Heterogeneous	71.53%	10.69%	71.61%	10.76%	49000	5.26%	10.43%	5.09%	10.36%	49000
Low	62.75%	11.42%	62.83%	11.53%	196000	5.03%	17.79%	4.95%	17.83%	196000
High	54.44%	13.21%	54.44%	13.23%	196000	4.95%	22.38%	4.98%	22.34%	196000
IE Size					441000					441000
Large IE	92.12%	10.14%	92.06%	10.26%	126000	4.98%	18.96%	5.09%	19.14%	126000
Medium IE	72.61%	15.61%	73.02%	15.96%	126000	5.06%	19.01%	4.92%	18.89%	126000
Small IE	42.70%	14.58%	42.79%	14.42%	126000	4.92%	19.13%	4.97%	19.01%	126000
IE Valence										
Positive IE	69.10%	0.17%	69.37%	0.16%	189000	4.93%	18.99%	4.97%	18.86%	189000
Negative IE	69.19%	26.72%	69.22%	26.93%	189000	5.04%	19.07%	5.01%	19.16%	189000
No IE	5.38%	4.26%	4.80%	4.13%	63000	5.22%	18.92%	4.92%	18.96%	63000
MEs Present					441000					441000
None	60.11%	-	60.05%	-	63000	4.97%	-	4.99%	-	63000
One	60.00%	12.17%	60.09%	12.16%	252000	5.05%	19.17%	4.97%	19.01%	252000
Both	-	12.09%	-	12.25%	126000	-	18.86%	-	19.00%	126000
ME Size					441000					441000
Large ME	59.99%	1.10%	60.08%	1.15%	189000	5.02%	0.00%	4.98%	0.00%	189000
Small ME	60.00%	23.16%	60.11%	23.25%	189000	5.07%	38.03%	4.97%	38.01%	189000

Table C3 Type-I and Type-II error for MEs of x_1 and x_2, estimated using step-wise analysis across all 441,000 datasets. Datasets were initially analyzed with the long model to determine whether the IE was significant. Following the step-wise approach, we reanalyzed datasets with significant IEs using the long model, and those with non-significant IEs using the short model. This process covers all 441,000 datasets.

Data Group	Short Model (IE Non-Sig.): Without Interaction					Long Model (IE Sig.): With Interaction					Total Datasets (Short + Long)
	x_1 Error Type		x_2 Error Type		No. Datasets	x_1 Error Type		x_2 Error Type		No. Datasets	
	I	II	I	II		I	II	I	II		
All	32.44%	10.95%	32.55%	10.86%	168441	5.18%	18.61%	5.19%	18.67%	272559	441000
Homogeneity											
Heterogeneous	30.01%	7.53%	29.78%	7.45%	13947	5.09%	10.26%	5.04%	10.20%	35053	49000
Low	31.40%	10.41%	31.59%	10.47%	69310	5.05%	17.72%	4.93%	17.83%	126690	196000
High	33.69%	11.95%	33.77%	11.74%	85184	5.35%	22.28%	5.54%	22.30%	110816	196000
IE Size											441000
Large IE	83.33%	19.23%	92.00%	10.53%	44	4.96%	18.96%	5.06%	19.13%	125956	126000
Medium IE	66.89%	14.10%	67.36%	14.65%	22808	3.88%	18.79%	3.77%	18.65%	103192	126000
Small IE	42.26%	14.74%	42.63%	14.54%	85722	6.75%	17.03%	6.92%	17.13%	40278	126000
IE Valence											
Positive IE	47.45%	0.43%	47.89%	0.38%	54386	4.79%	21.74%	4.82%	21.65%	134614	189000
Negative IE	47.41%	28.81%	47.82%	28.80%	54188	4.84%	15.47%	4.87%	15.65%	134812	189000
No IE	5.33%	4.30%	4.82%	4.13%	59867	35.94%	18.99%	34.76%	20.06%	3133	63000
MEs Present											441000
None	32.64%	-	32.88%	-	24088	5.16%	-	5.13%	-	38912	63000
One	32.34%	10.99%	32.38%	10.82%	95888	5.19%	18.80%	5.22%	18.78%	156112	252000
Both	-	10.91%	-	10.90%	48465	-	18.42%	-	18.55%	77535	126000
ME Size											441000
Large ME	31.99%	0.00%	32.71%	0.00%	72205	5.24%	0.00%	5.24%	0.00%	116795	189000
Small ME	32.69%	21.89%	32.05%	21.71%	72148	5.13%	37.22%	5.20%	37.34%	116852	189000

Table C4 Type-I and Type-II error for MEs of x_1 and x_2, estimated using long model with coefficients β_1 and β_2, for the 168,441 datasets where the IE was estimated as non-significant, i.e. the counterfactual full model analysis of the step-wise short model.

Datasets: IE Non-Sig.	**Long Model: With Interaction [Full Model]**				
Data Group	**x_1 Error Type**		**x_2 Error Type**		**No. Datasets**
	I	**II**	**I**	**II**	
All	4.76%	19.67%	4.64%	19.55%	168441
Homogeneity					
Heterogeneous	5.67%	10.84%	5.21%	10.76%	13947
Low	4.98%	17.93%	4.99%	17.82%	69310
High	4.44%	22.52%	4.26%	22.40%	85184
IE Size					
Large IE	72.22%	23.08%	64.00%	42.11%	44
Medium IE	10.44%	19.98%	10.13%	19.95%	22808
Small IE	4.05%	20.11%	4.04%	19.88%	85722
IE Valence					
Positive IE	5.29%	12.20%	5.36%	11.97%	54386
Negative IE	5.55%	27.98%	5.35%	27.86%	54188
No IE	3.58%	18.92%	3.34%	18.90%	59867
MEs Present					
None	4.65%	-	4.76%	-	24088
One	4.82%	19.77%	4.58%	19.37%	95888
Both	-	19.57%	-	19.72%	48465
ME Size					
Large ME	4.67%	0.00%	4.55%	0.00%	72205
Small ME	4.98%	39.33%	4.60%	39.09%	72148

IE: Small effect size & high homogeneity reduce detection, independent of method & MEs

Table C5 Association of all combinations of the MEs Present and size, with the IE's Type-II Error, from both marginal means (DiD) and regression coefficients approach, subset by the three IE sizes constructed

IE Size	**ME**			**Type-II Error of IE**	
	Present	**Size**			
		ME of x_1	**ME of x_2**	**DiD**	$\widehat{\boldsymbol{\beta}}_3$
Small IE	Both	Large ME	Large ME	68.18%	68.32%
Small IE	Both	Small ME	Small ME	68.40%	68.50%
Small IE	One	No ME	Large ME	67.79%	67.87%
Small IE	One	No ME	Small ME	67.91%	67.94%
Small IE	One	Large ME	No ME	68.02%	68.11%
Small IE	One	Small ME	No ME	67.74%	67.81%
Small IE	None	No ME	No ME	67.58%	67.69%
Medium IE	Both	Large ME	Large ME	18.29%	18.34%
Medium IE	Both	Small ME	Small ME	18.66%	18.72%

IE Size	ME Present	ME Size: ME of x_1	ME Size: ME of x_2	Type-II Error of IE: DiD	Type-II Error of IE: $\hat{\beta}_3$
Medium IE	One	No ME	Large ME	17.78%	17.86%
Medium IE	One	No ME	Small ME	17.45%	17.53%
Medium IE	One	Large ME	No ME	17.94%	17.99%
Medium IE	One	Small ME	No ME	17.72%	17.76%
Medium IE	None	No ME	No ME	18.44%	18.51%
Large IE	Both	Large ME	Large ME	0.03%	0.03%
Large IE	Both	Small ME	Small ME	0.03%	0.03%
Large IE	One	No ME	Large ME	0.01%	0.01%
Large IE	One	No ME	Small ME	0.04%	0.04%
Large IE	One	Large ME	No ME	0.05%	0.05%
Large IE	One	Small ME	No ME	0.04%	0.04%
Large IE	None	No ME	No ME	0.05%	0.05%

ME: Small effect size & high homogeneity reduce detection, independent of method & IE

Table C6 Difference in estimating Type-I and II errors for the ME of x_1 and x_2, using the marginal means pairwise comparison (PC) and the regression coefficients ($\hat{\beta}$) approach. Differences are estimates separately using the unadjusted and adjusted* PC. Results are presented for the entire data, and separately for subsets varying according to: homogeneity, IE's size & valence, and ME's presence & size.

Error Difference Data Group	PC - $\hat{\beta}$: x_1 Type-I	PC - $\hat{\beta}$: x_1 Type-II	PC - $\hat{\beta}$: x_2 Type-I	PC - $\hat{\beta}$: x_2 Type-II	Adjusted PC- $\hat{\beta}$: x_1 Type-I	Adjusted PC- $\hat{\beta}$: x_1 Type-II	Adjusted PC- $\hat{\beta}$: x_2 Type-I	Adjusted PC- $\hat{\beta}$: x_2 Type-II
All	0.03%	-0.04%	0.03%	-0.04%	-1.31%	3.39%	-1.29%	3.44%
Homogeneity								
Heterogeneity	0.04%	-0.05%	0.04%	-0.03%	-1.18%	2.66%	-1.13%	2.89%
Low	0.04%	-0.03%	0.03%	-0.04%	-1.25%	3.28%	-1.27%	3.33%
High	0.03%	-0.03%	0.02%	-0.03%	-1.39%	3.68%	-1.35%	3.69%
IE Size								
Large IE	0.03%	-0.04%	0.01%	-0.03%	-0.95%	1.76%	-0.94%	1.83%
Medium IE	0.05%	-0.03%	0.02%	-0.04%	-0.99%	3.18%	-0.87%	3.20%
Small IE	0.02%	-0.04%	0.04%	-0.03%	-1.49%	4.60%	-1.59%	4.70%
IE Valence								
Positive IE	0.03%	-0.03%	0.03%	-0.03%	-0.87%	1.93%	-0.90%	1.95%
Negative IE	0.04%	-0.04%	0.02%	-0.04%	-1.42%	4.43%	-1.36%	4.54%
No IE	0.03%	-0.04%	0.04%	-0.05%	-2.30%	4.66%	-2.23%	4.63%
MEs Present								
None	0.03%	0.00%	0.02%	0.00%	-2.09%	0.00%	-2.06%	0.00%
One	0.04%	-0.03%	0.03%	-0.04%	-0.91%	3.50%	-0.90%	3.62%
Both	0.00%	-0.04%	0.00%	-0.04%	0.00%	3.29%	0.00%	3.26%
ME Size								
Large ME	0.04%	0.00%	0.04%	0.00%	-0.42%	0.00%	-0.45%	0.00%

Small ME	0.03%	-0.07%	0.03%	-0.07%	-1.41%	6.78%	-1.35%	6.88%

*The adjusted PC adjusts the p-value for multiple comparison, as against the (unadjusted) PC where each PC tests a separate hypothesis and does not require any adjustment. We adjusted PC using the methods of Benjamini, Hochberg, and Yekutieli which is found to be more powerful and less conservative than other methods like the Bonferroni correction (R Core Team 2024).

Note[1]: While differences with the (unadjusted) PC were negligible, the PC approach consistently seemed to have higher Type-I and lower Type-II errors than the coefficient approach. This is seen by the positive and negative values in the Type-I and Type-II columns, across data groups, for both x_1 & x_2.

This pattern switched when using adjusted PC, where the differences between the two approaches were also seen to be more substantial. The Type-I error difference across data groups, for both x_1 & x_2, was consistently negative, implying that adjusted PC is more conservative or has lower Type-I error than the coefficient approach. This is understandable given that p-value adjustment methods are known to result in conservative estimates.

Note[2]: The Type-I and II error pattern seen with the adjusted PC suggests that in the absence of specific hypotheses, where p-value adjustment is necessary, the approach choice for MEs should be based on the research interest. If the focus is on having a low Type-I error, e.g. intervention effectiveness research, marginal means is more suitable. Whereas if it is more important to detect an effect, e.g. discrimination research, the regression coefficient method is more suitable.

Table C7 Type-II Error for MEs estimated using both marginal means (PC) and the regression coefficients approach ($\hat{\beta}$) for the data subsets varied as per the three homogeneity levels of the CAD interaction group (x_1x_2), and the MEs' presence.

Homogeneity	MEs Present	Type-II Error for ME of x_1		x_2	
		PC:00-10	$\hat{\beta}_1$	PC:00-01	$\hat{\beta}_2$
Heterogeneous	One	10.41%	10.48%	10.29%	10.31%
Heterogeneous	Both	10.34%	10.37%	10.37%	10.41%
Low	One	17.86%	17.88%	17.88%	17.92%
Low	Both	17.66%	17.71%	17.68%	17.73%
High	One	22.60%	22.63%	22.22%	22.26%
High	Both	22.10%	22.14%	22.40%	22.42%

Table C8 Type-II Error for MEs estimated using both marginal means (PC) and the regression coefficients approach ($\hat{\beta}$) for the data subsets varied as per MEs' size and presence

ME Variation		Type-II Error for ME of x_1		x_2	
Size	MEs Present	PC:00-10	$\hat{\beta}_1$	PC:00-01	$\hat{\beta}_2$
Large ME	One	0.00%	0.00%	0.00%	0.00%
Large ME	Both	0.00%	0.00%	0.00%	0.00%
Small ME	One	38.27%	38.33%	37.93%	38.01%
Small ME	Both	37.64%	37.72%	37.93%	38.00%

Table C9 Association of all combinations of the estimated IE size and valence, with the Type-II Error for MEs, from both marginal means (PC) and the regression coefficients ($\hat{\beta}$) approach, subset by the two ME sizes

ME Size	IE's		Type-II Error for ME of			
	Valence	Size	x_1		x_2	
			PC:00-10	$\hat{\beta}_1$	PC:00-01	$\hat{\beta}_2$
Small ME	No IE	No IE	37.75%	37.84%	37.82%	37.92%
Small ME	Negative IE	Large IE	37.77%	37.87%	38.50%	38.55%
Small ME	Negative IE	Medium IE	38.24%	38.30%	38.13%	38.23%
Small ME	Negative IE	Small IE	38.15%	38.24%	38.12%	38.18%
Small ME	Positive IE	Large IE	37.89%	37.97%	37.94%	38.01%
Small ME	Positive IE	Medium IE	37.67%	37.72%	37.26%	37.32%
Small ME	Positive IE	Small IE	38.22%	38.27%	37.77%	37.84%
Large ME	No IE	No IE	0.00%	0.00%	0.00%	0.00%
Large ME	Negative IE	Large IE	0.01%	0.01%	0.00%	0.00%
Large ME	Negative IE	Medium IE	0.00%	0.00%	0.00%	0.00%
Large ME	Negative IE	Small IE	0.01%	0.01%	0.00%	0.00%
Large ME	Positive IE	Large IE	0.00%	0.00%	0.00%	0.00%
Large ME	Positive IE	Medium IE	0.00%	0.00%	0.00%	0.00%
Large ME	Positive IE	Small IE	0.00%	0.00%	0.01%	0.01%

Appendix C2. Shorthand and Acronyms

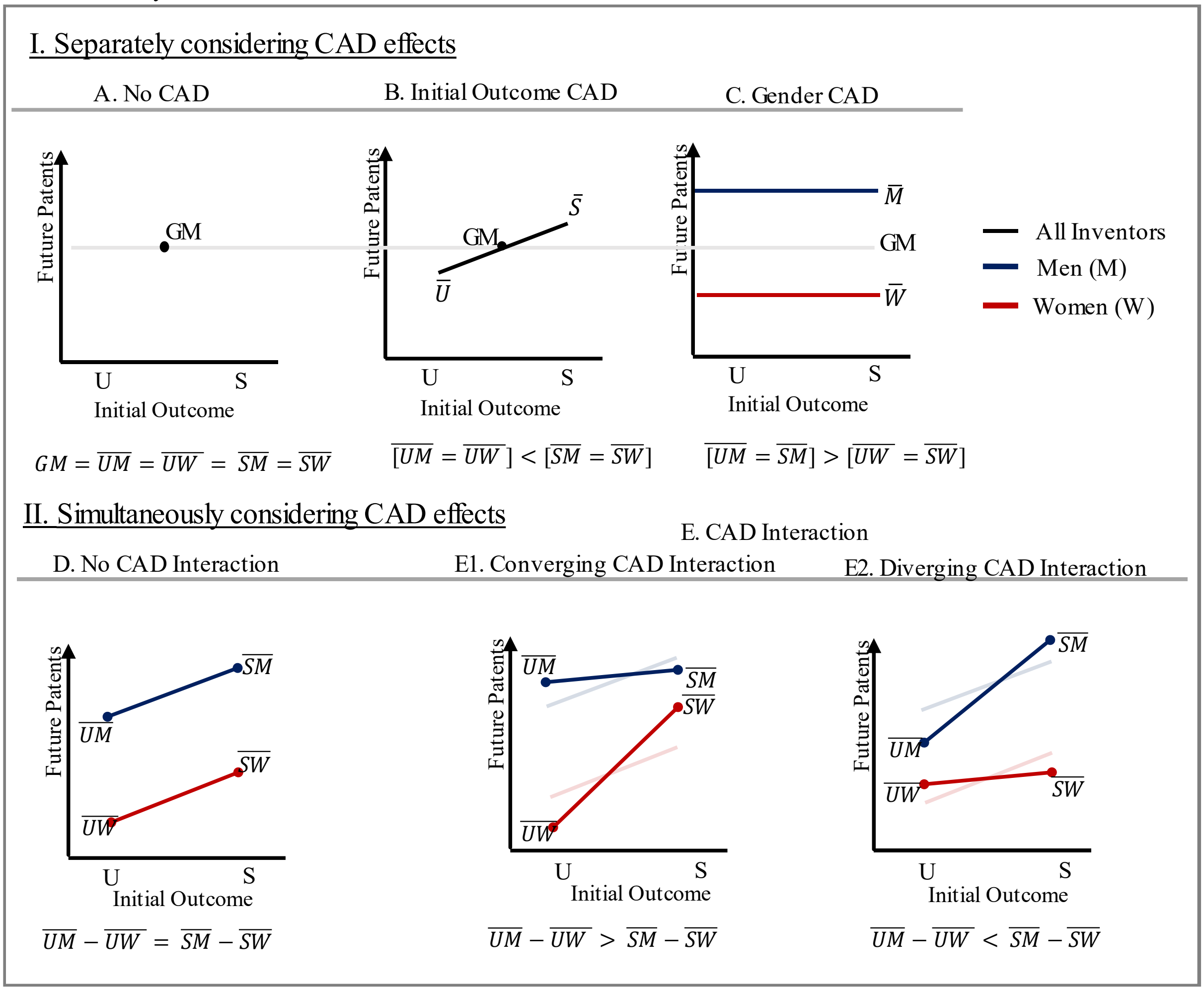
Shorthand	Explanation
CAD	cumulative advantage/disadvantage
CA	cumulative advantage
CD	cumulative disadvantage
L	Low status CAD group: Actors from poor Low-status households
H	High status CAD group: Actors from rich High-status households
S	Successful path CAD group: Actors who were initially Successful at work
U	Unsuccessful path CAD group: Actors who were initially Unsuccessful at work
SL	Successful Path and Low Status CAD interaction group: Actors from High-status households who were initially Unsuccessful at work
UH	Unsuccessful Path and High Status CAD interaction group: Actors from High-status households who were initially unsuccessful at work
UL	Unuccessful Path and Low Status CAD interaction group: Actors from Lw-status households who were initially Unsuccessful at work
SH	Successful path and High status CAD interaction group: Actors from High-status households who were initially Successful at work
$\widehat{\beta_\iota}$	Estimated Coefficient
$\beta_1 = \overline{SL} - \overline{UL}$	Path CAD's ME (for Success)
$\beta_2 = \overline{UH} - \overline{UL}$	Status CAD's ME (for High status)
$\beta_3 = [\overline{SH} - \overline{UH}] - [\overline{SL} - \overline{UL}] = [\overline{SH} - \overline{SL}] - [\overline{UH} - \overline{UL}]$	IE of Path and Status CAD or DiD comparison of the four groups
PC	Pairwise Comparison
DiD	Difference-in-Difference comparison
ME	Main effect
IE	Interaction effect
CI	Confidence interval
NHHI	Normalized Herfindahl–Hirschman index
HHI	Herfindahl–Hirschman index
pp	Percentage points

Equation(s)	
(i)	Regression equation for the step-wise approach's 'long' or 'full model', including the IE
(ii)	Regression equation for the step-wise approach's 'short model', excluding the IE
(iii)	Seven possible CAD effect estimates or possible group comparisons for a binary CAD interaction
(iv)	Three equations representing the group comparisons estimable by both coefficients and marginal means approaches
(v)	Standardized form of regression equation (i) used in simulation analysis

Appendix D

Supplementary material for Chapter 5: Cumulative Advantage as a Mechanism for Career (In)Equality: Can Some Women Benefit?

Figure D1 Identical inventors average future outcomes when CAD effects are considered (I) Separately, and (II) Simultaneously



All graphs in Figure D1 show the same relationship: future outcome or patents accumulated in seven years on the y-axis, initial success or path CAD on the x-axis and gender CAD as a grouping variable. The lines thus depict the CAD effect through the relation between inventors' CAD variables and their future outcomes. The equal world scenario, when there are no group differences is illustrated in Panel-A where all four groups have the same future patent average which is equivalent to the grand mean (GM), i.e. all seven group comparisons are 0.

A positive slope indicates path CAD, where unsuccessful inventors accumulate lesser patents than those who were successful, i.e. $\bar{U} < \bar{S}$ as in Panel-B . Non-overlapping lines indicate gender CAD, where women accumulate lesser patents than men, i.e. $\bar{W} < \bar{M}$ as in panels C, D and E. When two CAD effects are simultaneously present, the CAD effects can be visualized through positively-sloped non-overlapping lines as in the bottom panel, taking one of three forms: (1) parallel lines as in panel D, or non-parallel lines that are either (2) converging or (3) diverging as in panels E1 and E2, respectively.

Variable Definitions and Sources

Table D1: Variable description and source

Variable name	Description & Construction	Source
	Inventor Variables	
Initial outcome	Unsuccessful = 0 | Successful = 1 An indicator variable equal to one if the inventor's first application was successful at getting a patent grant and zero otherwise.	PatEx & PatentsView
Gender	Women = 0 | Men = 1 An indicator variable equal to one if the inventor is classified by the matching algorithm as a man and zero otherwise.	PatentsView
Future outcome		
No. of applications	Number of patent applications filed by inventor i between year of first-action decision t and year t+k, where k = 2 : 20 years.	PatEx & PatentsView
No. of patens	Number of successful patents granted to inventor i between year of first action t and year t+k, where k = 2 : 20 years.	PatentsView
	Examiner Variables	
Prior applications reviewed	Number of patent applications reviewed by the examiner, prior to the focal application	PatEx
Leniency	Examiner approval rate (%) Number of patent applications rejected by the examiner divided by the number of applications reviewed, prior to the focal application	PatEx
Art Unit		PatEx

Data Construction

Our data is primarily compiled using the June 2023 PatentsView release and the June 2022 PatEx dataset, the latest release at the time.

The PatEx is based on the data from the Patent Examination Data System (PEDS), downloaded by the Office of the Chief Economist at the USPTO. Focusing on applications' examination, the data provides information such as uniquely identified examiner names, their communication with the applicants, their decisions at every stage, the names of applicant inventors, etc. During robustness checks we found some data to be lost in the newer PatEx data versions. Hence combined Patex's application data releases from 2019, 2020, 2021 and 2022, such that the latest available information was retained.

The PatentsView is an award-winning visualization, data dissemination, and analysis platform hosted again by the Office of the Chief Economist at the USPTO to disseminate raw data on patent applications and grants. However, the platform benefits from multiple institutional and individual collaborators providing free data dissemination and value-added analyses.

Inventor disambiguation: Employing a series of algorithms and post-processing techniques, PatentsView provides unique identifiers for all patent inventors allowing us to track individual inventors through their careers, regardless of where or whom they work for.

Inventor's gender attribution: PatentsView also predicts the most likely gender of inventors based on the name provided in the application form. We conducted robustness checks on the gender attribution using other sources such as the World Gender Name Dictionary (WGND 2.0), built specifically by the UN's World Intellectual Property Organization (WIPO) for intellectual property data (Lax Martínez et al. 2016), and genderize.io, a gender prediction API. However found PatentsView's gender attribution algorithm to have the highest gender identification rate, with only about 10% of inventor names being unassigned a gender. Thus, we build on the PatentsView's gender attribution to identify and track those inventors for whom a clear gender assignment was possible.

Inventor's first application: To study the impact of inventors' initial outcome on their future outcome, we limit our sample to those inventors who filed a single first patent application after the American Inventors Protection Act was enforced in November 2000. Multiple first applications could have mixed outcomes which can contaminate our CAD estimates. To ensure we capture the start of all inventor careers, we exclude inventors with any applications filed or patents granted prior to the Act. Additionally, to ensure that these inventors are pursuing a career in innovation, we only include those inventors who submit at least two or more applications, filed at least a year apart.

Application Filing Dates

Start date: While PatentsView provides data from 1976, we only include applications filed after the AIPA on November 30, 2000. This ensures full inventor population coverage and avoids selection bias. However we still use pre-AIPA data to construct our instrument and identifying the start of inventor careers.

Note that our timing consideration is based on applications' first-action date and not its filing date. This means that the initial outcome considered does not technically follow from the first application filed, but the first application the inventor received an action letter for from the USPTO. However, this did not occur in our data subset.

End date: The latest PatEx release provides outcome data through June 1, 2023 under the "Application status description." Given that it takes an average of three years to receive a final decision (Farre-Mensa et al. 2020), we only include applications filed before June 1, 2020. this cutoff ensures that applications in our sample have had sufficient time to reach an outcome.

Application Type: Patent applications can be categorized under non-provisional, provisional, re-examination, re-issue, and PCT applications. Only non-provisional applications specify the invention's subject matter:

- <u>Utility patents:</u> New and useful process, machine, manufacture, or composition of matter.

- <u>Design patents:</u> New, original, and ornamental designs, embodied in or applied to an article of manufacture.
- <u>Plant patents</u>: New and distinct, invented or discovered asexually reproduced plants including cultivated sports, mutants, hybrids, and newly found seedlings.

Following the patenting literature, we limit our sample to non-provisional utility applications filed after enactment of the AIPA in November 2000 to minimize missing. Additionally we use information from PatEx to weed out inventors who made other types of applications.

To ensure that we only include applications with examiner decisions, we exclude provisional, re-examination, re-issue, and PCT applications. This avoids double-counting and incomplete data. Provisional applications, for instance, expire twelve months after filing and must be filed as a corresponding regular non-provisional application during the 12-month pendency period. The initial outcome of these applications would follow only after they transform into a non-provisional application, which is already captured in our data.

Applications Filed Under the Patent Cooperation Treaty (PCT) allows an applicant to file one "international application" and to have that application acknowledged as a basis for filing regular nonprovisional national phase or regional filings in as many PCT member states as the applicant designates. As our study focuses on US inventors and already includes all non-provisional applications filed, it would be redundant to also consider PCT. Moreover, the PCT data is found to have curious holes with about 18% of the PCT filings from 2000 and 2001 being missing (Graham, Marco, and Miller 2015).

Design patents are excluded as they are not covered by the AIPA, and plant patents are omitted because they may be granted upon discovery rather than innovation, and often lack pre-grant publication.

Application Outcome: Technically, the USPTO does not reject patent applications; instead, applications are abandoned by applicants after receiving rejections that could be appealed (Lemley and Sampat 2008). Thus we construct patent applications to have a binary outcome, where they are either successful or unsuccessful.

> A successful outcome is relatively easy to recognize as it is transformed into granted patent(s), which are well documented and publicized. We use information on patents granted from both PatentsView and PatEx to ensure comprehensive data coverage
>
> An unsuccessful outcome is however more difficult to establish. While the absence of a successful patent grant implies that the application was unsuccessful, the application could also have a different status, e.g. it could be still under processing or have missing information. To minimize this risk of misidentifying applications as unsuccessful we took two key steps. Firstly, we only include applications submitted before June 2019, despite our data having patent grant information up till June 2023, from PatentsView. Secondly we extracted and analyzed the content of the first-office action letter, communicated by the USPTO to the inventor, describing their decision on the patentability of the application. This was primarily done using the 2022 PatEx Transaction History dataset, which tracked information on all pre-examination and examination events. We started by filtering this data to only retain office action events for all applications submitted by our inventors, where the USPTO officially

communicates the applications' status to the applicant. Next we use the date on which the event occurred along with the USPTO's event codes to categorize the applications based on their latest available status: abandoned, appealed, rejected, patented, and other. 99.3% of applications that were not successfully patented were abandoned, which, as described earlier, occurs when an application is unsuccessful as the USPTO does not typically reject an application. Finally, to ensure maximum coverage, when outcome information or review date was missing per the latest PatEx release, we dove back into the earlier data releases, i.e. 2021, 2020 and 2019, to improve identification of the outcome variable.

Inventor Outcomes: To study how inventors' initial outcomes affect their future outcomes, we formulate the above described application outcomes as below:

Initial outcomes: Inventors' first ever application outcome, i.e. the binary successful or unsuccessful application outcome described above.

Future outcome: We quantify two types of future outcomes. The first refers to the inventor's the total number of patent applications submitted after receiving their initial outcome. The main study uses this number of subsequent applications made by an inventor as the future outcome measure.

As a robustness check, we also study a second measure of future outcomes which refers to the total number of successful patents granted to an inventor after receiving their initial outcome.

While CAD processes can drive both estimates, the first measure more accurately quantities the future innovation an inventor engages in, regardless of the outcome. Thus, throughout our analysis we use the application number based estimate to study the effects of CAD. In the next Appendix section, we use the success based estimate as a robustness check, and found our results to remain largely unchanged.

As we expect the effect of CAD to vary over time, the main study uses quantile regression to look at how inventors' subsequent future outcomes accumulate after 3, 5, 7, 9 and 15 years. We present only the select years to highlight the trend with brevity. Quantile regressions were estimated for all twenty years in our data and is available upon request.

Examiner Experience: The examiner experience is repeatedly estimated for each examiner when reviewing an application, based on the number of previous applications they reviewed. This means that this variable has a different value for each application, varying with each application and examiner. As inexperience can hinder examiner decision-making, we exclude first-time inventors whose application is assigned to an examiner with fewer than 10 prior reviews. This exclusion also allows for more robust examiner leniency estimation, our IV, which is central to our identification strategy.

Additional Summary Statistics & Results

Table D2 Share of Inventors with Successful and Unsuccessful First-Time Applications by Gender

	Women	**Men**
No. Inventors	75594	410550
No. Applications	68576	328803
No. Successful Applications	51951	259671
No. Unsuccessful Applications	16625	69132
% initially Successful	75.757%	78.975%
% initially Unsuccessful	24.243%	21.025%

CAD Effects After Seven Years

Table D3 Seven comparison and the main regression coefficient estimates from equation 1 using four variations of the dependent variable: 1. the number of subsequent applications submitted by an inventor 7 years after receiving their first-action decision (Column-A), 2. number of subsequent patents granted to an inventor 7 years after receiving their first-action decision (Column-B), 3. Natural log of Column-A, and 4. Natural log of Column-B. Panel A presents the seven comparison estimates based on the corresponding regression coefficients from panel B which presents the main second stage results of equation 1. All variables are defined in Table D1. Standard errors reported in parentheses are clustered at the art unit-year level.

		Future Outcome: Seven Years after First-Action Decision on Inventor's First Application			
	Hypothesis	**A. No. subsequent patent applications filed**	**B. No. subsequent patents granted**	**log(A)**	**log(B)**
Panel-A : Comparison Estimates Predicted					
1: Interaction Effect	$< 0\ (H_1)$	-4.486** (1.511)	-2.071** (0.684)	-0.379*** (0.083)	-0.212* (0.083)
2: SW-UW	$> 0\ (H_2)$	5.468** (1.815)	4.548*** (0.829)	0.538*** (0.100)	0.776*** (0.098)
3: UM-UW	$> 0\ (H_3)$	3.633** (1.150)	1.764*** (0.533)	0.331*** (0.063)	0.195** (0.067)
4: SM-UW	$> 0\ (H_4)$	4.616** (1.540)	4.242*** (0.721)	0.491*** (0.086)	0.759*** (0.086)
5: SM-UM	$\geq 0\ (H_5)$	0.982 (1.112)	2.478*** (0.556)	0.159* (0.065)	0.564*** (0.064)
6: SM-SW	$\geq 0\ (H_6)$	-0.852* (0.384)	-0.306 (0.162)	-0.047* (0.021)	-0.016 (0.017)
7: SW-UM	$= 0\ (H_7)$	1.835 (1.161)	2.784*** (0.564)	0.207** (0.066)	0.580*** (0.064)
Panel-B : Regression Output					
Initial outcome: Successful		5.468** (1.815)	4.548*** (0.829)	0.538*** (0.100)	0.776*** (0.098)
Gender: Man		3.633** (1.150)	1.764*** (0.533)	0.331*** (0.063)	0.195** (0.067)
Successful x Man		-4.486** (1.511)	-2.071** (0.684)	-0.379*** (0.083)	-0.212* (0.083)
Control: Examiner experience		Yes	Yes	Yes	Yes

Hypothesis	A. No. subsequent patent applications filed	B. No. subsequent patents granted	log(A)	log(B)
	Future Outcome: Seven Years after First-Action Decision on Inventor's First Application			
Fixed-Effects: Examiner art unit x application year	Yes	Yes	Yes	Yes
Observations	167,753	206,971	167,753	200,257
R2	0.074	0.137	0.086	0.159
Within R2	-0.003	0.006	-0.007	0.006
F-test (IV only)	5.911	18.664	29.772	54.643
Wald (IV only), p-value	0.002	0.000	0.000	0.000

Significance codes: $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***)

These results are largely similar to Table 5.4, except for H_6 and H_7. There was partial support for H_6, with non-significant difference in the patents subsequently granted to SM and SW. We also find evidence of the SM-UM difference being absent through considerable overlap in their CIs, see Figure D2. However, this differential result continues to align with our theory of SW overcoming their gender disadvantages by leveraging their initial success.

While H_7 was supported with evidence of there being no difference in the patent applications submitted by SM and UM, other outcome measures suggest the contrary. For instance, SW had almost 3 more patents on average, than UM. However, these differential result still aligns with our theory on men being able to leverage their higher status gender to reduce disadvantage from initial failure. This becomes more apparent when considering the SM-UM comparison trend of both their applications and patents together, as in Figure D2. UM's reduced disadvantage first reflects in their opportunities, i.e. they would first see an increase in application numbers and then the patents granted. This implies a lag in advantages accumulated, with the application measure increasing before the patents measure. Figure D2 illustrates precisely this, where UM's application outcome increases at a faster rate, with the blue line flattening more in the first row, as compared to their patent outcome in the second row.

Figure D2 CAD interaction trend: Inventors' predicted future outcomes after 3, 5, 7, and 9 years, based on the 2SLS IV regression from equation 1.

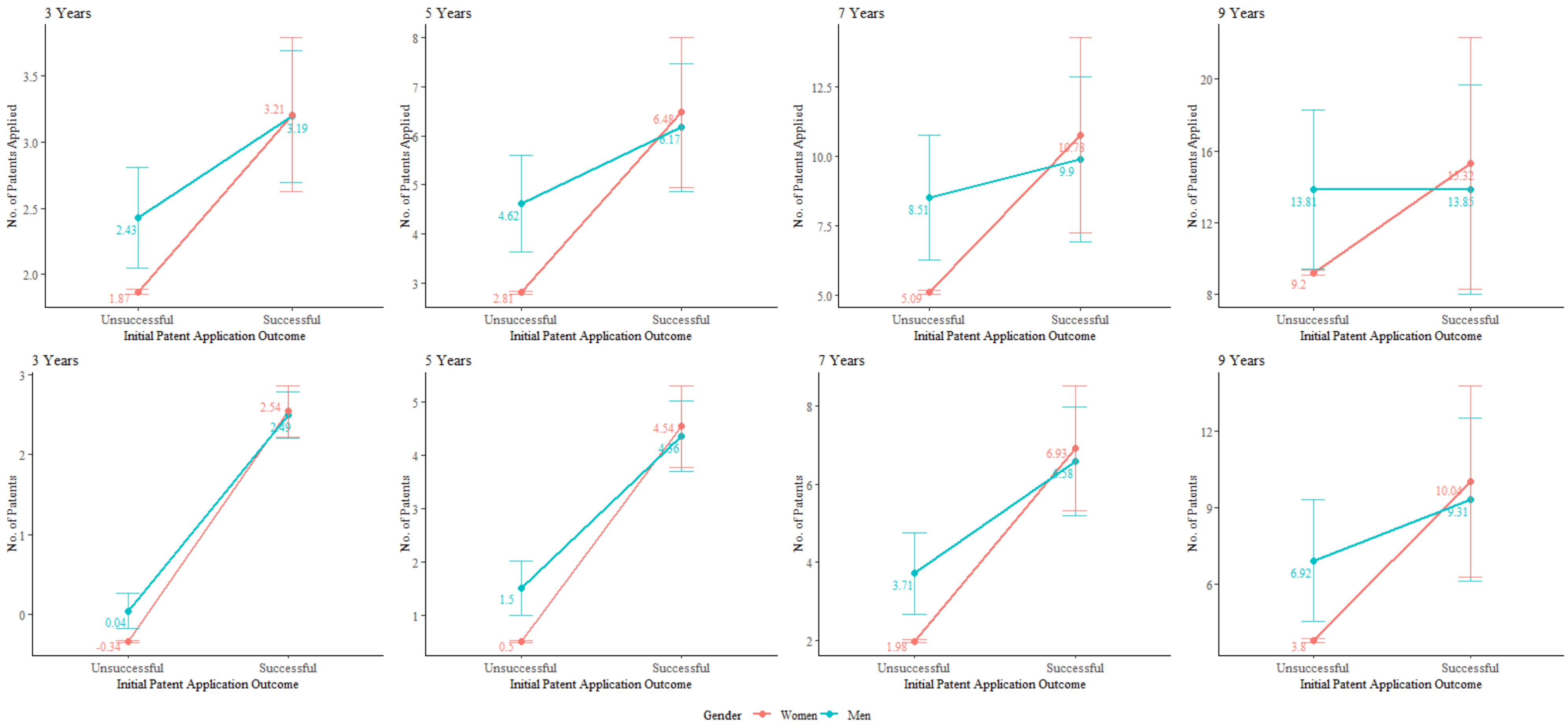


Figure D3 Seven possible comparisons of the four CAD interaction groups—successful men (SM), successful women (SW), unsuccessful men (UM), and unsuccessful women (UW)—quantifying the predicted difference in patents received seven years after the inventor's first application decision. The plot visualize these estimates also at the third, fifth, and ninth years.

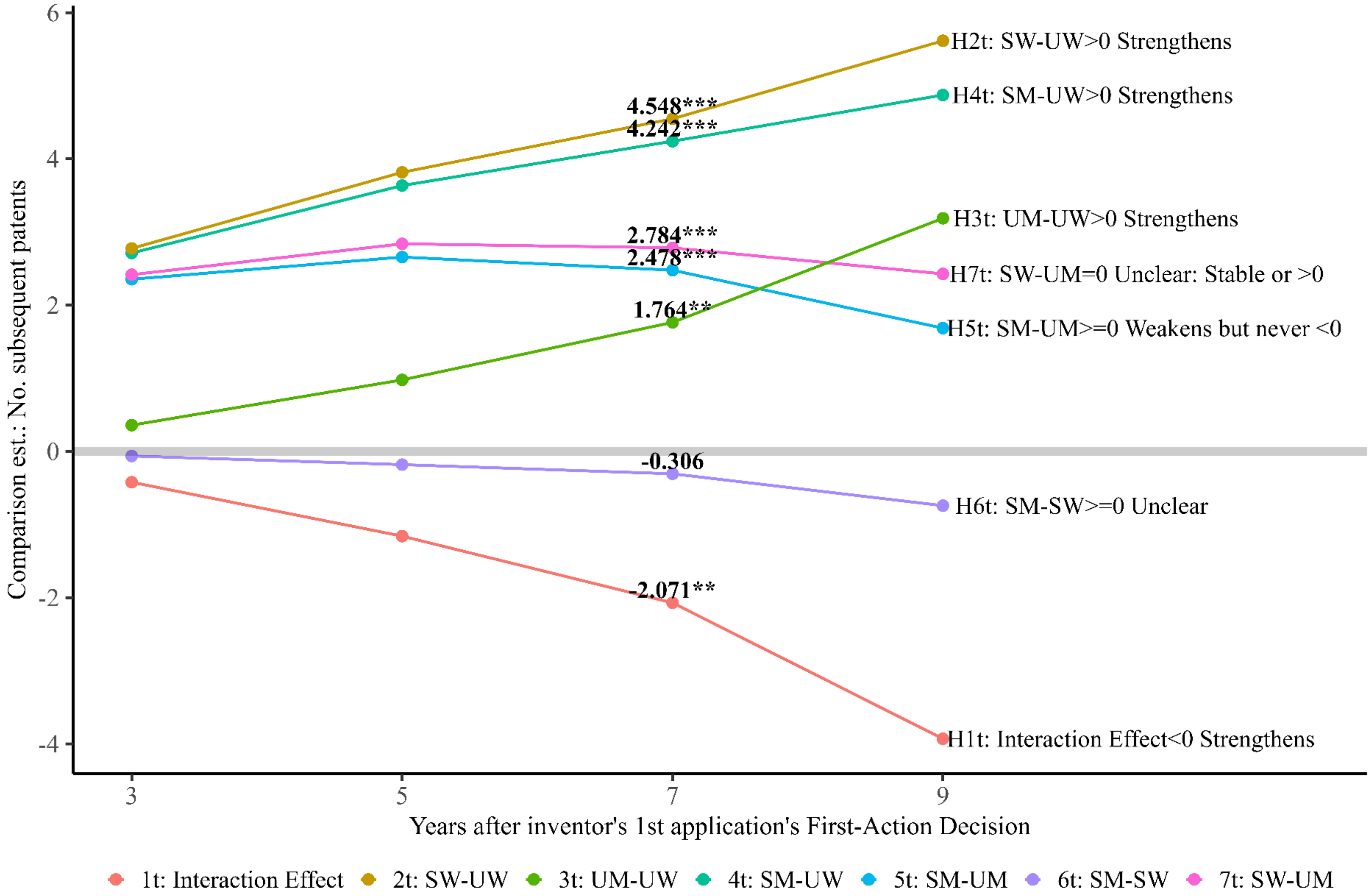


Table D4 Seven comparison and the main regression coefficient estimate trend based on the main second stage results of equation 1. The dependent variable is the number of subsequent patents granted to an inventor, k years after receiving their first-action, where k reflects the different columns of: 3, 5, 7 and 9 years. All variables are defined in Table D1. Standard errors reported in parentheses are clustered at the art unit-year level.

	Future Outcome: Number of subsequent patents granted after			
Estimate	**3 Years**	**5 Years**	**7 Years**	**9 Years**
Panel-A : Comparison Estimates Predicted				
1: Interaction Effect	-0.422** (0.139)	-1.159*** (0.332)	-2.071** (0.684)	-3.930* (1.565)
2: SW-UW	2.776*** (0.168)	3.816*** (0.399)	4.548*** (0.829)	5.615** (1.926)
3: UM-UW	0.359** (0.111)	0.978*** (0.260)	1.764*** (0.533)	3.188** (1.215)
4: SM-UW	2.713*** (0.149)	3.636*** (0.347)	4.242*** (0.721)	4.874** (1.650)
5: SM-UM	2.354*** (0.118)	2.658*** (0.263)	2.478*** (0.556)	1.686 (1.136)
6: SM-SW	-0.063* (0.030)	-0.181* (0.077)	-0.306 (0.162)	-0.742* (0.369)
7: SW-UM	2.417*** (0.118)	2.839*** (0.267)	2.784*** (0.564)	2.427* (1.200)
Panel-B : Regression Output				
Initial outcome: Successful	2.776*** (0.1683)	3.816*** (0.3986)	4.548*** (0.8286)	5.615** (1.926)
Gender: Man	0.3588** (0.1108)	0.9779*** (0.2604)	1.764*** (0.5331)	3.188** (1.215)
Successful x Man	-0.4218** (0.1387)	-1.159*** (0.3318)	-2.071** (0.6839)	-3.930* (1.565)
Control: Examiner experience	Yes	Yes	Yes	Yes
Fixed-Effects: Examiner art unit x application year	Yes	Yes	Yes	Yes
Observations	414,059	307,985	206,971	136,432
R2	0.11391	0.13342	0.13679	0.13921
Within R2	-0.01313	0.00408	0.00601	0.00301
F-test (IV only)	227.36	60.028	18.664	5.5957
Wald (IV only), p-value	9.4e-96	2.67e-29	1.97e-9	0.01136

Significance codes: $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***)

We found a similar pattern for all comparisons as with the future outcome based on application numbers in Figure 5.4 and Table 5.5. The only exception was SM-UM (H_{5t}), which weakened and remained positive as expected. This aligns with our earlier robustness observation of a lag between men's applications and success; illustrated by the differential evolution of men's slope for the two outcomes in Figure D2.

Table D5 Dropout rates between time intervals indicated in years following first-action decision for all groups per summary statistics, i.e. broken down by gender, initial outcome, and the combination of the two.

	Dropout rates between time intervals (years) following first-action decision					
Groups	**0-3**	**3-5**	**5-7**	**7-9**	**0-7**	**0-9**
Women	12.50%	41.27%	41.23%	36.45%	69.80%	80.81%
Men	11.13%	37.03%	36.92%	32.58%	64.70%	76.20%
Unsuccessful	10.70%	37.10%	38.07%	32.94%	65.21%	76.67%
Successful	11.51%	37.84%	37.40%	33.15%	65.57%	76.98%
Unsuccessful Women	12.13%	41.03%	41.28%	35.11%	69.58%	80.26%
Successful Women	12.62%	41.35%	41.22%	36.88%	69.87%	80.99%
Unsuccessful Men	10.38%	36.25%	37.43%	32.54%	64.25%	75.88%
Successful Men	11.32%	37.23%	36.79%	32.59%	64.81%	76.28%

Note: Dropout rates represent the percentage reduction in group size between two time points after inventors receive their first-action letter. They are calculated as: Dropout Rate (%) = 100 × (j − k) / j, where j and k refer to the number of inventors in the corresponding group at the start and end of each interval. This provides a pseudo estimate based on the submission of inventor applications, given the absence of comprehensive information on inventors' layoff or resignation.

Women show consistently higher dropout rates across all groups—not only compared to men overall but also within subgroups. Over a nine-year period, both UW and SW have dropout rates exceeding 80%, while UM and SM drop out at a lower rate of around 75%.

Supplementary Regression Output

Table D6 First stage of the 2SLS regression results presented in Table 5.5, measuring future outcome using number of subsequent applications. The use of interaction introduces three dependent variables in the first stage for every regression: inventors' initial outcome, their gender and the interaction of the two. While the use of gender as a dependent variable is not intuitive, it is necessary for estimating unbiased effects in the second stage e.g. as applied by Farre-Mensa et al., (2020). The results are presented in four different panels, with each panel corresponding to first stage results of the future outcome or dependent variable, i.e. for each column of k in Table 5.5. In other words, the dependent variables for each panel in the second stage refers to the number of subsequent applications submitted by an inventor, k years after receiving their first-action, where k = 3, 5, 7, and 9. All variables are defined in Table D1. Standard errors reported in parentheses are clustered at the art unit-year level.

	Dependent Variable		
Future Outcome: Number of subsequent applications	**Initial outcome: Successful**	**Gender: Man**	**Successful x Man**
Panel I: Future outcome after **3 Years**			
Examiner leniency	0.0032*** (0.0001)	-1.47e-17*** (7.2e-19)	-0.0005*** (7e-5)
Gender: Man	-0.0015 (0.0067)	1.000*** (4.84e-17)	0.5031*** (0.0056)
Examiner Leniency x Man	0.0001 (8.77e-5)	-3.16e-18*** (7.23e-19)	0.0039*** (7.13e-5)
Control: Examiner experience	1.93e-5*** (1.62e-6)	2.12e-21 (3.05e-21)	1.65e-5*** (1.41e-6)
Fixed-Effects: Examiner art unit x application year	Yes	Yes	Yes
Observations	430,999	430,999	430,999
R2	0.13224	1	0.44796
Within R2	0.01726	1	0.38049
F-test (IV only)	1,887.5	4.39e+34	103,314.9
Panel II: Future outcome after **5 Years**			
Examiner leniency	0.0029*** (0.0001)	2.53e-17*** (5.66e-19)	-0.0005*** (8.27e-5)
Gender: Man	-0.0100 (0.0089)	1.000*** (4.04e-17)	0.5127*** (0.0065)
Examiner Leniency x Man	0.0002* (0.0001)	-3.22e-17*** (6.66e-19)	0.0036*** (8.04e-5)
Control: Examiner experience	2.01e-5*** (1.89e-6)	-2.54e-21. (1.36e-21)	1.74e-5*** (1.66e-6)
Fixed-Effects: Examiner art unit x application year	Yes	Yes	Yes
Observations	268,592	268,592	268,592
R2	0.13490	1	0.43592
Within R2	0.01463	1	0.36141
F-test (IV only)	1,066.3	4.47e+34	62,533.4
Panel III: Future outcome after **7 Years**			
Examiner leniency	0.0026*** (0.0002)	2.68e-17*** (2.49e-18)	-0.0006*** (0.0001)
Gender: Man	-0.0167 (0.0115)	1.000*** (1.63e-16)	0.5074*** (0.0082)

Examiner Leniency x Man	0.0003* (0.0001)	-4.3e-17*** (2.88e-18)	0.0036*** (9.9e-5)
Control: Examiner experience	2.31e-5*** (2.26e-6)	3.18e-21 (8.3e-21)	1.97e-5*** (1.99e-6)
Fixed-Effects: Examiner art unit x application year	Yes	Yes	Yes
Observations	167,753	167,753	167,753
R2	0.14466	1	0.43302
Within R2	0.01321	1	0.34962
F-test (IV only)	601.35	4.95e+33	36,500.5
Panel IV: Future outcome after **9 Years**			
Examiner leniency	0.0026*** (0.0002)	-3.13e-17*** (1.26e-18)	-0.0008*** (0.0001)
Gender: Man	-0.0220 (0.0151)	1.000*** (1.07e-16)	0.4800*** (0.0106)
Examiner Leniency x Man	0.0005* (0.0002)	2.61e-17*** (1.24e-18)	0.0039*** (0.0001)
Control: Examiner experience	2.69e-5*** (2.66e-6)	-2.12e-21 (1.87e-21)	2.33e-5*** (2.42e-6)
Fixed-Effects: Examiner art unit x application year	Yes	Yes	Yes
Observations	112,222	112,222	112,222
R2	0.15497	1	0.42982
Within R2	0.01338	1	0.34022
F-test (IV only)	391.63	2.56e+34	22,986.3

Table D7 Average group estimates based on the regression coefficient output presented in Table 5.5. The dependent variable was the number of subsequent applications submitted by an inventor k years after receiving their first-action, where k reflects the different columns of: 3, 5, 7 and 9 years.

	Future Outcome: Number of subsequent applications submitted after			
	3 Years	**5 Years**	**7 Years**	**9 Years**
Successful Men	3.194	6.173	9.904	13.845
Unsuccessful Women	1.866	2.809	5.094	9.202
Successful Women	3.208	6.477	10.777	15.319
Unsuccessful Men	2.428	4.620	8.507	13.815

DUTCH SUMMARY

Dutch Summary

In mijn proefschrift, Structural Gender Inequality at Work: How Integration and the Accumulation of Advantages Shape (In)equality, onderzoek ik hoe structurele mechanismen genderongelijkheid op de werkvloer vormgeven en in stand houden. Mijn centrale stelling is dat de kern van het probleem niet uitsluitend ligt in zichtbare ondervertegenwoordiging van vrouwen op de werkvloed, maar in subtiele, relationele processen die bepalen wie toegang heeft tot informatie, kansen en erkenning. Twee processen staan in deze dissertatie centraal: integratie, de mate waarin werknemers daadwerkelijk sociaal en professioneel zijn ingebed in hun teams en organisaties, en cumulative (dis)advantage (CAD), de dynamiek waarbij kleine verschillen aan het begin van een loopbaan zich in de tijd opstapelen tot aanzienlijke ongelijkheden. Deze combinatie van een relationeel en temporeel perspectief stelt mij in staat om te analyseren hoe ongelijkheid ontstaat, voortduurt en, onder de juiste voorwaarden, ook kan afnemen.

Dit onderzoek is ingebed in het interdisciplinaire programma SCOOP (Sustainable Cooperation: Roadmaps to a Resilient Society). Binnen dat programma richt ik mij op duurzame samenwerking in organisaties en op de vraag onder welke voorwaarden samenwerking tussen mannen en vrouwen niet alleen mogelijk is, maar ook rechtvaardig en veerkrachtig wordt. Ik vertrek vanuit een observatie die mij al vroeg in mijn onderzoek raakte: we tellen wel het aantal vrouwen, maar we meten zelden de relaties die ze hebben op de werkvloer. Representatiecijfers zijn onmisbaar, maar zij verhullen vaak wat er werkelijk gebeurt in de sociale infrastructuur van alledag op het werk. Aanwezigheid is niet hetzelfde als participatie, en participatie is niet hetzelfde als integratie. Wie niet geïntegreerd is in de sociale en professionele netwerken waar informatie circuleert en kansen ontstaan, blijft, ondanks aanwezigheid, aan de randen van de organisatie staan.

In het theoretisch deel van mijn dissertatie beschrijf ik hoe organisaties gendered contexten zijn—plekken waarin normen, taal en routines historisch door mannen zijn gevormd, en waarin vrouwen vaak als 'buitenstaander' moeten navigeren. Dat is niet slechts een kwestie van cultuur; het is óók een netwerkrealiteit. Wie met wie werkt, met wie luncht, wie een mentor heeft of wie door een leidinggevende informeel wordt meegenomen naar een klantbespreking: het zijn precies die kleine, relationele keuzes die cumuleren tot grote verschillen in loopbanen. Tegelijkertijd beargumenteer ik dat ongelijkheid zelden plotseling ontstaat. Door CAD begrijp ik ongelijkheid als een proces dat zich in de tijd stapelt: een vroege tegenslag kan via beschadigd vertrouwen, verminderde reputatie, of minder toegang tot middelen en samenwerkingskansen jarenlang doorwerken. Vanuit deze dubbele lens toets ik in vier empirische hoofdstukken, die zich in schaal en tijdsbestek opbouwen, genderongelijkheid op de werkvloer: van individuele netwerken (hoofdstuk 2), via organisatiebeleid (hoofdstuk 3) en methodologische verfijning (hoofdstuk 4), naar langjarige loopbanen op grote schaal (hoofdstuk 5).

In het tweede hoofdstuk onderzoek ik hoe de gendercompositie van individuele netwerken samenhangt met carrière-uitkomsten voor betaald werkende mannen en vrouwen. Ik maak gebruik van de European Sustainable Workforce Survey (ESWS), een rijke dataset met gegevens van 4.345 werknemers, georganiseerd in honderden teams verspreid over negen

Europese landen en zes sectoren. Deze survey bevat niet alleen achtergrondkenmerken en uitkomsten (inkomen, tevredenheid, gepercipieerde loopbaanvooruitzichten), maar ook relationele informatie waarmee affectieve samenwerkingsnetwerken kunnen worden gereconstrueerd: wie werkt graag met wie samen, met wie socialiseert men buiten formele setting om. Die gegevens maken het mogelijk om integratie niet indirect via functie of hiërarchie te benaderen, maar rechtstreeks via het patroon van relaties waarin werknemers feitelijk zijn ingebed.

Om de tegenstrijdige bevindingen in de literatuur te adresseren—sommige studies vinden voordelen van netwerken van gelijken, andere van ongelijke netwerken—introduceer ik een typologie van drie netwerkconfiguraties: homophilous (overwegend contacten met hetzelfde gender), heterophilous (overwegend contacten met de andere gender) en genderdiverse netwerken (een relatief evenwicht). Vervolgens koppel ik die configuraties aan twee soorten uitkomsten: een objectieve maat (het inkomensniveau) en een subjectieve maat (het eigen oordeel over toekomstperspectief binnen de organisatie). Cruciaal is dat ik, naast klassieke predictoren, de team- en organisatieniveaus expliciet meeneem in de modellen, zodat rekening wordt gehouden met contextkenmerken als sector of teamgrootte.

Methodologisch combineer ik een multilevel model met Bayesiaanse informatieve hypothesetoetsing om tegelijk meerdere, theoretisch gefundeerde hypothesen tegen elkaar te wegen. Deze aanpak is belangrijk, omdat netwerkprocessen vaak subtiel zijn en klassieke Null Hypotheses Testing-framings (met afzonderlijke p-waarden per contrast) weinig robuust blijken bij overlappende hypothesen. De Bayesiaanse benadering dwingt mij om a priori plausibele relaties expliciet te maken en deze tegen elkaar af te wegen. Zo kan ik onderscheid maken tussen scenario's waarin genderdiversiteit per se samenhangt met betere uitkomsten, en scenario's waarin vooral heterofiele contacten voor vrouwen in mannelijke contexten renderen.

De resultaten laten een helder patroon zien. Over de hele linie hangen genderdiverse netwerken samen met betere uitkomsten, zowel subjectief (een positiever toekomstperspectief) als objectief (hoger inkomen). Dit geldt niet exclusief voor vrouwen: ook mannen lijken te profiteren van breed samengestelde netwerken, plausibel omdat diversiteit in contacten toegang geeft tot uiteenlopende informatiebronnen en kansen. Tegelijkertijd zie ik een belangrijke nuancering voor het inkomensdomein van vrouwen: in sterk mannelijke teams blijkt een gemengd netwerk (relatief veel mannelijke contacten) de hoogste inkomensopbrengst te hebben. Mijn interpretatie is dat vrouwen in zulke contexten via mannelijke contacten toegang krijgen tot schaarse en invloedrijke middelen—sponsorschap en vroege signalen van competentie—die in een hiërarchie gedomineerd door mannen cruciaal blijken. Zodra vrouwen echter niet langer een kleine minderheid vormen, verdwijnen deze verschillen: de opbrengsten worden dan vergelijkbaar.

Deze bevindingen helpen de schijnbare inconsistenties in de literatuur verklaren. Studies die gendergelijke netwerken als waardevoller beschouwden, en studies die juist genderdiverse contacten belangrijker vonden, lijken tegengesteld totdat we rekening houden met de statushiërarchie waarin die netwerken opereren. Genderdiverse netwerken zijn in principe het gunstigst, maar onder condities van extreme ondervertegenwoordiging blijft toegang tot machtige (veelal mannelijke) circuits voor vrouwen een noodzakelijke randvoorwaarde om niet structureel buiten de boot te vallen. Daarmee biedt dit hoofdstuk een relationele verklaring voor het fenomeen dat 'meedoen' in de praktijk vaak betekent: opgenomen worden in specifieke, historisch gegroeide netwerken.

Met deze micro-inzichten ga ik in het derde hoofdstuk naar het mesoniveau van Gender Diversity Management (GDM): beleid en praktijken waarmee organisaties gendergelijkheid proberen te bevorderen. De centrale vraag is eenvoudig maar urgent: doen deze maatregelen wat zij beloven? Veel gangbaar beleid wordt geëvalueerd op basis van representatie (meer vrouwen in teams of management), maar zegt weinig over de vraag of dat beleid ook relaties op de werkvloer verandert. Daarom richt ik mij in dit hoofdstuk op de verhouding tussen representatie en integratie: ik vergelijk uitkomsten gemeten met klassieke representatie-indicatoren met uitkomsten gemeten met een op netwerken gebaseerde integratiemaat.

Ik onderscheid drie categorieën GDM-maatregelen: interventies op individueel niveau (mentoring, netwerkprogramma's, leiderschapstrainingen), op leidinggevend niveau (biastrainingen, evaluatieprocedures die discriminatie moeten tegengaan) en op organisatorisch niveau (quota, verantwoording via rapportages, taskforces en formele verantwoordelijkheidstoedeling). Met multilevel modellen analyseer ik vervolgens in hoeverre deze strategieën de integratie van vrouwen in teams beïnvloeden, bovenop eventuele effecten op aantallen. Omdat integratie geen vanzelfsprekende uitkomst is van representatie, is de vraag of beleid relationele structuren daadwerkelijk openbreekt, en niet alleen 'de aantallen' bijstuurt.

De uitkomsten zijn ontnuchterend. Wanneer ik enkel naar representatie kijk, lijken vooral interventies op individueel niveau bescheiden positieve effecten te hebben: mentorprogramma's en gerichte netwerkactiviteiten vergroten de zichtbaarheid van vrouwen en zetten soms beweging in gang aan de 'toeleveringskant'. Maar zodra ik integratie meeneem, verdwijnt veel van die winst. In de meeste organisaties blijken GDM-maatregelen wel aantallen te beïnvloeden, maar nauwelijks de sociale architectuur waarbinnen die aantallen betekenis krijgen. Leidinggevenden volgen een training, er komt een taskforce of een quotum, maar wie daadwerkelijk met wie kennis deelt, wie werk met grote zichtbaarheid krijgt en wie als 'beloftevol' wordt gezien, verandert vaak niet. De conclusie is tweeledig: representatie is nodig, maar onvoldoende; en beleid dat niet expliciet op relaties stuurt, raakt de kern van structurele ongelijkheid niet.

Deze bevinding dwingt tot een verschuiving in evaluatiekaders. Ik pleit ervoor om integratie niet langer te behandelen als een impliciet neveneffect van representatie, maar als een expliciet beleidsdoel. Organisaties zouden routineus sociale netwerken (geanonimiseerd en met waarborgen) moeten monitoren—bijvoorbeeld via periodieke netwerkvragenlijsten of via geaggregeerde communicatiepatronen—en beleid moeten richten op relationele bruggen: gemengde projectteams met echte taakinterdependentie, formele sponsorschapstrajecten waarin senioren junioren introduceren in cruciale circuits, en beoordelingssystemen die samenwerking en kennisdeling expliciet belonen.

Het vierde hoofdstuk betreft een methodologisch onderzoek naar de onbetrouwbaarheid van stap-voor-stap regressies bij het detecteren en interpreteren van interactie-effecten in ongelijkheidsonderzoek. Wie ongelijkheid onderzoekt, stuit onvermijdelijk op interacties tussen verschillende factoren die de kans op succes vergroten of verkleinen (bijvoorbeeld tussen gender en netwerkpositie of tussen gender en vroege loopbaanuitkomsten). De bevindingen in de literatuur waren echter tegenstrijdig of niet robuust. Daarom heb ik een grootschalige simulatiestudie gedaan om te onderzoeken hoe verschillende analysemethoden presteren wanneer de 'ware' wereld bekend is, en juist dit is belangrijk in het genderonderzoek dat ik doe.

Ik simuleer 441 unieke scenario's van groepsverschillen en interacties, en trek voor elk scenario 1.000 steekproeven van 1.000 observaties—goed voor 441.000 schattingen. Vervolgens vergelijk ik twee analysetechnieken. De eerste is de wijdverbreide stepwise-aanpak, waarin men stelselmatig variabelen toevoegt en verwijdert op basis van significantie. De tweede is de volledig gespecificeerde modelbenadering, waarin hoofd- en interactietermen blijven staan en interpretatie niet afhangt van het al dan niet 'overleven' van afzonderlijke termen. De resultaten zijn helder: stepwise mist in circa 68% van de gevallen echte interacties (Type II-fouten) en rapporteert in circa 47% van de gevallen schijn-effecten op de hoofdeffecten (Type I-fouten) die verdwijnen zodra de interactie correct wordt gemodelleerd. Dit probleem wordt acuter naarmate groepen kleiner of homogener zijn—condities die in ongelijkheidsonderzoek (bijvoorbeeld bij vrouwen in STEM of minderheden in management) vaak juist voorkomen.

Deze resultaten hebben mij ertoe gebracht expliciete richtlijnen te formuleren voor ons onderzoek. Theoretiseer eerst het volledige 2×2- of k×m-interactieraamwerk—denk daarbij in termen van groepen en hun relatieve posities—en leg dan in één volledig model de hypothesen vast, zonder p-waarden te corrigeren voor alle onderlinge vergelijkingen als die vergelijkingen inhoudelijk deel uitmaken van één globaal toetsingskader. Visualiseer vervolgens de zeven relevante groepsvergelijkingen (de zes directe paren plus de difference-in-differences) in één figuur; zo wordt zichtbaar waar ongelijkheden werkelijk ontstaan. En wees terughoudend met het 'wegpoetsen' van interacties omwille van modelzuinigheid: in ongelijkheidsonderzoek is de interactie vaak het fenomeen, niet de ruis.

Met deze methodologische bagage keer ik in het vijfde hoofdstuk terug naar de vraag hoe CAD zich in echte loopbanen ontvouwt. Ik bestudeer hiervoor de innovatie-infrastructuur van de Verenigde Staten aan de hand van 2,8 miljoen octrooi aanvragen bij de United States Patent and Trademark Office (USPTO) tussen 2001 en 2020. Centraal staat de vraag hoe een vroege uitkomst—een eerste octrooi dat wordt goedgekeurd of afgewezen—doorwerkt in het vervolg van iemands loopbaan, en of die effecten verschillen naar gender. Om causaliteit te benaderen maak ik gebruik van een quasi-experimenteel design waarin de strengheid van de eerste octrooi inspecteur fungeert als instrumentele variabele: omdat aanvragen willekeurig aan inspecteurs worden toegewezen, en inspecteurs systematisch verschillen in de neiging om octrooien af te keuren, kan ik de kans op vroeg succes als exogeen behandelen en de impact op latere uitkomsten schatten met 2SLS-IV met passende fixed effects modellen.

De resultaten bevestigen de logica van CAD en laten tegelijk een complex verhaal zien wat betreft gender. In algemene zin geldt: wie vroeg succes boekt, krijgt zichtbaarheid, vertrouwen en toegang tot meer samenwerkingen, en die combinatie leidt tot hogere latere output (meer aanvragen, meer citaties). Wie vroeg tegenslag ervaart, ziet juist kansen verschralen—partners haken af, financiering stokt, het zelfvertrouwen daalt—en de kans op hernieuwd succes neemt af. Maar uitgesplitst naar gender ontstaan contrasten. Mannen die vroege tegenslag ervaren, blijken in de regel te kunnen herstellen; hun lagere statuskosten na falen en de beschikbaarheid van mannelijke netwerken maken het waarschijnlijker dat zij alsnog worden gesponsord of opnieuw kansen krijgen. Vrouwen daarentegen die vroege tegenslag ervaren laten een dalend traject zien: ze dienen minder vervolgpatenten in, hun netwerken krimpen en de kans op herintrede daalt. Omgekeerd geldt dat vrouwen met vroeg succes vaak dichter naar de topgroep toe bewegen en soms zelfs de prestaties van succesvolle mannen benaderen of overtreffen—een stereotype-brekend signaal dat deuren opent die anders gesloten blijven.

Deze bevinding nuanceert het idee dat CAD per definitie ongelijkheid vergroot. Mijn resultaten laten zien dat CAD-interacties ongelijkheid kunnen vergroten (voor vrouwen met vroege tegenslag), maar ook verkleinen (doordat vrouwen met vroeg succes en mannen met vroege tegenslag over de tijd naar de topgroep van initieel succesvolle mannen toe bewegen). Daarmee lever ik een theoretische correctie op een te eendimensionale lezing van cumulatie: het mechanisme is niet uniform, maar gevoelig voor status, context en timing. Voor beleid betekent dit dat vroege, rechtvaardige en transparante evaluatiemomenten—met mogelijkheden tot herstel—essentieel zijn, en dat sponsorrelaties niet aan toevallige netwerktoegang of eerste successen mogen worden overgelaten.

Na deze vier studies keer ik terug naar de overkoepelende vraag: wat betekenen integratie en cumulatie samen voor ons begrip van structurele (gender)ongelijkheid op het werk, en voor de manier waarop we meer (gender)gelijkheid proberen te realiseren? Theoretisch laat mijn werk zien dat ongelijkheid relationeel is: het ontstaat en verschuift in netwerken, waar toegang tot informatie en erkenning nooit gelijk verdeeld is. Het is bovendien temporeel: de gevolgen van een beslissing of uitkomst in jaar één werken—via CAD—door in toekomstige loopbanen. En het is contextueel: de opbrengsten van dezelfde netwerkconfiguratie verschillen naar gelang de statushiërarchie van het team of de sector. Samen maken deze inzichten begrijpelijk waarom 'meer vrouwen aannemen' niet automatisch leidt tot gelijke kansen voor vrouwen op de werkvloer.

Praktisch mondt dit uit in drie aanbevelingen. Ten eerste: institutionaliseer integratie als beleidsdoel. Dat betekent dat organisaties niet alleen werven op diversiteit, maar ook actief bouwen aan bruggen—bijvoorbeeld door gemengde teams te organiseren rond echte taakinterdependentie, door belonings- en beoordelingssystemen te koppelen aan samenwerking en kennisdeling, en door sponsors expliciet verantwoordelijk te maken voor het openbreken van gesloten circuits. Ten tweede: investeer vroeg. De eerste evaluatiemomenten, de eerste zichtbare projecten, de eerste mentorrelaties—ze zetten via CAD een pad uit dat moeilijk te hertekenen is. Maak herstel mogelijk, bouw vangnetten in en voorkom dat een enkel incident een loopbaan duurzaam schaadt. Ten derde: evalueer beleid op relaties. Een programma dat aantallen verhoogt maar integratie verslechtert, boekt geen vooruitgang; het verplaatst het probleem.

Natuurlijk kent deze dissertatie ook beperkingen. De netwerken die ik bestudeer zijn primair affectieve samenwerkingsnetwerken; andere netwerken—advies, vriendschap, formele hiërarchische lijnen—kunnen aanvullende of afwijkende patronen laten zien. De GDM-taxonomie die ik hanteer is noodzakelijkerwijs grof; fijnmazige evaluaties van specifieke interventies in één organisatie zouden de mechanieken van relationele verandering nog preciezer kunnen blootleggen. En hoewel ik in hoofdstuk vijf quasi-experimenteel te werk ga, blijven er altijd randvoorwaarden aan instrumentele identificatie; aanvullende designs (bijvoorbeeld natuurlijke experimenten in beoordelingscommissies) kunnen het causale verhaal verder versterken. Ten slotte heb ik mij hier primair op gender gericht; intersectionele dimensies (ras, migratiestatus, leeftijd, klasse) verdienen systematische aandacht, juist omdat statusprocessen elkaar kruisen en versterken.

REFERENCES

Acker, Joan. 1990. "Hierarchies, Jobs, Bodies: A Theory of Gendered Organizations." *Gender & Society* 4(2):139–58. doi:10.1177/089124390004002002.

Acker, Joan. 2006. "Inequality Regimes: Gender, Class, and Race in Organizations." *Gender & Society* 20(4):441–64. doi:10.1177/0891243206289499.

Acker, Joan, and Donald R. Van Houten. 1974. "Differential Recruitment and Control: The Sex Structuring of Organizations." *Administrative Science Quarterly* 152–63.

Adams, Renée B., and Daniel Ferreira. 2009. "Women in the Boardroom and Their Impact on Governance and Performance." *Journal of Financial Economics* 94(2):291–309.

Albrecht, James, Mary Ann Bronson, Peter Skogman Thoursie, and Susan Vroman. 2018. "The Career Dynamics of High-Skilled Women and Men: Evidence from Sweden." *European Economic Review* 105:83–102. doi:10.1016/j.euroecorev.2018.03.012.

Alegria, Sharla. 2019. "Escalator or Step Stool? Gendered Labor and Token Processes in Tech Work." *Gender & Society* 33(5):722–45. doi:10.1177/0891243219835737.

Amis, John M., Johanna Mair, and Kamal A. Munir. 2020. "The Organizational Reproduction of Inequality." *Academy of Management Annals* 14(1):195–230. doi:10.5465/annals.2017.0033.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton university press.

Annis, Barbara, and John Gray. 2013. *Work with Me: The 8 Blind Spots Between Men and Women in Business*. Macmillan.

Arel-Bundock, Vincent, Noah Greifer, and Andrew Heiss. 2024. "How to Interpret Statistical Models Using Marginaleffects for R and Python." *Journal of Statistical Software* 111:1–32. doi:10.18637/jss.v111.i09.

Arthur, Michael B., Svetlana N. Khapova, and Celeste PM Wilderom. 2005. "Career Success in a Boundaryless Career World." *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior* 26(2):177–202.

Arulampalam, Wiji, Alison L. Booth, and Mark L. Bryan. 2007. "Is There a Glass Ceiling over Europe? Exploring the Gender Pay Gap across the Wage Distribution." *ILR Review* 60(2):163–86. doi:10.1177/001979390706000201.

Ashforth, Blake E., and Fred Mael. 1989. "Social Identity Theory and the Organization." *Academy of Management Review* 14(1):20–39.

Aten, Kathryn, Marco DiRenzo, and Dina Shatnawi. 2017. "Gender and Professional E-Networks: Implications of Gender Heterophily on Job Search Facilitation and Outcomes." *Computers in Human Behavior* 72:470–78.

Averett, Susan L., Laura M. Argys, and Saul D. Hoffman. 2018. *The Oxford Handbook of Women and the Economy*. Oxford University Press.

Baltagi, Badi H. 2021. *Econometric Analysis of Panel Data*. Springer Texts in Business and Economics. Cham: Springer International Publishing.

Bantel, Karen A., and Susan E. Jackson. 1989. "Top Management and Innovations in Banking: Does the Composition of the Top Team Make a Difference?" *Strategic Management Journal* 10(S1):107–24.

Baranger, David A. A., Megan C. Finsaas, Brandon L. Goldstein, Colin E. Vize, Donald R. Lynam, and Thomas M. Olino. 2023. "Tutorial: Power Analyses for Interaction Effects in Cross-Sectional Regressions." *Advances in Methods and Practices in Psychological Science* 6(3):25152459231187531. doi:10.1177/25152459231187531.

Barth, Erling, Sari Pekkala Kerr, and Claudia Olivetti. 2021. "The Dynamics of Gender Earnings Differentials: Evidence from Establishment Data." *European Economic Review* 134:103713. doi:10.1016/j.euroecorev.2021.103713.

Bauer, Greta R., Siobhan M. Churchill, Mayuri Mahendran, Chantel Walwyn, Daniel Lizotte, and Alma Angelica Villa-Rueda. 2021. "Intersectionality in Quantitative Research: A

Systematic Review of Its Emergence and Applications of Theory and Methods." *SSM - Population Health* 14:100798. doi:10.1016/j.ssmph.2021.100798.

Becker, Gary S. 1964. "Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education."

Bell, Alex, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen. 2019. "Who Becomes an Inventor in America? The Importance of Exposure to Innovation*." *The Quarterly Journal of Economics* 134(2):647–713. doi:10.1093/qje/qjy028.

Belliveau, Maura A. 2005. "Blind Ambition? The Effects of Social Networks and Institutional Sex Composition on the Job Search Outcomes of Elite Coeducational and Women's College Graduates." *Organization Science* 16(2):134–50.

Bellotti, Elisa, Dominika Czerniawska, Martin G. Everett, and Luigi Guadalupi. 2022. "Gender Inequalities in Research Funding: Unequal Network Configurations, or Unequal Network Returns?" *Social Networks* 70:138–51. doi:10.1016/j.socnet.2021.12.007.

Ben-Shachar, Mattan S., Indrajeet Patil, Rémi Thériault, Brenton M. Wiernik, and Daniel Lüdecke. 2023. "Phi, Fei, Fo, Fum: Effect Sizes for Categorical Data That Use the Chi-Squared Statistic." *Mathematics* 11(9):1982. doi:10.3390/math11091982.

Berger, Joseph, M. Hamit Fisek, Robert Z. Norman, and Morris Zelditch. 1977. "Status Characteristics in Social Interactions: An Expectations States Approach."

Berger, Joseph, Susan J. Rosenholtz, and Morris Zelditch. 1980. "Status Organizing Processes." *Annual Review of Sociology* 6(1):479–508. doi:10.1146/annurev.so.06.080180.002403.

Bernardi, Fabrizio. 2014. "Compensatory Advantage as a Mechanism of Educational Inequality: A Regression Discontinuity Based on Month of Birth." *Sociology of Education* 87(2):74–88. doi:10.1177/0038040714524258.

Bernardi, Fabrizio, and Moris Triventi. 2020. "Compensatory Advantage in Educational Transitions: Trivial or Substantial? A Simulated Scenario Analysis." *Acta Sociologica* 63(1):40–62. doi:10.1177/0001699318780950.

Bettio, Francesca, and Alina Verashchagina. 2009. *Gender Segregation in the Labour Market : Root Causes, Implications and Policy Responses in the EU*. LU: Publications Office.

Bian, Bo, Jingjing Li, and Kai Li. 2024. "Does Mandating Women on Corporate Boards Backfire?"

Biegert, Thomas, Michael Kühhirt, and Wim Van Lancker. 2023. "They Can't All Be Stars: The Matthew Effect, Cumulative Status Bias, and Status Persistence in NBA All-Star Elections." *American Sociological Review* 00031224231159139. doi:10.1177/00031224231159139.

Blau, Francine D., and Lawrence M. Kahn. 2012. "'The Gender Pay Gap: Have Women Gone as Far as They Can?'" in *Inequality in the United States*. Routledge.

Blau, Francine D., and Lawrence M. Kahn. 2017. "The Gender Wage Gap: Extent, Trends, and Explanations." *Journal of Economic Literature* 55(3):789–865.

Blau, Peter M., and Otis Dudley Duncan. 1967. *The American Occupational Structure*. John Wiley & Sons, Inc.

Blickenstaff, Jacob Clark. 2005. "Women and Science Careers: Leaky Pipeline or Gender Filter?" *Gender and Education* 17(4):369–86. doi:10.1080/09540250500145072.

Block, Per, and Thomas Grund. 2014. "Multidimensional Homophily in Friendship Networks*." *Network Science* 2(2):189–212. doi:10.1017/nws.2014.17.

Bohnet, Iris, Alexandra van Geen, and Max Bazerman. 2016. "When Performance Trumps Gender Bias: Joint vs. Separate Evaluation." *Management Science* 62(5):1225–34.

Bol, Thijs, Mathijs de Vaan, and Arnout van de Rijt. 2018. "The Matthew Effect in Science Funding." *Proceedings of the National Academy of Sciences* 115(19):4887–90. doi:10.1073/pnas.1719557115.

Bowen, Chieh-Chen, Janet K. Swim, and Rick R. Jacobs. 2000. "Evaluating Gender Biases on Actual Job Performance of Real People: A Meta-Analysis." *Journal of Applied Social Psychology* 30(10):2194–2215. doi:10.1111/j.1559-1816.2000.tb02432.x.

Bowlby, John. 2005. *The Making and Breaking of Affectional Bonds*. London: Routledge.

Bowles, Hannah Riley, Linda Babcock, and Lei Lai. 2007. "Social Incentives for Gender Differences in the Propensity to Initiate Negotiations: Sometimes It Does Hurt to Ask." *Organizational Behavior and Human Decision Processes* 103(1):84–103. doi:10.1016/j.obhdp.2006.09.001.

Bowles, Hannah Riley, and Kathleen L. McGinn. 2008. "Gender in Job Negotiations: A Two-level Game." *Negotiation Journal* 24(4):393–410.

Boyle, Matthew and Green, Jeff. 2023. "Work Shift: Women CEOs (Finally) Outnumber Those Named John." *Bloomberg.Com*, April 25.

Braeken, Johan, Joris Mulder, and Stephen Wood. 2015. "Relative Effects at Work: Bayes Factors for Order Hypotheses." *Journal of Management* 41(2):544–73.

Brambor, Thomas, William Roberts Clark, and Matt Golder. 2006. "Understanding Interaction Models: Improving Empirical Analyses." *Political Analysis* 14(1):63–82.

Burt, Ronald S. 1998. "The Gender of Social Capital." *Rationality and Society* 10(1):5–46. doi:10.1177/104346398010001001.

van Buuren, Stef, and Karin Groothuis-Oudshoorn. 2010. "Mice: Multivariate Imputation by Chained Equations in R." *Journal of Statistical Software* 1–68.

van Buuren, Stef, Karin Groothuis-Oudshoorn, Alexander Robitzsch, Gerko Vink, Lisa Doove, and Shahab Jolani. 2015. "Package 'Mice.'" *Computer Software*.

Campbell, Lesley G., Siya Mehtani, Mary E. Dozier, and Janice Rinehart. 2013. "Gender-Heterogeneous Working Groups Produce Higher Quality Science." *PLOS ONE* 8(10):e79147. doi:10.1371/journal.pone.0079147.

Carley, Michael, Deepak Hedge, and Alan Marco. 2015. "What Is the Probability of Receiving a U.S. Patent." *Yale Journal of Law and Technology* 17:203–82.

Cerda-Jara, Michael, David J. Harding, and The Underground Scholars Research Cohort. 2024. "Criminal Record Stigma in the Labor Market for College Graduates: A Mixed Methods Study." *Sociological Science* 11:42–66. doi:10.15195/v11.a2.

Cha, Youngjoo, Kim A. Weeden, and Landon Schnabel. 2023. "Is the Gender Wage Gap Really a Family Wage Gap in Disguise?" *American Sociological Review* 88(6):972–1001. doi:10.1177/00031224231212464.

Cheryan, Sapna, Sianna A. Ziegler, Amanda K. Montoya, and Lily Jiang. 2017. "Why Are Some STEM Fields More Gender Balanced than Others?" *Psychological Bulletin* 143(1):1–35. doi:10.1037/bul0000052.

Cho, Sumi, Kimberlé Williams Crenshaw, and Leslie McCall. 2013. "Toward a Field of Intersectionality Studies: Theory, Applications, and Praxis." *Signs: Journal of Women in Culture and Society* 38(4):785–810. doi:10.1086/669608.

Cockburn, Cynthia. 1988. "Machinery of Dominance : Women, Men and Technical Know-How." doi:10.3406/apre.1988.864.

Cohen, Bernard P., and Xueguang Zhou. 1991. "Status Processes in Enduring Work Groups." *American Sociological Review* 179–88.

Cohen, Jacob. 1992. "A Power Primer." *Psychological Bulletin* 112(1):155.

Cohen, Jacob. 1994. "The Earth Is Round (P<. 05)." *American Psychologist* 49(12):997.

Cole, Jonathan R., and Burton Singer. 1991. "A Theory of Limited Differences: Explaining the Productivity Puzzle in Science." Pp. 277–310 in *The outer circle: Women in the scientific community*. Norton, New York.

Colyvas, Jeannette A., Kaisa Snellman, Janet Bercovitz, and Maryann Feldman. 2012. "Disentangling Effort and Performance: A Renewed Look at Gender Differences in Commercializing Medical School Research." *The Journal of Technology Transfer* 37(4):478–89. doi:10.1007/s10961-011-9235-6.

Costa Dias, Monica, Robert Joyce, and Francesca Parodi. 2018. *Wage Progression and the Gender Wage Gap: The Causal Impact of Hours of Work*. doi:10.1920/BN.IFS.2018.BN0223.

Cracau, Daniel, and José E. Durán Lima. 2016. “On the Normalized Herfindahl-Hirschman Index: A Technical Note.” *Int. J. Food System Dynamics* 7(4):382–86.

Cullen, Zoë B., and Ricardo Perez-Truglia. 2019. *The Old Boys' Club: Schmoozing and the Gender Gap*. No. 26530. National Bureau of Economic Research.

Cullen, Zoë, and Ricardo Perez-Truglia. 2023. “The Old Boys Club: Schmoozing and the Gender Gap.” *American Economic Review* 113(7):1703–40. doi:10.1257/aer.20210863.

Danbold, Felix, and Corinne Bendersky. 2020. “Balancing Professional Prototypes Increases the Valuation of Women in Male-Dominated Professions.” *Organization Science* 31(1):119–40. doi:10.1287/orsc.2019.1288.

Dannefer, Dale. 2003. “Cumulative Advantage/Disadvantage and the Life Course: Cross-Fertilizing Age and Social Science Theory.” *The Journals of Gerontology: Series B* 58(6):S327–37. doi:10.1093/geronb/58.6.S327.

Dannefer, Dale. 2020. “Systemic and Reflexive: Foundations of Cumulative Dis/Advantage and Life-Course Processes.” *The Journals of Gerontology: Series B* 75(6):1249–63. doi:10.1093/geronb/gby118.

Dechezleprêtre, Antoine, Elias Einiö, Ralf Martin, Kieu-Trang Nguyen, and John Van Reenen. 2023. “Do Tax Incentives Increase Firm Innovation? An RD Design for R&D, Patents, and Spillovers.” *American Economic Journal: Economic Policy* 15(4):486–521. doi:10.1257/pol.20200739.

DiPrete, Thomas A., and Gregory M. Eirich. 2006. “Cumulative Advantage as a Mechanism for Inequality: A Review of Theoretical and Empirical Developments.” *Annual Review of Sociology* 32:271–97.

Dobbin, Frank, Alexandra Kalev, and Erin Kelly. 2007. “Diversity Management in Corporate America.” *Contexts* 6(4):21–27. doi:10.1525/ctx.2007.6.4.21.

DORA. 2025. “About DORA.” https://sfdora.org/about-dora/.

Dover, Tessa L., Cheryl R. Kaiser, and Brenda Major. 2019. “Mixed Signals: The Unintended Effects of Diversity Initiatives.” *Social Issues and Policy Review*.

Duchin, Ran, Mikhail Simutin, and Denis Sosyura. 2021. “The Origins and Real Effects of the Gender Gap: Evidence from CEOs' Formative Years.” *The Review of Financial Studies* 34(2):700–762. doi:10.1093/rfs/hhaa068.

Eagly, Alice H., and Linda L. Carli. 2007. *Through the Labyrinth: The Truth About How Women Become Leaders*. Harvard Business Review Press.

Eagly, Alice H., Christa Nater, David I. Miller, Michèle Kaufmann, and Sabine Sczesny. 2020. “Gender Stereotypes Have Changed: A Cross-Temporal Meta-Analysis of U.S. Public Opinion Polls from 1946 to 2018.” *American Psychologist* 75(3):301–15. doi:10.1037/amp0000494.

Egan, Mark, Gregor Matvos, and Amit Seru. 2022. “When Harry Fired Sally: The Double Standard in Punishing Misconduct.” *Journal of Political Economy* 130(5):1184–1248. doi:10.1086/718964.

Elaine Spector. 2024. “Challenging the ‘Diversity Hire’ Label: The Need for Equal Standards in the Workplace.” https://harrityllp.com/challenging-the-diversity-hire-label-the-need-for-equal-standards-in-the-workplace/.

Ellemers, N. (2018). Gender stereotypes. *Annual review of psychology*, *69*(1), 275-298.

Ellwood, David T. 1982. “Teenage Unemployment: Permanent Scars or Temporary Blemishes?” Pp. 349–90 in *The Youth Labor Market Problem: Its Nature, Causes, and Consequences*. University of Chicago Press.

Ely, Robin J., and Debra E. Meyerson. 2000. “Theories of Gender in Organizations: A New Approach to Organizational Analysis and Change.” *Research in Organizational Behavior* 22:103–51.

England, Paula, Jonathan Bearak, Michelle J. Budig, and Melissa J. Hodges. 2016. “Do Highly Paid, Highly Skilled Women Experience the Largest Motherhood Penalty?” *American Sociological Review* 81(6):1161–89. doi:10.1177/0003122416673598.

Engqvist, Leif. 2005. “The Mistreatment of Covariate Interaction Terms in Linear Model Analyses of Behavioural and Evolutionary Ecology Studies.” *Animal Behaviour* 70(4):967–71. doi:10.1016/j.anbehav.2005.01.016.

Erola, Jani, and Elina Kilpi-Jakonen. 2017. “Compensation and Other Forms of Accumulation in Intergenerational Social Inequality.” *Social Inequality Across the Generations* 3–24.

Ertug, Gokhan, Martin Gargiulo, Charles Galunic, and Tengjian Zou. 2018. “Homophily and Individual Performance.” *Organization Science* 29(5):912–30. doi:10.1287/orsc.2018.1208.

Etzkowitz, Henry, Carol Kemelgor, and Brian Uzzi. 2000. *Athena Unbound: The Advancement of Women in Science and Technology*. Cambridge University Press.

European Commission, Directorate-General for Justice and Consumers. 2018. *Report on Equality between Women and Men in the EU 2018*. DS-AU-18-001-EN-N. Luxembourg.

Fairhurst, Gail Theus, and B. Kay Snavely. 1983. “A Test of the Social Isolation of Male Tokens.” *Academy of Management Journal* 26(2):353–61.

Farre-Mensa, Joan, Deepak Hegde, and Alexander Ljungqvist. 2020. “What Is a Patent Worth? Evidence from the U.S. Patent ‘Lottery.’” *The Journal of Finance* 75(2):639–82. doi:10.1111/jofi.12867.

Feng, Josh, and Xavier Jaravel. 2020. “Crafting Intellectual Property Rights: Implications for Patent Assertion Entities, Litigation, and Innovation.” *American Economic Journal: Applied Economics* 12(1):140–81. doi:10.1257/app.20180361.

Fine, Cordelia, Victor Sojo, and Holly Lawford-Smith. 2020. “Why Does Workplace Gender Diversity Matter? Justice, Organizational Benefits, and Policy.” *Social Issues and Policy Review* 14(1):36–72. doi:10.1111/sipr.12064.

Finsaas, Megan C., and Brandon L. Goldstein. 2021. “Do Simple Slopes Follow-up Tests Lead Us Astray? Advancements in the Visualization and Reporting of Interactions.” *Psychological Methods* 26(1):38–60. doi:10.1037/met0000266.

Forret, Monica L., and Thomas W. Dougherty. 2004. “Networking Behaviors and Career Outcomes: Differences for Men and Women?” *Journal of Organizational Behavior* 25(3):419–37. doi:10.1002/job.253.

Fortin, Nicole M., and Michael Huberman. 2002. “Occupational Gender Segregation and Women’s Wages in Canada: An Historical Perspective.” *Canadian Public Policy / Analyse de Politiques* 28:S11–39. doi:10.2307/3552342.

Fox, Mary Frank, Carolyn Fonseca, and Jinghui Bao. 2011. “Work and Family Conflict in Academic Science: Patterns and Predictors among Women and Men in Research Universities.” *Social Studies of Science* 41(5):715–35. doi:10.1177/0306312711417730.

Fry, Rick, Brian Kennedy, and Cary Funk. 2021. *STEM Jobs See Uneven Progress in Increasing Gender, Racial and Ethnic Diversity*. Pew Research Center. https://www.pewresearch.org/wp-content/uploads/sites/20/2021/03/PS_2021.04.01_diversity-in-STEM_REPORT.pdf.

Ganzeboom, Harry B. G., and Donald J. Treiman. 2003. “Three Internationally Standardised Measures for Comparative Research on Occupational Status.” Pp. 159–93 in *Advances in Cross-National Comparison: A European Working Book for Demographic and Socio-Economic Variables*, edited by J. H. P. Hoffmeyer-Zlotnik and C. Wolf. Boston, MA: Springer US.

Garofalo, Sara, Sara Giovagnoli, Matteo Orsoni, Francesca Starita, and Mariagrazia Benassi. 2022. “Interaction Effect: Are You Doing the Right Thing?” *PLoS ONE* 17(7):e0271668. doi:10.1371/journal.pone.0271668.

Garrison, Gary, Robin L. Wakefield, Xiaobo Xu, and Sang Hyun Kim. 2010. "Globally Distributed Teams: The Effect of Diversity on Trust, Cohesion and Individual Performance." *ACM SIGMIS Database: The DATABASE for Advances in Information Systems* 41(3):27–48.

Gaulé, Patrick. 2018. "Patents and the Success of Venture-Capital Backed Startups: Using Examiner Assignment to Estimate Causal Effects." *The Journal of Industrial Economics* 66(2):350–76. doi:10.1111/joie.12168.

Georgeac, Oriane A. M., and Aneeta Rattan. 2023. "The Business Case for Diversity Backfires: Detrimental Effects of Organizations' Instrumental Diversity Rhetoric for Underrepresented Group Members' Sense of Belonging." *Journal of Personality and Social Psychology* 124(1):69–108. doi:10.1037/pspi0000394.

Gertsberg, Marina, Johanna Mollerstrom, and Michaela Pagel. 2021. "Gender Quotas and Support for Women in Board Elections."

Giannetti, Mariassunta, and Tracy Yue Wang. 2023. "Public Attention to Gender Equality and Board Gender Diversity." *Journal of Financial and Quantitative Analysis* 58(2):485–511. doi:10.1017/S0022109022000400.

Gibson, Carter, Jay H. Hardy III, and M. Ronald Buckley. 2014. "Understanding the Role of Networking in Organizations." *Career Development International* 19(2):146–61. doi:10.1108/CDI-09-2013-0111.

Giczy, Alexander V., Nicholas A. Pairolero, and Andrew A. Toole. 2024. "Discovering Value: Women's Participation in University and Commercial AI Invention." *Nature Biotechnology* 42(1):26–29. doi:10.1038/s41587-023-02075-1.

Goldin, Claudia. 2014. "A Grand Gender Convergence: Its Last Chapter." *American Economic Review* 104(4):1091–1119. doi:10.1257/aer.104.4.1091.

Goldin, Claudia Dale. 2021. *Career and Family: Women's Century-Long Journey toward Equity*. Princeton, New Jersey: Princeton University Press.

González, M. José, Clara Cortina, and Jorge Rodríguez. 2019. "The Role of Gender Stereotypes in Hiring: A Field Experiment." *European Sociological Review* 35(2):187–204. doi:10.1093/esr/jcy055.

Gould, R. V. 2002. "The Origins of Status Hierarchies: A Formal Theory and Empirical Test." *American Journal of Sociology* 107(5):1143–78. doi:10.1086/341744.

Graham, Stuart JH, Alan C. Marco, and Richard Miller. 2015. "The USPTO Patent Examination Research Dataset: A Window on the Process of Patent Examination." *Georgia Tech Scheller College of Business Research Paper No. WP* 43. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2729322.

Green, Clifton, Narasimhan Jegadeesh, and Yue Tang. 2009. "Gender and Job Performance: Evidence from Wall Street." *Financial Analysts Journal* 65(6):65–78. doi:10.2469/faj.v65.n6.1.

Greenwald, Ag, and Mr Banaji. 1995. "Implicit Social Cognition - Attitudes, Self-Esteem, and Stereotypes." *Psychological Review* 102(1):4–27. doi:10.1037/0033-295X.102.1.4.

Grissom, Jason A., Jill Nicholson-Crotty, and Lael Keiser. 2012. "Does My Boss's Gender Matter? Explaining Job Satisfaction and Employee Turnover in the Public Sector." *Journal of Public Administration Research and Theory* 22(4):649–73.

Gu, Xin, Joris Mulder, and Herbert Hoijtink. 2018. "Approximated Adjusted Fractional Bayes Factors: A General Method for Testing Informative Hypotheses." *British Journal of Mathematical and Statistical Psychology* 71(2):229–61. doi:10.1111/bmsp.12110.

Gupta, Anil K., and Vijay Govindarajan. 2000. "Knowledge Flows within Multinational Corporations." *Strategic Management Journal* 21(4):473–96.

Hall, Bronwyn H., and Josh Lerner. 2010. "Chapter 14 - The Financing of R&D and Innovation." Pp. 609–39 in *Handbook of the Economics of Innovation*. Vol. 1, *Handbook of The Economics of Innovation, Vol. 1*, edited by B. H. Hall and N. Rosenberg. North-Holland.

Hall, William M., Toni Schmader, and Elizabeth Croft. 2015. "Engineering Exchanges: Daily Social Identity Threat Predicts Burnout Among Female Engineers." *Social Psychological and Personality Science* 6(5):528–34. doi:10.1177/1948550615572637.

Hansen, Morten T. 1999. "The Search-Transfer Problem: The Role of Weak Ties in Sharing Knowledge across Organization Subunits." *Administrative Science Quarterly* 44(1):82–111. doi:10.2307/2667032.

Harrison, David A., and Katherine J. Klein. 2007. "What's the Difference? Diversity Constructs as Separation, Variety, or Disparity in Organizations." *Academy of Management Review* 32(4):1199–1228.

He, Joyce C., Sonia K. Kang, and Nicola Lacetera. 2021. "Opt-out Choice Framing Attenuates Gender Differences in the Decision to Compete in the Laboratory and in the Field." *Proceedings of the National Academy of Sciences* 118(42):e2108337118. doi:10.1073/pnas.2108337118.

Hegde, Deepak, Alexander Ljungqvist, and Manav Raj. 2022. "Quick or Broad Patents? Evidence from U.S. Startups." *The Review of Financial Studies* 35(6):2705–42. doi:10.1093/rfs/hhab097.

Heilman, Madeline E., Aaron S. Wallen, Daniella Fuchs, and Melinda M. Tamkins. 2004. "Penalties for Success: Reactions to Women Who Succeed at Male Gender-Typed Tasks." *Journal of Applied Psychology* 89(3):416–27. doi:10.1037/0021-9010.89.3.416.

Herbaut, Estelle. 2021. "Overcoming Failure in Higher Education: Social Inequalities and Compensatory Advantage in Dropout Patterns." *Acta Sociologica* 64(4):383–402. doi:10.1177/0001699320920916.

Herring, Cedric. 2009. "Does Diversity Pay?: Race, Gender, and the Business Case for Diversity." *American Sociological Review* 74(2):208–24.

Herring, Cedric, and Loren Henderson. 2015. "Diversity in Organizations." *A Critical Examination*.

Hill, Aaron D., Arun D. Upadhyay, and Rafik I. Beekun. 2015. "Do Female and Ethnically Diverse Executives Endure Inequity in the CEO Position or Do They Benefit from Their Minority Status? An Empirical Examination." *Strategic Management Journal* 36(8):1115–34. doi:10.1002/smj.2274.

Hinchliffe, Emma. 2017. "What Ellen Pao's Case—and Advocacy—Means to Women." https://mashable.com/feature/ellen-pao-women-in-tech.

Hofstra, Bas, Vivek V. Kulkarni, Sebastian Munoz-Najar Galvez, Bryan He, Dan Jurafsky, and Daniel A. McFarland. 2020. "The Diversity–Innovation Paradox in Science." *Proceedings of the National Academy of Sciences* 117(17):9284–91. doi:10.1073/pnas.1915378117.

Hoijtink, Herbert. 2011. *Informative Hypotheses: Theory and Practice for Behavioral and Social Scientists*. CRC Press.

Hoijtink, Herbert, Xin Gu, and Joris Mulder. 2019. "Bayesian Evaluation of Informative Hypotheses for Multiple Populations." *British Journal of Mathematical and Statistical Psychology* 72(2):219–43.

Hoisl, Karin, and Myriam Mariani. 2017. "It's a Man's Job: Income and the Gender Gap in Industrial Research." *Management Science* 63(3):766–90. doi:10.1287/mnsc.2015.2357.

Horak, Sven, and Yuliani Suseno. 2023. "Informal Networks, Informal Institutions, and Social Exclusion in the Workplace: Insights from Subsidiaries of Multinational Corporations in Korea." *Journal of Business Ethics* 186(3):633–55. doi:10.1007/s10551-022-05244-5.

Huang, Junming, Alexander J. Gates, Roberta Sinatra, and Albert-László Barabási. 2020. "Historical Comparison of Gender Inequality in Scientific Careers across Countries and Disciplines." *Proceedings of the National Academy of Sciences* 117(9):4609–16. doi:10.1073/pnas.1914221117.

Huang, Ruidi, Erik J. Mayer, and Darius P. Miller. 2024. "Gender Bias in Promotions: Evidence from Financial Institutions." *The Review of Financial Studies* 37(5):1685–1728. doi:10.1093/rfs/hhad079.

Huber, John C. 1998. "Cumulative Advantage and Success-Breeds-Success: The Value of Time Pattern Analysis." *Journal of the American Society for Information Science* 49(5):471–76. doi:10.1002/(SICI)1097-4571(19980415)49:5<471::AID-ASI8>3.0.CO;2-T.

Huffman, Matt L., Philip N. Cohen, and Jessica Pearlman. 2010. "Engendering Change: Organizational Dynamics and Workplace Gender Desegregation, 1975–2005." *Administrative Science Quarterly* 55(2):255–77.

Huisman, Mark. 2014. "Imputation of Missing Network Data: Some Simple Procedures." Pp. 707–15 in *Encyclopedia of Social Network Analysis and Mining*, edited by R. Alhajj and J. Rokne. New York, NY: Springer New York.

Hunt, Jennifer, Jean-Philippe Garant, Hannah Herman, and David J. Munroe. 2012. "Why Don't Women Patent?"

Hyde, Janet S., Elizabeth Fennema, and Susan J. Lamon. 1990. "Gender Differences in Mathematics Performance: A Meta-Analysis." *Psychological Bulletin* 107(2):139–55. doi:10.1037/0033-2909.107.2.139.

Ibarra, Herminia. 1992. "Homophily and Differential Returns: Sex Differences in Network Structure and Access in an Advertising Firm." *Administrative Science Quarterly* 37(3):422. doi:10.2307/2393451.

Ibarra, Herminia. 1997. "Paving an Alternative Route: Gender Differences in Managerial Networks." *Social Psychology Quarterly* 91–102.

Ibarra, Herminia, Nancy M. Carter, and Christine Silva. 2010. "Why Men Still Get More Promotions than Women." *Harvard Business Review* 88(9):80–85.

ILO. 1931. *Yearbook of Labour Statistics*. 1. Geneva.

ILO. 2016. *Key Indicators of the Labour Market (KILM), Ninth Edition*. 9th ed.

IWPR. 2024. *Women Earn Less than Men Whether They Work in the Same or Different Occupations*. IWPR #C521. https://iwpr.org/equal-pay-day-2024/, https://iwpr.org/equal-pay-day-2024/.

Jones, Stacy. 2017. "White Men Account for 72% of Corporate Leadership at 16 of the Fortune 500 Companies." *Fortune*, June 9.

Judge, Timothy A., Daniel M. Cable, John W. Boudreau, and Robert D. Bretz Jr. 1995. "An Empirical Investigation of the Predictors of Executive Career Success." *Personnel Psychology* 48(3):485–519.

Juhász, Borbála, and Enikő Pap. 2018. *Backlash in Gender Equality and Women's and Girls' Rights*. PE 604.955. Women's Rights & Gender Equality. Brussels.

Kahan, Brennan C. 2013. "Bias in Randomised Factorial Trials." *Statistics in Medicine* 32(26):4540–49. doi:10.1002/sim.5869.

Kalev, Alexandra, Frank Dobbin, and Erin Kelly. 2006. "Best Practices or Best Guesses? Assessing the Efficacy of Corporate Affirmative Action and Diversity Policies." *American Sociological Review* 71(4):589–617.

Kanter, Rosabeth Moss. 1977a. "(1977a). Men and Women of the Corporation. New York: Basic Books."

Kanter, Rosabeth Moss. 1977b. "Some Effects of Proportions on Group Life." Pp. 53–78 in *The Gender Gap in Psychotherapy*. Springer.

Kanter, Rosabeth Moss. 1977c. "Some Effects of Proportions on Group Life: Skewed Sex Ratios and Responses to Token Women." *American Journal of Sociology* 82:965–90.

Karasek Jr, Robert A. 1979. "Job Demands, Job Decision Latitude, and Mental Strain: Implications for Job Redesign." *Administrative Science Quarterly* 285–308.

Kasper Gossink-Melenhorst and NWO. 2019. "Quality over Quantity: How the Dutch Research Council Is Giving Researchers the Opportunity to Showcase Diverse Types of Talent." https://sfdora.org/2019/11/14/quality-over-quantity-how-the-dutch-research-council-is-giving-researchers-the-opportunity-to-showcase-diverse-types-of-talent/.

Kelley, Heather, Quinn Galbraith, and Jessica Strong. 2020. "Working Moms: Motherhood Penalty or Motherhood Return?" *The Journal of Academic Librarianship* 46(1):102075. doi:10.1016/j.acalib.2019.102075.

Kelly, Erin L., Samantha K. Ammons, Kelly Chermack, and Phyllis Moen. 2010. "Gendered Challenge, Gendered Response: Confronting the Ideal Worker Norm in a White-Collar Organization." *Gender & Society* 24(3):281–303. doi:10.1177/0891243210372073.

Kerin, Ra, Pr Varadarajan, and Ra Peterson. 1992. "1st-Mover Advantage - a Synthesis, Conceptual-Framework, and Research Propositions." *Journal of Marketing* 56(4):33–52. doi:10.2307/1251985.

Kleinert, Simon, and Kazem Mochkabadi. 2022. "Gender Stereotypes in Equity Crowdfunding: The Effect of Gender Bias on the Interpretation of Quality Signals." *The Journal of Technology Transfer* 47(6):1640–61. doi:10.1007/s10961-021-09892-z.

Kluytmans, Anouck, Rens Van De Schoot, and Herbert Hoijtink. 2012. "Illustrating Bayesian Evaluation of Informative Hypotheses for Regression Models." *Frontiers in Psychology* 3:2.

Kolb, Deborah M. 2009. "Too Bad for the Women or Does It Have to Be? Gender and Negotiation Research over the Past Twenty-five Years." *Negotiation Journal* 25(4):515–31.

Köllen, Thomas, Marja-Liisa Kakkuri-Knuuttila, and Regine Bendl. 2018. "An Indisputable 'Holy Trinity'? On the Moral Value of Equality, Diversity, and Inclusion." *Equality, Diversity and Inclusion: An International Journal* 37(5):438–49. doi:10.1108/EDI-04-2018-0072.

Koning, Rembrand, Sampsa Samila, and John-Paul Ferguson. 2021. "Who Do We Invent for? Patents by Women Focus More on Women's Health, but Few Women Get to Invent." *Science* 372(6548):1345–48. doi:10.1126/science.aba6990.

Kram, Kathy E. 1988. *Mentoring at Work: Developmental Relationships in Organizational Life*. University Press of America.

Krivkovich, A., E. Field, L. Yee, M. McConnell, and H. Smith. 2024. "Women in the Workplace 2024: The 10th-Anniversary Report." *McKinsey & Company*.

Kumar, Alok. 2010. "Self-Selection and the Forecasting Abilities of Female Equity Analysts." *Journal of Accounting Research* 48(2):393–435. doi:10.1111/j.1475-679X.2009.00362.x.

Kwiek, Marek, and Wojciech Roszka. 2021. "Gender-Based Homophily in Research: A Large-Scale Study of Man-Woman Collaboration." *Journal of Informetrics* 15(3):101171. doi:10.1016/j.joi.2021.101171.

LaFrance, Adrienne. 2015. "'Manned' Spaceflight Is So 20th Century." https://www.theatlantic.com/technology/archive/2015/12/astronauts-not-astronauttes/419388/.

Lax Martínez, Gema, Julio Raffo, and Kaori Saito. 2016. *Identifying the Gender of PCT Inventors*. World Intellectual Property Organization-Economics and Statistics Division.

Leeb, Hannes, and Benedikt M. Pötscher. 2005. "Model Selection and Inference: Facts and Fiction." *Econometric Theory* 21(01). doi:10.1017/S0266466605050036.

Leibbrandt, Andreas, Liang Choon Wang, and Cordelia Foo. 2018. "Gender Quotas, Competitions, and Peer Review: Experimental Evidence on the Backlash Against Women." *Management Science* 64(8):3501–16. doi:10.1287/mnsc.2017.2772.

Leigh, Angelica, and Sreedhari D. Desai. 2023. "What's Race Got to Do with It? The Interactive Effect of Race and Gender on Negotiation Offers and Outcomes." *Organization Science* 34(2):935–58. doi:10.1287/orsc.2022.1629.

Lemley, Mark A., and Bhaven Sampat. 2008. "Is the Patent Office a Rubber Stamp The Future of Law, Religion, and the Family - A 25th Anniversary Symposium: Essay." *Emory Law Journal* 58(1):181–206.

Lemley, Mark A., and Bhaven Sampat. 2012. "Examiner Characteristics and Patent Office Outcomes." *The Review of Economics and Statistics* 94(3):817–27. doi:10.1162/REST_a_00194.

Leslie, Lisa M., Colleen Flaherty Manchester, and Patricia C. Dahm. 2017. "Why and When Does the Gender Gap Reverse? Diversity Goals and the Pay Premium for High Potential Women." *Academy of Management Journal* 60(2):402–32. doi:10.5465/amj.2015.0195.

Leszczyńska, Monika. 2018. "Mandatory Quotas for Women on Boards of Directors in the European Union: Harmful to or Good for Company Performance?" *European Business Organization Law Review* 19(1):35–61.

Lindberg, Sara M., Janet Shibley Hyde, Jennifer L. Petersen, and Marcia C. Linn. 2010. "New Trends in Gender and Mathematics Performance: A Meta-Analysis." *Psychological Bulletin* 136(6):1123–35. doi:10.1037/a0021276.

Long, J. Scott. 1992. "Measures of Sex Differences in Scientific Productivity*." *Social Forces* 71(1):159–78. doi:10.1093/sf/71.1.159.

Long, J. Scott, and Mary Frank Fox. 1995. "Scientific Careers: Universalism and Particularism." *Annual Review of Sociology* 21(Volume 21, 1995):45–71. doi:10.1146/annurev.so.21.080195.000401.

Lovell, Bronwyn. 2021. "Sex and the Stars: The Enduring Structure of Gender Discrimination in the Space Industry." *Journal of Feminist Scholarship* 18(18):61–77. doi:10.23860/jfs.2021.18.04.

Luhr, Sigrid. 2024. "Engineering Inequality: Informal Coaching, Glass Walls, and Social Closure in Silicon Valley." *American Journal of Sociology*. doi:10.1086/729506.

Manning, Alan, and Joanna Swaffield. 2008. "The Gender Gap in Early-Career Wage Growth." *The Economic Journal* 118(530):983–1024.

Mayrhofer, Wolfgang, Michael Meyer, Michael Schiffinger, and Angelika Schmidt. 2008. "The Influence of Family Responsibilities, Career Fields and Gender on Career Success: An Empirical Study." *Journal of Managerial Psychology*.

McClelland, Gary H., and Charles M. Judd. 1993. "Statistical Difficulties of Detecting Interactions and Moderator Effects." *Psychological Bulletin* 114(2):376–90. doi:10.1037/0033-2909.114.2.376.

McGuire, Gail M. 2002. "Gender, Race, and the Shadow Structure: A Study of Informal Networks and Inequality in a Work Organization." *Gender & Society* 16(3):303–22. doi:10.1177/0891243202016003003.

Melamed, Tuvia. 1996. "Career Success: An Assessment of a Gender-specific Model." *Journal of Occupational and Organizational Psychology* 69(3):217–42.

Merton, Robert K. 1968. "The Matthew Effect in Science." *Science* 159(3810):56–63.

Michael, James, and Gary Yukl. 1993. "Managerial Level and Subunit Function as Determinants of Networking Behavior in Organizations." *Group & Organization Management* 18(3):328–51. doi:10.1177/1059601193183005.

Mincer, Jacob, and Solomon Polachek. 1978. "An Exchange: The Theory of Human Capital and the Earnings of Women: Women's Earnings Reexamined." *The Journal of Human Resources* 13(1):118–34. doi:10.2307/145305.

Moore, Elaine. 2019. "How Ellen Pao Lost a Lawsuit, but Won a Hearing." *Financial Times*, March 8.

Morgan, Allison C., Samuel F. Way, Michael J. D. Hoefer, Daniel B. Larremore, Mirta Galesic, and Aaron Clauset. 2021. "The Unequal Impact of Parenthood in Academia." *Science Advances* 7(9):eabd1996. doi:10.1126/sciadv.abd1996.

Muralidharan, Karthik, Mauricio Romero, and Kaspar Wüthrich. 2023. "Factorial Designs, Model Selection, and (Incorrect) Inference in Randomized Experiments." *The Review of Economics and Statistics* 1–44. doi:10.1162/rest_a_01317.

Murphy, Kevin R., and Craig J. Russell. 2017. "Mend It or End It: Redirecting the Search for Interactions in the Organizational Sciences." *Organizational Research Methods* 20(4):549–73. doi:10.1177/1094428115625322.

Muzio, Daniel, and Jennifer Tomlinson. 2012. "Editorial: Researching Gender, Inclusion and Diversity in Contemporary Professions and Professional Organizations." *Gender, Work & Organization* 19(5):455–66. doi:10.1111/j.1468-0432.2012.00608.x.

NASA. 2019. "Rocket Fuel in Her Blood: The Story of JoAnn Morgan - NASA." https://www.nasa.gov/people-of-nasa/women-at-nasa/rocket-fuel-in-her-blood-the-story-of-joann-morgan/.

NASA. 2021. *NASA Model Equal Employment Opportunity Program Status Report: Fy 2021*.

Neumark, David, Roy J. Bank, and Kyle D. Van Nort. 1996. "Sex Discrimination in Restaurant Hiring: An Audit Study*." *The Quarterly Journal of Economics* 111(3):915–41. doi:10.2307/2946676.

Ng, Thomas WH, Lillian T. Eby, Kelly L. Sorensen, and Daniel C. Feldman. 2005. "Predictors of Objective and Subjective Career Success: A Meta-analysis." *Personnel Psychology* 58(2):367–408.

Nie, Dan, Anna-Maija Lämsä, and Raminta Pučėtaitė. 2018. "Effects of Responsible Human Resource Management Practices on Female Employees' Turnover Intentions." *Business Ethics: A European Review* 27(1):29–41. doi:10.1111/beer.12165.

Nishii, L. H. (2013). The benefits of climate for inclusion for gender-diverse groups. *Academy of Management Journal, 56*(6), 1754–1774.

O'Connor, Lindsey Trimble. 2013. "Ask and You Shall Receive: Social Network Contacts' Provision of Help during the Job Search." *Social Networks* 35(4):593–603.

OECD. 2014. "PISA 2012 Technical Report."

OECD. 2022. *Enhancing Gender Diversity on Boards and in Senior Management of Listed Companies*. Vol. 28. *OECD Corporate Governance Working Papers*. 28. OECD Corporate Governance Working Papers. doi:10.1787/4f7ca695-en.

Orpen, Christopher. 1996. "Dependency as a Moderator of the Effects of Networking Behavior on Managerial Career Success." *The Journal of Psychology* 130(3):245–48.

Ostroff, Cheri, and Leanne E. Atwater. 2003. "Does Whom You Work with Matter? Effects of Referent Group Gender and Age Composition on Managers' Compensation." *Journal of Applied Psychology* 88(4):725.

Pairolero, Nicholas A., Andrew A. Toole, Peter-Anthony Pappas, Charles A. W. de Grazia, and Mike H. M. Teodorescu. 2025. "Closing the Gender Gap in Patenting: Evidence from a Randomized Control Trial at the USPTO." *American Economic Journal: Economic Policy*. doi:10.1257/pol.20230253.

Paustian-Underdahl, Samantha C., Lisa Slattery Walker, and David J. Woehr. 2014. "Gender and Perceptions of Leadership Effectiveness: A Meta-Analysis of Contextual Moderators." *Journal of Applied Psychology* 99(6):1129–45. doi:10.1037/a0036751.

Pedro Conceição, Jon Hall, Yu-Chieh Hsu, Admir Jahic, Milorad Kovacevic, Tanni Mukhopadhyay, Anna Ortubia, Carolina Rivera, and Heriberto Tapia. 2020. *Tackling Social Norms: A Game Changer for Gender Inequalities*. 2020 Human Development Perspectives. New York: United Nations Development Programme.

Pedulla, David S. 2016. "Penalized or Protected? Gender and the Consequences of Nonstandard and Mismatched Employment Histories." *American Sociological Review* 81(2):262–89. doi:10.1177/0003122416630982.

Pedulla, David S. 2018. "How Race and Unemployment Shape Labor Market Opportunities: Additive, Amplified, or Muted Effects?" *Social Forces* 96(4):1477–1506. doi:10.1093/sf/soy002.

Pelletier, Rachelle, Martha Patterson, and Melissa Moyser. 2019. "The Gender Wage Gap in Canada: 1998 to 2018."

Perry-Jenkins, Maureen, and Naomi Gerstel. 2020. "Work and Family in the Second Decade of the 21st Century." *Journal of Marriage and Family* 82(1):420–53. doi:10.1111/jomf.12636.

Petersen, Trond, and Ishak Saporta. 2004. "The Opportunity Structure for Discrimination." *American Journal of Sociology* 109(4):852–901. doi:10.1086/378536.

Pfeffer, Jeffrey. 1991. "Organization Theory and Structural Perspectives on Management." *Journal of Management* 17(4):789–803.

Pini, Barbara, Kerry Brown, and Chris Ryan. 2004. "Women-only Networks as a Strategy for Change? A Case Study from Local Government." *Women in Management Review* 19(6):286–92. doi:10.1108/09649420410555051.

Pistilli, Luca, Alessia Paccagnini, Stefano Breschi, and Franco Malerba. 2023. "Gender Bias in Entrepreneurship: What Is the Role of the Founders' Entrepreneurial Background?" *Journal of Business Ethics* 187(2):325–46. doi:10.1007/s10551-022-05275-y.

Post, Corinne, and Kris Byron. 2015. "Women on Boards and Firm Financial Performance: A Meta-Analysis." *Academy of Management Journal* 58(5):1546–71. doi:10.5465/amj.2013.0319.

Pradhan, Rabindra Kumar, and Lalatendu Kesari Jena. 2017. "Employee Performance at Workplace: Conceptual Model and Empirical Validation." *Business Perspectives and Research* 5(1):69–85. doi:10.1177/2278533716671630.

Preston, McKenzie C. 2025. "The Moral Case Revisited: Moral Framing as a Double-Edged Sword for Motivating Majority Group Leaders to Support DEI." *Academy of Management Journal* amj.2023.1421. doi:10.5465/amj.2023.1421.

R Core Team. 2024. "R: A Language and Environment for Statistical Computing,."

Reagans, Ray, and Bill McEvily. 2003. "Network Structure and Knowledge Transfer: The Effects of Cohesion and Range." *Administrative Science Quarterly* 48(2):240–67. doi:10.2307/3556658.

Reagans, Ray, Ezra Zuckerman, and Bill McEvily. 2004. "How to Make the Team: Social Networks vs. Demography as Criteria for Designing Effective Teams." *Administrative Science Quarterly* 49(1):101–33.

Reynolds, Sade'. 2021. "Perceptions of NASA as a Gendered Organization." *Walden Dissertations and Doctoral Studies*. https://scholarworks.waldenu.edu/dissertations/10316.

Ridgeway, Cecilia L. 1991. "The Social Construction of Status Value: Gender and Other Nominal Characteristics." *Social Forces* 70(2):367–86.

Ridgeway, Cecilia L. 2001. "Gender, Status, and Leadership." *Journal of Social Issues* 57(4):637–55.

Ridgeway, Cecilia L. 2014. "Why Status Matters for Inequality." *American Sociological Review* 79(1):1–16. doi:10.1177/0003122413515997.

Ridgeway, Cecilia L., Cathryn Johnson, and David Diekema. 1993. "External Status, Legitimacy and Compliance in Male and Female Groups Gender." *Social Forces* 72(4):1051–78.

Rivera, Lauren A., and András Tilcsik. 2016. "Class Advantage, Commitment Penalty: The Gendered Effect of Social Class Signals in an Elite Labor Market." *American Sociological Review* 81(6):1097–1131. doi:10.1177/0003122416668154.

Rivera, Lauren A., and András Tilcsik. 2019. "Scaling down Inequality: Rating Scales, Gender Bias, and the Architecture of Evaluation." *American Sociological Review* 84(2):248–74.

Rogelberg, Steven G., and Steven M. Rumery. 1996. "Gender Diversity, Team Decision Quality, Time on Task, and Interpersonal Cohesion." *Small Group Research* 27(1):79–90.

Rogers, William M. 2002. "Theoretical and Mathematical Constraints of Interactive Regression Models." *Organizational Research Methods* 5(3):212–30. doi:10.1177/1094428102005003002.

Ross, Catherine E., and John Mirowsky. 2006. "Sex Differences in the Effect of Education on Depression: Resource Multiplication or Resource Substitution?" *Social Science & Medicine* 63(5):1400–1413. doi:10.1016/j.socscimed.2006.03.013.

Rothman, Stanley, Seymour Martin Lipset, and Neil Nevitte. 2003. “Does Enrollment Diversity Improve University Education?” *International Journal of Public Opinion Research* 15(1):8–26.

Rubery, Jill, and Colette Fagan. 1993. *Occupational Segregation of Women and Men in the European Community*. Publications Office of the European Union.

Rudman, Laurie A., Corinne A. Moss-Racusin, Julie E. Phelan, and Sanne Nauts. 2012. “Status Incongruity and Backlash Effects: Defending the Gender Hierarchy Motivates Prejudice against Female Leaders.” *Journal of Experimental Social Psychology* 48(1):165–79. doi:10.1016/j.jesp.2011.10.008.

Sabharwal, Meghna. 2014. “Is Diversity Management Sufficient? Organizational Inclusion to Further Performance.” *Public Personnel Management* 43(2):197–217.

Saha, Anamik, and Sandra and van Lente. 2022. “Diversity, Media and Racial Capitalism: A Case Study on Publishing.” *Ethnic and Racial Studies* 45(16):216–36. doi:10.1080/01419870.2022.2032250.

Sampat, Bhaven N. 2005. *Determinants of Patent Quality: An Empirical Analysis*. New York: Columbia University.

Sampat, Bhaven, and Heidi L. Williams. 2019. “How Do Patents Affect Follow-On Innovation? Evidence from the Human Genome.” *American Economic Review* 109(1):203–36. doi:10.1257/aer.20151398.

Sarsons, Heather. 2017. “Interpreting Signals in the Labor Market: Evidence from Medical Referrals.”

Schoen, Constantin, Katja Rost, and David Seidl. 2018. “The Influence of Gender Ratios on Academic Careers: Combining Social Networks with Tokenism.” *PloS One* 13(11):e0207337.

Seron, Carroll, Susan S. Silbey, Erin Cech, and Brian Rubineau. 2016. “Persistence Is Cultural: Professional Socialization and the Reproduction of Sex Segregation.” *Work and Occupations* 43(2):178–214. doi:10.1177/0730888415618728.

Shore, L. M., Randel, A. E., Chung, B. G., Dean, M. A., Ehrhart, K. H., & Singh, G. (2011). Inclusion and diversity in work groups: A review and model for future research. *Journal of Management, 37*(4), 1262–1289.

Shu, Tao, Xuan Tian, and Xintong Zhan. 2021. “Patent Quality, Firm Value, and Investor Underreaction: Evidence from Patent Examiner Busyness.”

Simmel, Georg. 1950. *The Sociology of Georg Simmel*. Vol. 92892. Simon and Schuster.

Smith, David. 2019. “Without These Women, Man Would Not Have Walked on the Moon.” *The Guardian*, July 19.

Smith, Rachel A., Timothy R. Levine, Kenneth A. Lachlan, and Thomas A. Fediuk. 2002. “The High Cost of Complexity in Experimental Design and Data Analysis: Type I and Type II Error Rates in Multiway ANOVA.” *Human Communication Research* 28(4):515–30. doi:10.1111/j.1468-2958.2002.tb00821.x.

de Solla Price, Derek John. 1963. “Little Science, Big Science–and Beyond.”

Son, Joonmo, and Nan Lin. 2012. “Network Diversity, Contact Diversity, and Status Attainment.” *Social Networks* 34(4):601–13.

Spence, Michael. 1973. “Job Market Signaling.” *The Quarterly Journal of Economics* 87(3):355. doi:10.2307/1882010.

Stamarski, Cailin S., and Leanne S. Son Hing. 2015. “Gender Inequalities in the Workplace: The Effects of Organizational Structures, Processes, Practices, and Decision Makers’ Sexism.” *Frontiers in Psychology* 6:1400. doi:10.3389/fpsyg.2015.01400.

Stienstra, Kim, Ineke Maas, Antonie Knigge, and Wiebke Schulz. 2021. “Resource Compensation or Multiplication? The Interplay between Cognitive Ability and Social Origin in Explaining Educational Attainment.” *European Sociological Review* 37(2):186–200. doi:10.1093/esr/jcaa054.

Stokke, Hildegunn E. 2021. "The Gender Wage Gap and the Early-Career Effect: The Role of Actual Experience and Education Level." *LABOUR* 35(2):135–62. doi:10.1111/labr.12191.

Subramani, Gauri S. 2021. *Representation and the Direction of Innovation: Evidence from US Patent Applications*. University of California, Berkeley.

Sutton, John. 1997. "Gibrat's Legacy." *Journal of Economic Literature* 35(1):40–59.

Tedder-King, Alyssa, and Elad N. Sherf. 2024. "Fairness Judgments in the Context of Structural Sexism: The Role of Beliefs in Individual and Structural Causes of Success." *Academy of Management Journal* amj.2022.0776. doi:10.5465/amj.2022.0776.

Toczek, Lisa, Hans Bosma, and Richard Peter. 2021. "The Gender Pay Gap: Income Inequality Over Life Course – A Multilevel Analysis." *Frontiers in Sociology* 6. https://www.frontiersin.org/articles/10.3389/fsoc.2021.815376.

Trzebiatowski, Tiffany, Kaifeng Jiang, Zhen Zhang, Rory Eckardt, and Yeongsu Anthony Kim. 2025. "A Diversity Signal Set Perspective: Examining Interactive Effects of Diversity Practices on Women and Racialized Non-Leader and Leader Turnover." *Academy of Management Journal* 68(1):191–220. doi:10.5465/amj.2020.1838.

Tsui, Anne S., and Charles A. O'Reilly. 1989. "Beyond Simple Demographic Effects: The Importance of Relational Demography in Superior-Subordinate Dyads." *The Academy of Management Journal* 32(2):402–23. doi:10.2307/256368.

UN Women. 2018. *Turning Promises into Action: Gender Equality in the 2030 Agenda for Sustainable Development*. UN Women.

United Nations,. 2001. *Road Map towards the Implementation of the United Nations Millennium Declaration. Report of the Secretary-General*. Item 40 of the provisional agenda. Fifty-Sixth Session. United Nations. https://www.un.org/millenniumgoals/sgreport2001.pdf?OpenElement.

United Nations,. 2022. *The Sustainable Development Goals Report 2022*. United Nations: United Nations publication issued by the Department of Economic and Social Affairs (DESA).

University of Essex, ;, ; Institute for Social and Economic Research, ; NatCen Social Research, and Kantar Public. 2019. *Understanding Society: Waves 1-9, 2009-2018 and Harmonised BHPS: Waves 1-18, 1991-2009*.

USPTO. 2020. *Progress and Potential: 2020 Update on U.S. Women Inventor-Patentees. IP DATA HIGHLIGHTS*. 4. U.S. Patent and Trademark Office. Office of the Chief Economist. https://www.uspto.gov/sites/default/files/documents/OCE-DH-Progress-Potential-2020.pdf.

Van Buuren, Stef. 2018. *Flexible Imputation of Missing Data*. CRC press.

Van der Lippe, Tanja, J. Lössbroek, A. van der Put, N. Vergeldt, J. Slabbekoorn, and T. Martens. 2018. *European Sustainable Workforce Survey [ESWS]. Second Wave.* Utrecht.

Van Knippenberg, Daan, Carsten KW De Dreu, and Astrid C. Homan. 2004. "Work Group Diversity and Group Performance: An Integrative Model and Research Agenda." *Journal of Applied Psychology* 89(6):1008.

Van Knippenberg, Daan, and Michaela C. Schippers. 2007. "Https://Psycnet.Apa.Org/Record/2004-21169-009." *Annual Review of Psychology* 58.

Vedres, Balazs, and Orsolya Vasarhelyi. 2022. *Inclusion Unlocks the Creative Potential of Gender Diversity in Teams. preprint*. SocArXiv. doi:10.31235/osf.io/a3wf9.

Vedres, Balázs, and Orsolya Vásárhelyi. 2023. "Inclusion Unlocks the Creative Potential of Gender Diversity in Teams." *Scientific Reports* 13(1):13757. doi:10.1038/s41598-023-39922-9.

Veit, Susanne, and Lex Thijsen. 2021. "Almost Identical but Still Treated Differently: Hiring Discrimination against Foreign-Born and Domestic-Born Minorities." *Journal of Ethnic and Migration Studies* 47(6):1285–1304. doi:10.1080/1369183X.2019.1622825.

Vink, Gerko, Goran Lazendic, and Stef van Buuren. 2015. "Partioned Predictive Mean Matching as a Large Data Multilevel Imputation Technique." *Psychological Test and Assessment Modeling* 57(4):577–94.

Von Hippel, Paul T. 2009. "8. How to Impute Interactions, Squares, and Other Transformed Variables." *Sociological Methodology* 39(1):265–91.

Waldman, David Andrew, and Jennifer L. Sparr. 2022. "Rethinking Diversity Strategies: An Application of Paradox and Positive Organization Behavior Theories." *Academy of Management Perspectives*. doi:10.5465/amp.2021.0183.

Weber, Lauren. 2018. "White Men Challenge Workplace Diversity Efforts." *Wall Street Journal*, March 14.

WEF. 2023. *Global Gender Gap Report 2023*. Geneva: World Economic Forum. https://www.weforum.org/publications/global-gender-gap-report-2023/.

White, Ian R., Patrick Royston, and Angela M. Wood. 2011. "Multiple Imputation Using Chained Equations: Issues and Guidance for Practice." *Statistics in Medicine* 30(4):377–99.

Whitely, William, Thomas W. Dougherty, and George F. Dreher. 1991. "Relationship of Career Mentoring and Socioeconomic Origin to Managers' and Professionals' Early Career Progress." *Academy of Management Journal* 34(2):331–50.

Whittington, Kjersten Bunker, and Laurel Smith-Doerr. 2005. "Gender and Commercial Science: Women's Patenting in the Life Sciences." *The Journal of Technology Transfer* 30(4):355–70. doi:10.1007/s10961-005-2581-5.

Williams, Christine L., Chandra Muller, and Kristine Kilanski. 2012. "Gendered Organizations in the New Economy." *Gender & Society* 26(4):549–73. doi:10.1177/0891243212445466.

Williams, Helen M., and Lindsey J. Meân. 2004. "Measuring Gender Composition in Work Groups: A Comparison of Existing Methods." *Organizational Research Methods* 7(4):456–74. doi:10.1177/1094428104269175.

Williams, Jamillah Bowman. 2022. "Beyond the Business Case: Moving from Transactional to Transformational Inclusion." *Seattle University Law Review* 46:299.

Williams, Katherine Y., and Charles A. O'Reilly III. 1998. "Demography and Diversity In Organizations: A Review of 40 Years of Research." *Research in Organizational Behavior* 20:77–140.

Wilson, Edward O. 2000. *Sociobiology: The New Synthesis, Twenty-Fifth Anniversary Edition*. Harvard University Press.

Woehler, Meredith L., Kristin L. Cullen-Lester, Caitlin M. Porter, and Katherine A. Frear. 2020. "Whether, How, and Why Networks Influence Men's and Women's Career Success: Review and Research Agenda." *Journal of Management* 0149206320960529.

Wolff, Hans-Georg, and Klaus Moser. 2009. "Effects of Networking on Career Success: A Longitudinal Study." *Journal of Applied Psychology* 94(1):196–206. doi:10.1037/a0013350.

Wuchty, Stefan, Benjamin F. Jones, and Brian Uzzi. 2007. "The Increasing Dominance of Teams in Production of Knowledge." *Science* 316(5827):1036–39. doi:10.1126/science.1136099.

Yzerbyt, Vincent Y., Dominique Muller, and Charles M. Judd. 2004. "Adjusting Researchers' Approach to Adjustment: On the Use of Covariates When Testing Interactions." *Journal of Experimental Social Psychology* 40(3):424–31. doi:10.1016/j.jesp.2003.10.001.

ACKNOWLEDGE -MENTS

This dissertation marks the end of a long and formative journey, shaped by the generosity, care, and support of many. I am deeply grateful to those who supported me academically, personally, and emotionally throughout these years. What follows is an attempt to acknowledge the many individuals and communities who made this PhD possible and meaningful. I say attempt as it is impossible to truly express my gratitude here to everyone who has been a part of this journey.

First and foremost, this dissertation would have been impossible without my Dutch *ouders*, my supervisors, Eva Jaspers and Tanja van der Lippe. Both of you went far beyond the role of supervisors, and I find it hard to put into words how grateful I am for your support (but here goes nothing). Thank you for always checking in on me, especially during the most isolating periods of the PhD, and for making sure I felt supported and connected. During Covid in particular, you took deliberate time to make sure I was not lonely, through walks, online pub quizzes, and regular check ins. I am deeply grateful for the way you encouraged me to pursue my (often wild and abstract) ideas while keeping me grounded, and for helping me see them through to completion. I cannot thank you enough for the care and trust you placed in me throughout this journey.

Eva, our many conversations were a steady source of support throughout this PhD. Whether we were talking through abstract ideas, trying to untangle half formed thoughts, or walking through the forest during Covid, you always made time for a chat. You had an uncanny ability to understand what I was trying to say even when I could not yet say it well, and you helped me learn how to communicate my ideas more clearly. I cherished the annual gazpacho evenings (despite never winning the funniest person award) and look forward to many more. Tanja, thank you for your generosity, care, and steady guidance. I will always remember how warmly you welcomed my mother during her visit and how you played the piano for us. Your calm presence and clear perspective helped me keep moving forward, even when things felt overwhelming. I feel lucky to have had the opportunity to learn so much from you. I can often hear your voice in my head when am writing (saying for example…, would your mum understand this…). I always admired your leadership style, engaging everyone around the table, making yourself approachable to everyone, especially over lunch. I do not know how you wear so many hats and with such ease.

I am also deeply thankful to the many academic mentors and collaborators who contributed to this dissertation. Vincent Buskens, while I got to know you over our many lovely lunch conversations (still can't believe I fell for you hiding my lunch), I could never have imagined how important your kindness would become during my PhD. What began as a small problem I casually mentioned over lunch turned into support far beyond what I could have imagined. Your kindness, generosity, and continued help in rebuilding and publishing the paper were critical to its completion, and I am immensely grateful.

Herbert Hoijtink, thank you for your patience, enthusiasm, and support in helping me navigate the many analytical questions I brought to you. Your guidance was critical in enabling me to answer my research questions, and you sparked a genuine love for BAIN, one strong enough that I ended up using it again in my second paper. Thank you for your encouragement and humility in helping me learn from my mistakes instead of losing too much time spiraling into them. Even when things were far from straightforward, you supported me with the same enthusiasm and kindness.

I would also like to thank Belle Derks and Paulina Pankowska for their time, feedback, and support. Although our collaboration was cut short, I really valued our brief time together. Belle, I always enjoyed running into you at SCOOP sessions, where your kind smile was always a welcome sight. Paulina, it was always lovely to see you in the office, and I really enjoyed our little walks and chats.

I am grateful to my committee members, Naomi Ellemers, Ellen Hamaker, Rebecca Kuiper, Ineke Maas, and Rafael Wittek, for the time and care you invested in evaluating my work. I feel fortunate to have had the opportunity to meet and learn from Rafael, Naomi, and Ineke in different capacities during my PhD. Rafael, thank you for your early input and encouragement, and for the thoughtful discussions at SCOOP events. Naomi, I am grateful for your generous advice on life after PhD and for sharing your perspective and experience in academia. Ineke, I truly enjoyed our lunch conversations and afternoon pantry run ins, and learned a great deal about tennis and Dutch culture from you. I will always appreciate how you so often picked us up for lunch, despite the long trek.

My PhD would not have been the same without the colleagues and peers who shared this journey with me. To my cohort and office mates, Anne, Kim, Ece, Emmanuel, Jan-Willem, Kasper, and Philipp, thank you for the many coffee breaks, treats, walks, and get togethers that sustained me. I loved our little office, with its Christmas lights, fancy tea setup, plants (that survived), and the iconic phoenix painting (or at least that is what I think it was). Thank you Kim, Ece and Jan-Willem, your presence and support made even the most challenging days lighter. I am glad I also had the chance to briefly share the office with Philipp and Amina, and later to be part of what was dubbed the most happening office in the department, surrounded by a dynamic yet wonderful group of people. Thank you, Gemma, for introducing me to the cutest farm animals on Uithof; Chloé, for organizing office outings and supplying us with crepes; Dominika, for your constant care and for always checking in on me; and Jos, for our neighborly walks (and conceding to the benefits of R over Python).

I am thankful to department colleagues past and present who shaped my daily life at work. Hendrik and Katerina, thank you for introducing me to German Christmas traditions in all its glory, and for the joy of baking cookies together, with the best little sous-chef one could ask for. I miss all of you every Christmas. Siyang, thank you for your support even before I started with the PhD and I am glad we have been able to stay in touch even after your move to Gouda. Our hotpot, meals, and walks always brighten up the Dutch winter for me. Aleida, there is far more to thank you for than I can say here. Your care and support have meant a great deal to me, and while I will miss our office conversations, I am hoping we will make up for it with more craft, chai and dinner time.

Marco, I am especially grateful for the chance to bond over our shared love of food, and for your continued support even after your retirement from Utrecht. Weverton, thank you for the warm lunch company and guidance. Huyen, thank you for the wonderful Vietnamese dinners and for making me fall in love with Lego all over again. Marjolein, I loved our Friday catch-ups and getting to witness the sorcery of your beautiful home expansion.

While I never fully adapted to the typical Dutch lunch and could manage at most two days of *kaasbroodje*, I will miss our department lunches. Starting with the much anticipated and personalized lunch calls, followed by gaming the microwave where the lines have only grown over time. Luckily the warm company makes the wait seamless, often with Isabelle, Miro, Jiamin, and Katia preparing their fancy lunches. Most of all I will miss the many lively, overlapping conversations with familiar faces drifting in and out over time, including Rense, Wojtek (my friendly neighbor, always willing to lend an ear), Thomas, Anne-Rigt, Rowan, Antonie, Mathijs, Joey, Ramona, Ana Macanovic, and Annelie. I will miss Amy making things international with her trusted Coke can, Disney discussions with Anne Brons, media binges with Tali, Maike's sailing adventures, sharing some Ghibli love with Joyce, and Paula's ice cream study updates.

There were also colleagues who were less frequent at the lunch table but no less supportive. Ellen, thank you for always looking out for me, especially during Covid. Sanne, thank you for always sharing snacks (yours and ours). I am also grateful to Marianne, Pim, and Sofie, for always being ready to help. Finally, I want to thank the entire department for the surprise citizenship celebration. Having given up my Indian citizenship, I felt mostly indifferent at first, and only truly processed the moment after the celebration. It made me feel incredibly lucky to be surrounded by such care, support, and love.

There were also many from the broader research community who enriched my PhD journey. As part of the ICS, Ashwin and Robert were among the first to help me realize just how wonderful the ICS community was. Alongside them were many others, including Lianne, Sofie, and Renae. I am especially grateful to Ashwin for regularly checking in on me throughout the PhD.

I also loved being part of the SCOOP community. The SCOOP Days and conferences were always a pleasure, made even more so by the many SCOOPies I got to know along the way, including Thomas, Hao, Damion, Julian, Marlou, Ewout, Anne, Piet, Nick, Esmee, Tara, Kevin, Larisa, Hendrik, and José Luis. None of this would have been possible without the many people working behind the scenes, especially Liesbet, Saskia, Geetha, and Daniela, who always went out of their way to support me.

Throughout my PhD, I was lucky to have friends outside of work who helped me get out of my head and remind me of the world beyond academia. Simone, Harm, Dignus, Jette, and Rhewein, thank you for letting me experience different worlds, timelines, creatures, characters and cuisines through our many RPG nights. I am very much looking forward to many more adventures together. Yvonne, Samhita, and Payal di, thank you for always being there for me despite the distance. Ram, thank you for your guidance and encouragement throughout this journey. Reeti, Zuno, and Edo, thank you for your friendship and warmth, I could really use it with the cold Dutch winters. Pheona, you taught me so much more than pottery, and I feel lucky to have you in my life. Ravi, thank you for being the best neighbor I could have asked for. I only wish we could have had more biryani nights and tested Utrecht's burgeoning food scene together.

Finally, I turn to my family, whose love and support carried me through this journey. I am deeply grateful to my parents for their constant support, and to my sister, Srishti, not only for her constant encouragement but also for patiently designing multiple versions of the cover page. I thank my nana, nani, mamas, and mamis, whose presence and love has always given me confidence. I will deeply miss not having my nani here to see me graduate. Benny, ATN, Zaadya and Kaddu, thank you for making (trying) CTP a reality in Utrecht. Shruthi, Pranav and Ashwini, you have long been family to me, and I cannot thank you enough for your unconditional love and friendship. Pranav and Ashwini, the fact that you were living under a bridge with a flamethrower certainly helped, and I am glad Aditya joined the mix to help keep things toasty. To all of you, and to everyone I may have unintentionally missed, thank you for making this journey possible.

ABOUT THE AUTHOR

Sanjana Singh was born in Chennai, India, on 20 March 1993. She obtained her bachelor's degree in Economics with distinction from Stella Maris College (Autonomous), Madras University, in 2013, during which she also interned with Goldman Sachs. In 2014, she completed her MSc in Economic Policy at University College London, after which she gained applied research experience as a social development Consultant at Athena Infonomics. She worked on projects related to financial inclusion, gender equality, health, and social norms for bilateral and multilateral clients, including the United Nations, the International Finance Corporation, and Oxfam.

In 2018, she started working as a Ph.D. candidate at the Interuniversity Centre for Social Science Theory and Methodology (ICS), the transdisciplinary SCOOP research program on sustainable cooperation, and the Department of Sociology at Utrecht University. She wrote her dissertation under the supervision of Prof. dr. Tanja van der Lippe (Department of Sociology, Utrecht University), and Prof. dr. Eva Jaspers (Department of Sociology, Utrecht University). During her PhD, she also taught courses in social networks and supervised student research. Her research focuses on gender, inequality, social networks, diversity management, and organizational behavior.

LIST OF ICS DISSERTATION SERIES

ICS Dissertation Series

The ICS series presents dissertations of the Interuniversity Center for Social Science Theory and Methodology. Each of these studies aims at integrating explicit theory formation with state-of-the-art empirical research or at the development of advanced methods for empirical research. The ICS was founded in 1986 as a cooperative effort of the universities of Groningen and Utrecht. Since 1992, the ICS has expanded to the University of Nijmegen and, since 2017, to the University of Amsterdam. Most of the projects are financed by the participating universities or by the Dutch Research Council. The international composition of the ICS graduate students is mirrored in the increasing international orientation of the projects and thus of the ICS series itself.

1. Kees van Liere (1990), *"Lastige leerlingen: Een empirisch onderzoek naar sociale oorzaken van probleemgedrag op basisscholen."* Amsterdam: Thesis Publishers
2. Marco van Leeuwen (1990), *"Bijstand in Amsterdam, ca. 1800-1850: Armenzorg als beheersings en overlevingsstrategie."* ICS dissertation, Utrecht
3. Ineke Maas (1990), *"Deelname aan podiumkunsten via de podia, de media en actieve beoefening: Substitutie of leereffecten?"* Amsterdam: Thesis Publishers
4. Marjolein Broese van Groenou (1991), *"Gescheiden netwerken: De relaties met vrienden en verwanten na echtscheiding."* Amsterdam: Thesis Publishers
5. Jan van den Bos (1991), *"Dutch EC policy making: A model guided approach to coordination and negotiation."* Amsterdam: Thesis Publishers
6. Karin Sanders (1991), *"Vrouwelijke pioniers: Vrouwen en mannen met een 'mannelijke' hogere beroepsopleiding aan het begin van hun loopbaan."* Amsterdam: Thesis Publishers
7. Sjerp de Vries (1991), *"Egoism, altruism, and social justice: Theory and experiments on cooperation in social dilemmas."* Amsterdam: Thesis Publishers
8. Ronald Batenburg (1991), *"Automatisering in bedrijf."* Amsterdam: Thesis Publishers
9. Rudi Wielers (1991), *"Selectie en allocatie op de arbeidsmarkt. Een uitwerking voor de informele en geïnstitutionaliseerde kinderopvang."* Amsterdam: Thesis Publishers
10. Gert Westert (1991), *"Verschillen in ziekenhuisgebruik."* ICS dissertation, Groningen
11. Hanneke Hermsen (1992), *"Votes and policy preferences: Equilibria in party systems."* Amsterdam: Thesis Publishers
12. Cora Maas (1992), *"Probleemleerlingen in het basisonderwijs."* Amsterdam: Thesis Publishers
13. Ed Boxman (1992), *"Contacten en carrière : Een empirisch theoretisch onderzoek naar de relatie tussen sociale netwerken en arbeidsmarktposities"* Amsterdam: Thesis Publishers
14. Conny Taes (1992), *"Kijken naar banen: Een onderzoek naar de inschatting van arbeidsmarktkansen bij schoolverlaters uit het middelbaar beroepsonderwijs."* Amsterdam: Thesis Publishers
15. Peter van Roozendaal (1992), *"Cabinets in multi party democracies: The effect of dominant and central parties on cabinet composition and durability."* Amsterdam: Thesis Publishers
16. Marcel van Dam (1992), *"Regio zonder regie: Verschillen in en effectiviteit van gemeentelijk arbeidsmarktbeleid."* Amsterdam: Thesis Publishers
17. Tanja van der Lippe (1993), *"Arbeidsverdeling tussen mannen en vrouwen."* Amsterdam: Thesis Publishers
18. Marc Jacobs (1993), *"Software: Kopen of kopiëren? Een sociaal wetenschappelijk onderzoek onder PC gebruikers."* Amsterdam: Thesis Publishers
19. Peter van der Meer (1993), *"Verdringing op de Nederlandse arbeidsmarkt: Sector- en sekseverschillen."* Amsterdam: Thesis Publishers
20. Gerbert Kraaykamp (1993), *"Over lezen gesproken: Een studie naar sociale differentiatie in leesgedrag."* Amsterdam: Thesis Publishers
21. Evelien Zeggelink (1993), *"Strangers into friends: The evolution of friendship networks using an individual oriented modeling approach."* Amsterdam: Thesis Publishers
22. Jaco Berveling (1994), *"Het stempel op de besluitvorming: Macht, invloed en besluitvorming op twee Amsterdamse beleidsterreinen."* Amsterdam: Thesis Publishers
23. Wim Bernasco (1994), *"Coupled careers: The effects of spouse's resources on success at work."* Amsterdam: Thesis Publishers

24. Liset van Dijk (1994), *"Choices in child care: The distribution of child care among mothers, fathers and non parental care providers."* Amsterdam: Thesis Publishers
25. Jos de Haan (1994), *"Research groups in Dutch sociology."* Amsterdam: Thesis Publishers
26. Kwasi Boahene (1995), *"Innovation adoption as a socioeconomic process: The case of the Ghanaian cocoa industry."* Amsterdam: Thesis Publishers
27. Paul Ligthart (1995), *"Solidarity in economic transactions: An experimental study of framing effects in bargaining and contracting."* Amsterdam: Thesis Publishers
28. Roger Leenders (1995), *"Structure and influence: Statistical models for the dynamics of actor attributes, network structure, and their interdependence."* Amsterdam: Thesis Publishers
29. Beate Volker (1995), *"Should auld acquaintance be forgot...? Institutions of communism, the transition to capitalism and personal networks: The case of East Germany."* Amsterdam: Thesis Publishers
30. Anneke Cancrinus-Matthijsse (1995), *"Tussen hulpverlening en ondernemerschap: Beroepsuitoefening en taakopvattingen van openbare apothekers in een aantal West-Europese landen."* Amsterdam: Thesis Publishers
31. Nardi Steverink (1996), *"Zo lang mogelijk zelfstandig: Naar een verklaring van verschillen in oriëntatie ten aanzien van opname in een verzorgingstehuis onder fysiek kwetsbare ouderen."* Amsterdam: Thesis Publishers
32. Ellen Lindeman (1996), *"Participatie in vrijwilligerswerk."* Amsterdam: Thesis Publishers
33. Chris Snijders (1996), *"Trust and commitments."* Amsterdam: Thesis Publishers
34. Koos Postma (1996), *"Changing prejudice in Hungary. A study on the collapse of state socialism and its impact on prejudice against Gypsies and Jews."* Amsterdam: Thesis Publishers
35. Jooske van Busschbach (1996), *"Uit het oog, uit het hart? Stabiliteit en verandering in persoonlijke relaties."* Amsterdam: Thesis Publishers
36. René Torenvlied (1996), *"Besluiten in uitvoering: Theorieën over beleidsuitvoering modelmatig getoetst op sociale vernieuwing in drie gemeenten."* Amsterdam: Thesis Publishers
37. Andreas Flache (1996), *"The double edge of networks: An analysis of the effect of informal networks on cooperation in social dilemmas."* Amsterdam: Thesis Publishers
38. Kees van Veen (1997), *"Inside an internal labor market: Formal rules, flexibility and career lines in a Dutch manufacturing company."* Amsterdam: Thesis Publishers
39. Lucienne van Eijk (1997), *"Activity and wellbeing in the elderly."* Amsterdam: Thesis Publishers
40. Róbert Gál (1997), *"Unreliability: Contract discipline and contract governance under economic transition."* Amsterdam: Thesis Publishers
41. Anne-Geerte van de Goor (1997), *"Effects of regulation on disability duration."* ICS dissertation, Utrecht
42. Boris Blumberg (1997), *"Das Management von Technologiekooperationen: Partnersuche und Verhandlungen mit dem Partner aus Empirisch Theoretischer Perspektive."* ICS dissertation, Utrecht
43. Marijke von Bergh (1997), *"Loopbanen van oudere werknemers."* Amsterdam: Thesis Publishers
44. Anna Petra Nieboer (1997), *"Life events and well-being: A prospective study on changes in well-being of elderly people due to a serious illness event or death of the spouse."* Amsterdam: Thesis Publishers
45. Jacques Niehof (1997), *"Resources and social reproduction: The effects of cultural and material resources on educational and occupational careers in industrial nations at the end of the twentieth century."* ICS dissertation, Nijmegen
46. Ariana Need (1997), *"The kindred vote: Individual and family effects of social class and religion on electoral change in the Netherlands, 1956 1994."* ICS dissertation, Nijmegen
47. Jim Allen (1997), *"Sector composition and the effect of education on Wages: An international comparison."* Amsterdam: Thesis Publishers
48. Jack Hutten (1998), *"Workload and provision of care in general practice: An empirical study of the relation between workload of Dutch general practitioners and the content and quality of their care."* ICS dissertation, Utrecht
49. Per Kropp (1998), *"Berufserfolg im Transformationsprozeß: Eine theoretisch empirische Studie über die Gewinner und Verlierer der Wende in Ostdeutschland."* ICS dissertation, Utrecht
50. Maarten Wolbers (1998), *"Diploma-inflatie en verdringing op de arbeidsmarkt: Een studie naar ontwikkelingen in de opbrengsten van diploma's in Nederland."* ICS dissertation, Nijmegen
51. Wilma Smeenk (1998), *"Opportunity and marriage: The impact of individual resources and marriage market structure on first marriage timing and partner choice in the Netherlands."* ICS dissertation, Nijmegen

52. Marinus Spreen (1999), *"Sampling personal network structures: Statistical inference in ego graphs."* ICS dissertation, Groningen
53. Vincent Buskens (1999), *"Social networks and trust."* ICS dissertation, Utrecht
54. Susanne Rijken (1999), *"Educational expansion and status attainment: A cross-national and over-time comparison."* ICS dissertation, Utrecht
55. Mérove Gijsberts (1999), *"The legitimation of inequality in state-socialist and market societies, 1987-1996."* ICS dissertation, Utrecht
56. Gerhard van de Bunt (1999), *"Friends by choice: An actor-oriented statistical network model for friendship networks through time."* ICS dissertation, Groningen
57. Robert Thomson (1999), *"The party mandate: Election pledges and government actions in the Netherlands, 1986 1998."* Amsterdam: Thela Thesis
58. Corine Baarda (1999), *"Politieke besluiten en boeren beslissingen: Het draagvlak van het mestbeleid tot 2000."* ICS dissertation, Groningen
59. Rafael Wittek (1999), *"Interdependence and informal control in organizations."* ICS dissertation, Groningen
60. Diane Payne (1999), *"Policy Making in the European Union: An analysis of the impact of the reform of the structural funds in Ireland."* ICS dissertation, Groningen
61. René Veenstra (1999), *"Leerlingen, Klassen, Scholen: Prestaties en vorderingen van leerlingen in het voortgezet onderwijs."* Amsterdam, Thela Thesis
62. Marjolein Achterkamp (1999), *"Influence strategies in collective decision making: A comparison of two models."* ICS dissertation, Groningen
63. Peter Mühlau (2000), *"The governance of the employment relation: Arelational signaling perspective."* ICS dissertation, Groningen
64. Agnes Akkerman (2000), *"Verdeelde vakbeweging en stakingen: Concurrentie om leden."* ICS dissertation, Groningen
65. Sandra van Thiel (2000), *"Quangocratization: Trends, causes and consequences."* ICS dissertation, Utrecht
66. Sylvia Peacock-Korupp (2000), *"Mothers and the process of social stratification."* ICS dissertation, Utrecht
67. Rudi Turksema (2000), *"Supply of day care."* ICS dissertation, Utrecht
68. Bernard Nijstad (2000), *"How the group affects the mind: Effects of communication in idea generating groups."* ICS dissertation, Utrecht
69. Inge de Wolf (2000), *"Opleidingsspecialisatie en arbeidsmarktsucces van sociale wetenschappers."* ICS dissertation, Utrecht
70. Jan Kratzer (2001), *"Communication and performance: An empirical study in innovation teams."* ICS dissertation, Groningen
71. Madelon Kroneman (2001), *"Healthcare systems and hospital bed use."* ICS/NIVEL-dissertation, Utrecht
72. Herman van de Werfhorst (2001), *"Field of study and social inequality: Four types of educational resources in the process of stratification in the Netherlands."* ICS dissertation, Nijmegen
73. Tamás Bartus (2001), *"Social capital and earnings inequalities: The role of informal job search in Hungary."* ICS dissertation Groningen
74. Hester Moerbeek (2001), *"Friends and foes in the occupational career: The influence of sweet and sour social capital on the labour market."* ICS dissertation, Nijmegen
75. Marcel van Assen (2001), *"Essays on actor perspectives in exchange networks and social dilemmas."* ICS dissertation, Groningen
76. Inge Sieben (2001), *"Sibling similarities and social stratification: The impact of family background across countries and cohorts."* ICS dissertation, Nijmegen
77. Alinda van Bruggen (2001), *"Individual production of social well-being: An exploratory study."* ICS dissertation, Groningen
78. Marcel Coenders (2001), *"Nationalistic attitudes and ethnic exclusionism in a comparative perspective: An empirical study of attitudes toward the country and ethnic immigrants in 22 countries."* ICS dissertation, Nijmegen
79. Marcel Lubbers (2001), *"Exclusionistic electorates: Extreme right-wing voting in Western Europe."* ICS dissertation, Nijmegen
80. Uwe Matzat (2001), *"Social networks and cooperation in electronic communities: A theoretical-empirical analysis of academic communication and internet discussion groups."* ICS dissertation, Groningen

81. Jacques Janssen (2002), *"Do opposites attract divorce? Dimensions of mixed marriage and the risk of divorce in the Netherlands."* ICS dissertation, Nijmegen
82. Miranda Jansen (2002), *"Waardenoriëntaties en partnerrelaties: Een panelstudie naar wederzijdse invloeden."* ICS dissertation, Utrecht
83. Anne-Rigt Poortman (2002), *"Socioeconomic causes and consequences of divorce."* ICS dissertation, Utrecht
84. Alexander Gattig (2002), *"Intertemporal decision making."* ICS dissertation, Groningen
85. Gerrit Rooks (2000), *"Contract en conflict: Strategisch management van inkooptransacties."* ICS dissertation, Utrecht
86. Károly Takács (2002), *"Social networks and intergroup conflict."* ICS dissertation, Groningen
87. Thomas Gautschi (2002), *"Trust and exchange, effects of temporal embeddedness and network embeddedness on providing and dividing a surplus."* ICS dissertation, Utrecht
88. Hilde Bras (2002), *"Zeeuwse meiden: Dienen in de levensloop van vrouwen, ca. 1850 – 1950."* Aksant Academic Publishers, Amsterdam
89. Merijn Rengers (2002), *"Economic lives of artists: Studies into careers and the labour market in the cultural sector."* ICS dissertation, Utrecht
90. Annelies Kassenberg (2002), *"Wat scholieren bindt: Sociale gemeenschap in scholen."* ICS dissertation, Groningen
91. Marc Verboord (2003), *"Moet de meester dalen of de leerling klimmen? De invloed van literatuuronderwijs en ouders op het lezen van boeken tussen 1975 en 2000."* ICS dissertation, Utrecht
92. Marcel van Egmond (2003), *"Rain falls on all of us (but some manage to get more wet than others): Political context and electoral participation."* ICS dissertation, Nijmegen
93. Justine Horgan (2003), *"High-performance human resource management in Ireland and the Netherlands: Adoption and effectiveness."* ICS dissertation, Groningen
94. Corine Hoeben (2003), *"LETS' be a community: Community in local exchange trading systems."* ICS dissertation, Groningen
95. Christian Steglich (2003), *"The framing of decision situations: Automatic goal selection and rational goal pursuit."* ICS dissertation, Groningen
96. Johan van Wilsem (2003), *"Crime and context: The impact of individual, neighborhood, city and country characteristics on victimization."* ICS dissertation, Nijmegen
97. Christiaan Monden (2003), *"Education, inequality and health: The impact of partners and life course."* ICS dissertation, Nijmegen
98. Evelyn Hello (2003), *"Educational attainment and ethnic attitudes: How to explain their relationship."* ICS dissertation, Nijmegen
99. Marnix Croes en Peter Tammes (2004). *"Gif laten wij niet voortbestaan: Een onderzoek naar de overlevingskansen van Joden in de Nederlandse gemeenten, 1940-1945."* Aksant Academic Publishers, Amsterdam.
100. Ineke Nagel (2004), *"Cultuurdeelname in de levensloop."* ICS dissertation, Utrecht
101. Marieke van der Wal (2004), *"Competencies to participate in life: Measurement and the impact of school."* ICS dissertation, Groningen
102. Vivian Meertens (2004), *"Depressive symptoms in the general population: A multifactorial social approach."* ICS dissertation, Nijmegen
103. Hanneke Schuurmans (2004), *"Promoting well-being in frail elderly people: Theory and intervention."* ICS dissertation, Groningen
104. Javier Arregui Moreno (2004), *"Negotiation in legislative decision-making in the European Union."* ICS dissertation, Groningen
105. Tamar Fischer (2004), *"Parental divorce, conflict and resources: The effects on children's behaviour problems, socioeconomic attainment, and transitions in the demographic career."* ICS dissertation, Nijmegen
106. René Bekkers (2004), *"Giving and volunteering in the Netherlands: Sociological and psychological perspectives."* ICS dissertation, Utrecht
107. Renée van der Hulst (2004), *"Gender differences in workplace authority: An empirical study on social networks."* ICS dissertation, Groningen

108. Rita Smaniotto (2004), *"'You scratch my back and I scratch yours' versus 'love thy neighbour': Two proximate mechanisms of reciprocal altruism."* ICS dissertation, Groningen
109. Maurice Gesthuizen (2004), *"The life course of the low-educated in the Netherlands: Social and economic risks."* ICS dissertation, Nijmegen
110. Carlijne Philips (2005), *"Vakantiegemeenschappen: Kwalitatief en kwantitatief onderzoek naar gelegenheid en refreshergemeenschap tijdens de vakantie."* ICS dissertation, Groningen
111. Esther de Ruijter (2005), *"Household outsourcing."* ICS dissertation, Utrecht
112. Frank van Tubergen (2005), *"The integration of immigrants in cross-national perspective: Origin, destination, and community effects."* ICS dissertation, Utrecht
113. Ferry Koster (2005), *"For the time being: Accounting for inconclusive findings concerning the effects of temporary employment relationships on solidary behavior of employees."* ICS dissertation, Groningen
114. Carolien Klein Haarhuis (2005), *"Promoting anti-corruption reforms: Evaluating the implementation of a World Bank anti-corruption program in seven African countries (1999-2001)."* ICS dissertation, Utrecht
115. Martin van der Gaag (2005), *"Measurement of individual social capital."* ICS dissertation, Groningen
116. Johan Hansen (2005), *"Shaping careers of men and women in organizational contexts."* ICS dissertation, Utrecht
117. Davide Barrera (2005), *"Trust in embedded settings."* ICS dissertation, Utrecht
118. Mattijs Lambooij (2005), *"Promoting cooperation: Studies into the effects of long-term and short-term rewards on cooperation of employees."* ICS dissertation, Utrecht
119. Lotte Vermeij (2006), *"What's cooking? Cultural boundaries among Dutch teenagers of different ethnic origins in the context of school."* ICS dissertation, Utrecht
120. Mathilde Strating (2006), *"Facing the challenge of rheumatoid arthritis: A 13-year prospective study among patients and cross-sectional study among their partners."* ICS dissertation, Groningen
121. Jannes de Vries (2006), *"Measurement error in family background variables: The bias in the intergenerational transmission of status, cultural consumption, party preference, and religiosity."* ICS dissertation, Nijmegen
122. Stefan Thau (2006), *"Workplace deviance: Four studies on employee motives and self-regulation."* ICS dissertation, Groningen
123. Mirjam Plantinga (2006), *"Employee motivation and employee performance in child care: The effects of the introduction of market forces on employees in the Dutch child-care sector."* ICS dissertation, Groningen
124. Helga de Valk (2006), *"Pathways into adulthood: A comparative study on family life transitions among migrant and Dutch youth."* ICS dissertation, Utrecht
125. Henrike Elzen (2006), *"Self-Management for chronically ill older people."* ICS dissertation, Groningen
126. Ayse Güveli (2007), *"New social classes within the service class in the Netherlands and Britain: Adjusting the EGP class schema for the technocrats and the social and cultural specialists."* ICS dissertation, Nijmegen
127. Willem-Jan Verhoeven (2007), *"Income attainment in post-communist societies."* ICS dissertation, Utrecht
128. Marieke Voorpostel (2007), *"Sibling support: The exchange of help among brothers and sisters in the Netherlands."* ICS dissertation, Utrecht
129. Jacob Dijkstra (2007), *"The effects of externalities on partner choice and payoffs in exchange networks."* ICS dissertation, Groningen
130. Patricia van Echtelt (2007), *"Time-greedy employment relationships: Four studies on the time claims of post-Fordist work."* ICS dissertation, Groningen
131. Sonja Vogt (2007), *"Heterogeneity in social dilemmas: The case of social support."* ICS dissertation, Utrecht
132. Michael Schweinberger (2007), *"Statistical methods for studying the evolution of networks and behavior."* ICS dissertation, Groningen
133. István Henrik Back (2007), *"Commitment and evolution: Connecting emotion and reason in long-term relationships."* ICS dissertation, Groningen
134. Ruben van Gaalen (2007), *"Solidarity and ambivalence in parent-child relationships."* ICS dissertation, Utrecht
135. Jan Reitsma (2007), *"Religiosity and solidarity: Dsimensions and relationships disentangled and tested."* ICS dissertation, Nijmegen
136. Jan Kornelis Dijkstra (2007), *"Status and affection among (pre)adolescents and their relation with antisocial and prosocial behavior."* ICS dissertation, Groningen
137. Wouter van Gils (2007), *"Full-time working couples in the Netherlands: Causes and consequences."* ICS dissertation, Nijmegen

138. Djamila Schans (2007), *"Ethnic diversity in intergenerational solidarity."* ICS dissertation, Utrecht
139. Ruud van der Meulen (2007), *"Brug over woelig water: Lidmaatschap van sportverenigingen, vriendschappen, kennissenkringen en veralgemeend vertrouwen."* ICS dissertation, Nijmegen
140. Andrea Knecht (2008), *"Friendship selection and friends' influence: Dynamics of networks and actor attributes in early adolescence."* ICS dissertation, Utrecht
141. Ingrid Doorten (2008), *"The division of unpaid work in the household: A stubborn pattern?"* ICS dissertation, Utrecht
142. Stijn Ruiter (2008), *"Association in context and association as context: Causes and consequences of voluntary association involvement."* ICS dissertation, Nijmegen
143. Janneke Joly (2008), *"People on our minds: When humanized contexts activate social norms."* ICS dissertation, Groningen
144. Margreet Frieling (2008), *"'Joint production' als motor voor actief burgerschap in de buurt."* ICS dissertation, Groningen
145. Ellen Verbakel (2008), *"The partner as resource or restriction? Labour market careers of husbands and wives and the consequences for inequality between couples."* ICS dissertation, Nijmegen
146. Gijs van Houten (2008), *"Beleidsuitvoering in gelaagde stelsels: De doorwerking van aanbevelingen van de Stichting van de Arbeid in het CAO-overleg."* ICS dissertation, Utrecht
147. Eva Jaspers (2008), *"Intolerance over time: Macro- and micro-level questions on attitudes towards euthanasia, homosexuality and ethnic minorities."* ICS dissertation, Nijmegen
148. Gijs Weijters (2008), *"Youth delinquency in Dutch cities and schools: A multilevel approach."* ICS dissertation, Nijmegen
149. Jessica Nooij (2009), *"The self in social rejection."* ICS dissertation, Groningen
150. Gerald Mollenhorst (2009), *"Networks in contexts: How meeting opportunities affect personal relationships."* ICS dissertation, Utrecht
151. Tom van der Meer (2009), *"States of freely associating citizens: Comparative studies into the impact of state institutions on social, civic and political participation."* ICS dissertation, Nijmegen
152. Manuela Vieth (2009), *"Commitments and reciprocity in trust situations: Experimental studies on obligation, indignation, and self-consistency."* ICS dissertation, Utrecht
153. Rense Corten (2009), *"Co-evolution of social networks and behavior in social dilemmas: Theoretical and empirical perspectives."* ICS dissertation, Utrecht
154. Arieke Rijken (2009), *"Happy families, high fertility? Childbearing choices in the context of family and partner relationships."* ICS dissertation, Utrecht
155. Jochem Tolsma (2009), *"Ethnic hostility among ethnic majority and minority groups in the Netherlands: An investigation into the impact of social mobility experiences, the local living environment and educational attainment on ethnic hostility."* ICS dissertation, Nijmegen
156. Freek Bucx (2009), *"Linked lives: Young adults' life course and relations with parents."* ICS dissertation, Utrecht
157. Philip Wotschack (2009), *"Household governance and time allocation: Four studies on the combination of work and care."* ICS dissertation, Groningen
158. Nienke Moor (2009), *"Explaining worldwide religious diversity: The relationship between subsistence technologies and ideas about the unknown in pre-industrial and (post-)industrial societies."* ICS dissertation, Nijmegen
159. Lieke ten Brummelhuis (2009), *"Family matters at work: Depleting and enriching effects of employees' family lives on work outcomes."* ICS dissertation, Utrecht
160. Renske Keizer (2010), *"Remaining childless: Causes and consequences from a life-course perspective."* ICS dissertation, Utrecht
161. Miranda Sentse (2010), *"Bridging contexts: The interplay between family, child, and peers in explaining problem behavior in early adolescence."* ICS dissertation, Groningen
162. Nicole Tieben (2010), *"Transitions, tracks and transformations: Social inequality in transitions into, through and out of secondary education in the Netherlands for cohorts born between 1914 and 1985."* ICS dissertation, Nijmegen

163. Birgit Pauksztat (2010), *"Speaking up in organizations: Four studies on employee voice."* ICS dissertation, Groningen
164. Richard Zijdeman (2010), *"Status attainment in the Netherlands, 1811-1941: Spatial and temporal variation before and during industrialization."* ICS dissertation, Utrecht
165. Rianne Kloosterman (2010), *"Social background and children's educational careers: The primary and secondary effects of social background over transitions and over time in the Netherlands."* ICS dissertation, Nijmegen
166. Olav Aarts (2010), *"Religious diversity and religious involvement: A study of religious markets in Western societies at the end of the twentieth century."* ICS dissertation, Nijmegen
167. Stephanie Wiesmann (2010), *"24/7 Negotiation in couples' transition to parenthood."* ICS dissertation, Utrecht
168. Borja Martinovic (2010), *"Interethnic contacts: A dynamic analysis of interaction between immigrants and natives in Western countries."* ICS dissertation, Utrecht
169. Anne Roeters (2010), *"Family life under pressure? Parents' paid work and the quantity and quality of parent-child and family time."* ICS dissertation, Utrecht
170. Jelle Sijtsema (2010), *"Adolescent aggressive behavior: Status and stimulation goals in relation to the peer context."* ICS dissertation, Groningen
171. Kees Keizer (2010), *"The spreading of disorder."* ICS dissertation, Groningen
172. Michael Mäs (2010), *"The diversity puzzle. explaining clustering and polarization of opinions."* ICS dissertation, Groningen
173. Marie-Louise Damen (2010), *"Cultuurdeelname en CKV: Studies naar effecten van kunsteducatie op de cultuurdeelname van leerlingen tijdens en na het voortgezet onderwijs."* ICS dissertation, Utrecht
174. Marieke van de Rakt (2011), *"Two generations of crime: The intergenerational transmission of convictions over the life course."* ICS dissertation, Nijmegen
175. Willem Huijnk (2011), *"Family life and ethnic attitudes: The role of the family for attitudes towards intermarriage and acculturation among minority and majority groups."* ICS dissertation, Utrecht
176. Tim Huijts (2011), *"Social ties and health in Europe: Individual associations, cross-national variations, and contextual explanations."* ICS dissertation, Nijmegen
177. Wouter Steenbeek (2011), *"Social and physical disorder: How community, business presence and entrepreneurs influence disorder in Dutch neighborhoods."* ICS dissertation, Utrecht
178. Miranda Vervoort (2011), *"Living together apart? Ethnic concentration in the neighborhood and ethnic minorities' social contacts and language practices."* ICS dissertation, Utrecht
179. Agnieszka Kanas (2011), *"The economic performance of immigrants: The role of human and social capital."* ICS dissertation, Utrecht
180. Lea Ellwardt (2011), *"Gossip in organizations: A social network study."* ICS dissertation, Groningen
181. Annemarije Oosterwaal (2011), *"The gap between decision and implementation: Decision making, delegation and compliance in governmental and organizational settings."* ICS dissertation, Utrecht
182. Natascha Notten (2011), *"Parents and the media: Causes and consequences of parental media socialization."* ICS dissertation, Nijmegen
183. Tobias Stark (2011), *"Integration in schools: A process perspective on students' interethnic attitudes and interpersonal relationships."* ICS dissertation, Groningen
184. Giedo Jansen (2011), *"Social cleavages and political choices: Large-scale comparisons of social class, religion and voting behavior in Western democracies."* ICS dissertation, Nijmegen
185. Ruud van der Horst (2011), *"Network effects on treatment results in a closed forensic psychiatric setting."* ICS dissertation, Groningen
186. Mark Levels (2011), *"Abortion laws in European Countries between 1960 and 2010: Legislative developments and their consequences for women's reproductive decision making."* ICS dissertation, Nijmegen
187. Marieke van Londen (2012), *"Exclusion of ethnic minorities in the Netherlands: The effects of individual and situational characteristics on opposition to ethnic policy and ethnically mixed neighbourhoods."* ICS dissertation, Nijmegen
188. Sigrid Mohnen (2012), *"Neighborhood context and health: How neighborhood social capital affects individual health."* ICS dissertation, Utrecht
189. Asya Zhelyazkova (2012), *"Compliance under controversy: Analysis of the transposition of European directives and their provisions."* ICS dissertation, Utrecht

190. Valeska Korff (2012), *"Between cause and control: Management in a humanitarian organization."* ICS dissertation, Groningen
191. Maike Gieling (2012), *"Dealing with diversity: Adolescents' support for civil liberties and immigrant rights."* ICS dissertation, Utrecht
192. Katya Ivanova (2012), *"From parents to partners: The impact of family on romantic relationships in adolescence and emerging adulthood."* ICS dissertation, Groningen
193. Jelmer Schalk (2012), *"The performance of public corporate actors: Essays on effects of institutional and network embeddedness in supranational, national, and local collaborative contexts."* ICS dissertation, Utrecht
194. Alona Labun (2012), *"Social networks and informal power in organizations."* ICS dissertation, Groningen
195. Michal Bojanowski (2012), *"Essays on social network formation in heterogeneous populations: Models, methods, and empirical analyses."* ICS dissertation, Utrecht
196. Anca Minescu (2012), *"Relative group position and intergroup attitudes in Russia."* ICS dissertation, Utrecht
197. Marieke van Schellen (2012), *"Marriage and crime over the life course: The criminal careers of convicts and their spouses."* ICS dissertation, Utrecht
198. Mieke Maliepaard (2012), *"Religious trends and social integration: Muslim minorities in the Netherlands."* ICS dissertation, Utrecht
199. Fransje Smits (2012), *"Turks and Moroccans in the Low Countries around the year 2000: Determinants of religiosity, trend in religiosity and determinants of the trend."* ICS dissertation, Nijmegen
200. Roderick Sluiter (2012), *"The diffusion of morality policies among Western European countries between 1960 and 2010: A comparison of temporal and spatial diffusion patterns of six morality and eleven non-morality policies."* ICS dissertation, Nijmegen
201. Nicoletta Balbo (2012), *"Family, friends and fertility."* ICS dissertation, Groningen
202. Anke Munniksma (2013), *"Crossing ethnic boundaries: Parental resistance to and consequences of adolescents' cross-ethnic peer relations"* ICS dissertation, Groningen
203. Anja-Kristin Abendroth (2013), *"Working women in Europe: How the country, workplace, and family context matter."* ICS dissertation, Utrecht
204. Katia Begall (2013), *"Occupational hazard? The relationship between working conditions and fertility."* ICS dissertation, Groningen
205. Hidde Bekhuis (2013), *"The popularity of domestic cultural products: Cross-national differences and the relation to globalization."* ICS dissertation, Utrecht
206. Lieselotte Blommaert (2013), *"Are Joris and Renske more employable than Rashid and Samira? A study on the prevalence and sources of ethnic discrimination in recruitment in the Netherlands using experimental and survey data."* ICS dissertation, Utrecht
207. Wiebke Schulz (2013), *"Careers of men and women in the 19th and 20th centuries."* ICS dissertation, Utrecht
208. Ozan Aksoy (2013), *"Essays on social preferences and beliefs in non-embedded social dilemmas."* ICS dissertation, Utrecht
209. Dominik Morbitzer (2013), *"Limited farsightedness in network formation."* ICS dissertation, Utrecht
210. Thomas de Vroome (2013), *"Earning your place: The relation between immigrants' economic and psychological integration in the Netherlands."* ICS dissertation, Utrecht
211. Marloes de Lange (2013), *"Causes and consequences of employment flexibility among young people: Recent developments in the Netherlands and Europe."* ICS dissertation, Nijmegen
212. Roza Meuleman (2014), *"Consuming the Nation: Domestic cultural consumption: Its stratification and relation with nationalist attitudes."* ICS dissertation, Utrecht
213. Esther Havekes (2014), *"Putting interethnic attitudes in context: The relationship between neighbourhood characteristics, interethnic attitudes and residential behaviour."* ICS dissertation, Utrecht
214. Zoltán Lippényi (2014), *"Transitions toward an open society? Intergenerational occupational mobility in Hungary in the 19th and 20th centuries."* ICS dissertation, Utrecht
215. Anouk Smeekes (2014), *"The presence of the past: Historical rooting of national identity and current group dynamics."* ICS dissertation, Utrecht
216. Michael Savelkoul (2014), *"Ethnic diversity and social capital: Testing underlying explanations derived from conflict and contact theories in Europe and the United States."* ICS dissertation, Nijmegen

217. Martijn Hogerbrugge (2014), *"Misfortune and family: How negative events, family ties, and lives are linked."* ICS dissertation, Utrecht
218. Gina-Felicia Potarca (2014), *"Modern love: Comparative insights in online dating preferences and assortative mating."* ICS dissertation, Groningen
219. Mariska van der Horst (2014), *"Gender, aspirations, and achievements: Relating work and family aspirations to occupational outcomes."* ICS dissertation, Utrecht
220. Gijs Huitsing (2014), *"A social network perspective on bullying"* ICS dissertation, Groningen
221. Thomas Kowalewski (2015), *"Personal growth in organizational contexts."* ICS dissertation, Groningen
222. Manuel Muñoz-Herrera (2015), *"The impact of individual differences on network relations: Social exclusion and inequality in productive exchange and coordination games."* ICS dissertation, Groningen
223. Tim Immerzeel (2015), *"Voting for a change: The democratic lure of populist radical right parties in voting behavior."* ICS dissertation, Utrecht
224. Fernando Nieto Morales (2015), *"The control imperative: Studies on reorganization in the public and private sectors."* ICS dissertation, Groningen
225. Jellie Sierksma (2015), *"Bounded helping: How morality and intergroup relations shape children's reasoning about helping."* ICS dissertation, Utrecht
226. Tinka Veldhuis (2015), *"Captivated by fear: An evaluation of terrorism detention policy."* ICS dissertation, Groningen
227. Miranda Visser (2015), *"Loyality in humanity: Turnover among expatriate humanitarian aid workers."* ICS dissertation, Groningen
228. Sarah Westphal (2015), *"Are the kids alright? Essays on postdivorce residence arrangements and children's well-being."* ICS dissertation, Utrecht
229. Britta Rüschoff (2015), *"Peers in careers: Peer relationships in the transition from school to work."* ICS dissertation, Groningen
230. Nynke van Miltenburg (2015), *"Cooperation under peer sanctioning institutions: Collective decisions, noise, and endogenous implementation."* ICS dissertation, Utrecht
231. Antonie Knigge (2015), *"Sources of sibling similarity: Status attainment in the Netherlands during modernization."* ICS dissertation, Utrecht
232. Sanne Smith (2015), *"Ethnic segregation in friendship networks: Studies of its determinants in English, German, Dutch, and Swedish school classes."* ICS dissertation, Utrecht
233. Patrick Präg (2015), *"Social stratification and health: Four essays on social determinants of health and wellbeing."* ICS dissertation, Groningen
234. Wike Been (2015), *"European top managers' support for work-life arrangements"* ICS dissertation, Utrecht
235. André Grow (2016), *"Status differentiation: New insights from agent-based modeling and social network analysis."* ICS dissertation, Groningen
236. Jesper Rözer (2016), *"Family and personal networks: How a partner and children affect social relationships."* ICS dissertation, Utrecht
237. Kim Pattiselanno (2016), *"At your own risk: The importance of group dynamics and peer processes in adolescent peer groups for adolescents' involvement in risk behaviors."* ICS- dissertation, Groningen
238. Vincenz Frey (2016), *"Network formation and trust."* ICS dissertation, Utrecht
239. Rozemarijn van der Ploeg (2016), *"Be a buddy, not a bully? Four studies on social and emotional processes related to bullying, defending, and victimization."* ICS dissertation, Groningen
240. Tali Spiegel (2016), *"Identity, career trajectories and wellbeing: A closer look at individuals with degenerative eye conditions."* ICS- dissertation, Groningen
241. Felix Christian Tropf (2016), *"Social science genetics and fertility."* ICS dissertation, Groningen
242. Sara Geven (2016), *"Adolescent problem behavior in school: The role of peer networks."* ICS dissertation, Utrecht
243. Josja Rokven (2016), *"The victimization-offending relationship from a longitudinal perspective."* ICS dissertation, Nijmegen
244. Maja Djundeva (2016), *"Healthy ageing in context: Family welfare state and the life course."* ICS dissertation, Groningen
245. Mark Visser (2017), *"Inequality between older workers and older couples in the Netherlands: A dynamic life course perspective on educational and social class differences in the late career."* ICS dissertation, Nijmegen

246. Beau Oldenburg (2017), *"Bullying in schools: The role of teachers and classmates."* ICS dissertation, Groningen
247. Tatang Muttaqin (2017), *"The education divide in Indonesia: Four essays on determinants of unequal access to and quality of education."* ICS dissertation, Groningen
248. Margriet van Hek (2017), *"Gender inequality in educational attainment and reading performance: A contextual approach."* ICS dissertation, Nijmegen
249. Melissa Verhoef-van Dorp (2017), *"Work schedules, childcare and well-being: Essays on the associations between modern-day job characteristics, childcare arrangements and the well-being of parents and children."* ICS dissertation, Utrecht
250. Timo Septer (2017), *"Goal priorities, cognition and conflict: Analyses of cognitive maps concerning organizational change."* ICS dissertation, Groningen
251. Bas Hofstra (2017), *"Online social networks: Essays on membership, privacy, and structure."* ICS dissertation, Utrecht
252. Yassine Khoudja (2018), *"Women's labor market participation across ethnic groups: The role of household conditions, gender role attitudes, and religiosity in different national contexts."* ICS dissertation, Utrecht
253. Joran Laméris (2018), *"Living together in diversity: Whether, why and where ethnic diversity affects social cohesion."* ICS dissertation, Nijmegen
254. Maaike van der Vleuten (2018), *"Gendered Choices: Fields of study of adolescents in the Netherlands."* ICS dissertation, Utrecht
255. Mala Sondang Silitonga (2018), *"Corruption in Indonesia: The impact of institutional change, norms, and networks."* ICS dissertation, Groningen
256. Manja Coopmans (2018), *"Rituals of the past in the context of the present: The role of Remembrance Day and Liberation Day in Dutch society."* ICS dissertation, Utrecht
257. Paul Hindriks (2018), *"The struggle for power: Attitudes towards the political participation of ethnic minorities."* ICS dissertation, Utrecht
258. Nynke Niezink (2018), *"Modeling the dynamics of networks and continuous behavior."* ICS dissertation, Groningen
259. Simon de Bruijn (2018), *"Reaching agreement after divorce and separation: Essays on the effectiveness of parenting plans and divorce mediation."* ICS dissertation, Utrecht
260. Susanne van 't Hoff-de Goede (2018), *"While you were locked up: An empirical study on the characteristics, social surroundings and wellbeing of partners of prisoners in The Netherlands."* ICS dissertation, Utrecht
261. Loes van Rijsewijk (2018), *"Antecedents and consequences of helping among adolescents."* ICS dissertation, Groningen
262. Mariola Gremmen (2018), *"Social network processes and academic functioning: The role of peers in students' school well-being, academic engagement, and academic achievement."* ICS dissertation, Groningen
263. Jeanette Renema (2018), *"Immigrants' support for welfare spending: The causes and consequences of welfare usage and welfare knowledgeability."* ICS dissertation, Nijmegen
264. Suwatin Miharti (2018), *"Community health centers in Indonesia in the era of decentralization: The impact of structure, staff composition and management on health outcomes."* ICS dissertation, Groningen
265. Chaïm la Roi (2019), *"Stigma and stress: Studies on attitudes towards sexual minority orientations and the association between sexual orientation and mental health."* ICS dissertation, Groningen
266. Jelle Lössbroek (2019), *"Turning grey into gold: Employer-employee interplay in an ageing workforce."* ICS dissertation, Utrecht
267. Nikki van Gerwen (2019), *"Employee cooperation through training: A multi-method approach."* ICS dissertation, Utrecht
268. Paula Thijs (2019), *"Trends in cultural conservatism: The role of educational expansion, secularisation, and changing national contexts."* ICS dissertation, Nijmegen
269. Renske Verweij (2019), *"Understanding childlessness: Unravelling the link with genes and the socio-environment."* ICS dissertation, Groningen
270. Niels Blom (2019), *"Partner relationship quality under pressing work conditions: Longitudinal and cross-national investigation."* ICS dissertation, Nijmegen

271. Müge Simsek (2019), *"The dynamics of religion among native and immigrant youth in Western Europe."* ICS dissertation, Utrecht
272. Leonie van Breeschoten (2019), *"Combining a career and childcare: The use and usefulness of work-family policies in European organizations."* ICS dissertation, Utrecht
273. Roos van der Zwan (2019), *"The political representation of ethnic minorities and their vote choice."* ICS dissertation, Nijmegen
274. Ashwin Rambaran (2019), *"The classroom as context for bullying: A social network approach."* ICS dissertation, Groningen
275. Dieko Bakker (2019), *"Cooperation and social control: Effects of preferences, institutions, and social structure."* ICS dissertation, Groningen
276. Femke van der Werf (2019), *"Shadow of a rainbow? National and ethnic belonging in Mauritius."* ICS dissertation, Utrecht
277. Robert Krause (2019), *"Multiple imputation for missing network data."* ICS dissertation, Groningen
278. Take Sipma (2020), *"Economic insecurity and populist radical right voting."* ICS dissertation, Nijmegen
279. Mathijs Kros (2020), *"The nature of negative contact: Studies on interethnic relations in Western societies."* ICS dissertation, Utrecht
280. Lonneke van den Berg (2020), *"Time to leave: Individual and contextual explanations for the timing of leaving home."* ICS dissertation, Amsterdam
281. Marianne Hooijsma (2020), *"Clashrooms: Interethnic peer relationships in schools."* ICS dissertation, Groningen
282. Marina Tulin (2020*), "Blind spots in social resource theory: Essays on the creation, maintenance and returns of social capital."* ICS dissertation, Amsterdam
283. Tessa Kaufman (2020), *"Toward tailored interventions: Explaining, assessing, and preventing persistent victimization of bullying."* ICS dissertation, Groningen
284. Lex Thijssen (2020), *"Racial and ethnic discrimination in western labor markets: Empirical evidence from field experiments."* ICS dissertation, Utrecht
285. Lukas Norbutas (2020), *"Trust on the dark web: An analysis of illegal online drug markets."* ICS dissertation, Utrecht
286. Tomáš Diviák (2020), *"Criminal networks: Actors, mechanisms, and structures."* ICS dissertation, Groningen
287. Tery Setiawan (2020), *"Support for interreligious conflict in Indonesia."* ICS dissertation, Nijmegen
288. Vera de Bel (2020), *"The ripple effect in family networks: Relational structures and well-being in divorced and non-divorced families."* ICS dissertation, Groningen
289. Diego Palacios (2020), *"How context and the perception of peers' behaviors shape relationships in adolescence: A multiplex social network perspective."* ICS dissertation, Groningen
290. Saskia Glas (2020), *"Where are the Muslim Feminists? Religiosity and support for gender equality in the Arab region."* ICS dissertation, Nijmegen
291. Tomas Turner-Zwinkels (2020), *"A new macro-micro approach to the study of political careers: Theoretical, Methodological and empirical challenges and solutions."* ICS dissertation, Groningen
292. Lotte Scheeren (2020), *"Not on the same track? Tracking age and gender inequality in education."* ICS dissertation, Amsterdam
293. Joris Broere (2020), *"Essays on how social network structure affects asymmetric coordination and trust."* ICS dissertation, Utrecht
294. Marcus Kristiansen (2021*), "Contact with benefits: How social networks affect benefit receipt dynamics in the Netherlands."* ICS dissertation, Utrecht
295. Judith Kas (2021), *"Trust and reputation in the peer-to-peer platform economy."* ICS dissertation, Utrecht
296. Andrea Forster (2021), *"Navigating educational institutions: Mechanisms of educational inequality and social mobility in different educational systems."* ICS dissertation, Amsterdam
297. Jannes ten Berge (2021), *"Technological change and work: The relation between technology implementation within organizations and changes in workers' employment."* ICS dissertation, Utrecht
298. Jolien Geerlings (2021), "*Teaching in culturally diverse classrooms: The importance of dyadic relations between teachers and children."* ICS dissertation, Utrecht
299. Kirsten van Houdt (2021), *"Stepfamilies in adulthood: Solidarity between parents and adult children."* ICS dissertation, Amsterdam

300. Suzanne de Leeuw (2021), *"The intergenerational transmission of educational attainment after divorce and remarriage."* ICS dissertation, Amsterdam
301. Fleur Goedkoop (2021), *"Involvement in bottom-up energy transitions: The role of local and contextual embeddedness."* ICS dissertation, Groningen
302. Eva Vriens (2021), *"Mutualism in the 21st century: The why, when, and how behind successful risk-sharing institutions."* ICS dissertation, Utrecht
303. Ardita Muja (2021), *"From school to work: The role of the vocational specificity of education in young people's labor market integration."* ICS dissertation, Nijmegen
304. Siyang Kong (2021), *"Women and work in contemporary China: The effect of market transition on women's employment, earnings, and status attainment."* ICS dissertation, Utrecht
305. Marijn Keijzer (2022), *"Opinion dynamics in online social media."* ICS dissertation, Groningen
306. Sander Kunst (2022), *"The educational divide in openness towards globalisation in Western Europe."* ICS dissertation, Amsterdam
307. Nella Geurts (2022), *"Puzzling pathways: The integration paradox among migrants in Western Europe."* ICS dissertation, Nijmegen
308. Dragana Stojmenovska (2022), *"Men's place: The incomplete integration of women in workplace authority."* ICS dissertation, Amsterdam
309. Bram Hogendoorn (2022), *"Divorce and inequality: Stratification in the risk and consequences of union dissolution."* ICS dissertation, Amsterdam
310. Tom Nijs (2022), *"This place is ours: Collective psychological ownership and its social consequences."* ICS dissertation, Utrecht
311. Nora Storz (2022), *"'This land is ours—but is it also theirs?' Collective ownership beliefs and reconciliation in territorial conflict regions."* ICS dissertation, Utrecht
312. Tara Koster (2022), *"Parenting and fairness in diverse families."* ICS dissertation, Utrecht
313. Danelien van Aalst (2022), *"Elements Contributing to Teachers' Role in Bullying."* ICS dissertation, Groningen
314. Wybren Nooitgedagt (2022), *"Who owns the country? Collective psychological ownership and intergroup relations in settler societies."* ICS dissertation, Utrecht
315. Marija Dangubić (2022), *"Rejecting Muslim minority practices: Principles and prejudices."* ICS dissertation, Utrecht
316. Sara Cvetkovska (2022), *"Lines in the shifting sand: The implications of being tolerated."* ICS dissertation, Utrecht
317. Maaike Hornstra (2022), *"Going beyond the dyad: Adult intergenerational closeness after divorce and remarriage."* ICS dissertation, Amsterdam
318. Wouter Kiekens (2022), *"Sexual and gender minority youth's mental health and substance use: Disparities, mechanisms, and protective factors."* ICS dissertation, Groningen
319. Carlijn Bussemakers (2022), *"Adversity and educational inequality: The interplay between adverse experiences and parental resources for children's educational attainment."* ICS dissertation, Nijmegen
320. Evi Velthuis (2022), *" 'To tolerate or not to tolerate?' Reasons for tolerance of minority group practices among majority members in the Netherlands and Germany."* ICS dissertation, Utrecht
321. Hendrik Nunner (2023), *"Co-evolution of social networks and infectious diseases."* ICS dissertation, Utrecht
322. Carly van Mensvoort (2023), *"Inspiring leaders: An empirical study of female supervisors at work."* ICS dissertation, Nijmegen
323. Anne van der Put (2023), *"Healthy at work: The role of the work environment in worksite health promotion."* ICS dissertation, Utrecht
324. Rowan ten Kate (2023), *"Understanding loneliness among older migrants."* ICS dissertation, Groningen
325. Sanne Kellij (2023), *"I see, I see what you don't see: Neural and behavioral social-cognitive processes underlying (persistent) victimization."* ICS dissertation, Groningen
326. Inge Hendriks (2023), *"Changes and contrasts in attitudes towards ethnic minorities."* ICS dissertation, Nijmegen
327. Eleonora Marucci (2023), *"Antecedents and consequences of teacher attunement in primary and secondary school classrooms."* ICS dissertation, Groningen

328. Kasper Otten (2023), *"Cooperation in changing groups: How newcomers and norms shape public good provision in the lab, online games, and the field."* ICS dissertation, Utrecht
329. Carlos de Matos Fernandes (2023), *"In or out? The paradox of exclusionary mechanisms in keeping cooperation going."* ICS dissertation, Groningen
330. Ruohuang Jiao (2023), *"Reputation effects in peer-to-peer online markets: meta-analyses and laboratory experiments."* ICS dissertation, Utrecht
331. Marlou Ramaekers (2024), *"Informal helping. Insights from a dyadic, family and societal perspective."* ICS dissertation, Nijmegen
332. Kim Stienstra (2024), *"Educational quality and inequality: The interplay between schools, families, and genes."* ICS dissertation, Utrecht
333. Ece Arat (2024), *"Diverse stepfamilies: Parenting and children's well-being."* ICS dissertation, Utrecht
334. Christian Fang (2024), *"Family life in postdivorce families."* ICS dissertation, Utrecht
335. Vera Buijs (2024), *"Close relationships and subjective well-being: A life course perspective on social needs, relations, and interactions."* ICS dissertation, Groningen
336. Twan Huijsmans (2024), *"Our place in politics: Urban-rural divergence and how place affects political attitudes."* ICS dissertation, Amsterdam
337. Thomas Teekens (2024), *"Sustainable collaboration in care: Joint production motivation and interprofessional learning in an interorganizational network."* ICS dissertation, Groningen
338. Ana Macanović (2024), *"Trust in the shadows: The role of communication in extra-legal contexts*." ICS dissertation, Utrecht
339. Christoph Janietz (2024), *"Inequality at work: Occupations, organizations, and the wage distribution."* ICS dissertation, Amsterdam
340. Philipp Schneider (2024), *"Social Influence and the Energy Transition: Leveraging social networks and norms."* ICS dissertation, Utrecht
341. Nick Wuestenenk (2024), "*Support for sexual liberalization among ethnic majorities and minorities in Europe: The role of social norms in the public expression of opinions.*" ICS dissertation, Utrecht
342. Renae Sze Ming Loh (2024), *"Copy Paste? Digital skills, social reproduction, and social mobility in education."* ICS dissertation, Nijmegen
343. Klara Raiber (2024), *"Beyond sacrifice? Long-term employment consequences of providing unpaid care."* ICS dissertation, Nijmegen
344. Sofie Lorijn (2024), "*Peer relationships in the transition from primary to secondary education.*" ICS dissertation, Groningen
345. Xingna Qin (2024), "*In connection – The role of peers, parents, and teachers in adolescent friendship dynamics."* ICS dissertation, Groningen
346. Echo Teng Li (2024), *"Simulation models of the collective consequences of bounded rationality in opinion formation in networks: Cases of market concentration and vaccination opinion polarization."* ICS dissertation, Groningen
347. Thomas Feliciani (2025), *"Divided spaces and divided opinions: Modeling the impact of residential segregation on opinion polarization."* ICS dissertation, Groningen
348. Thijmen Jeroense (2025), *"Social network segregation: Studies on network homogeneity in sociodemographic characteristics and political attitudes."* ICS dissertation, Nijmegen
349. Dieuwke Zwier (2025), "*Rhythms of class: Socio-economic disparities and peer dynamics in secondary school choice*." ICS dissertation, Amsterdam
350. Lian van Vemde (2025), *"Creating inclusive societies: Fostering belonging and positive intergroup relations in culturally diverse classrooms."* ICS dissertation, Utrecht
351. Rob Franken (2025), *"Social networks and sports participation: The interplay of selection and influence processes."* ICS dissertation, Nijmegen
352. Julian Rengers (2025), *"Selective disclosure at work: Lesbian, gay, and bisexual employees' disclosure decisions, motivations, and approaches."* ICS dissertation, Groningen
353. Sara Wiertsema (2025), "*Young adults in motion: Understanding the impact of the school-to-work transition on sports and physical activity."* ICS dissertation, Nijmegen
354. Sofie Wiersma (2025), *"Imprints at work: How the pasts of organizations and leaders shape workplace precarity and inequality."* ICS dissertation, Groningen

355. Tessa Ubels (2025), *"Stuck in migration: Pathways to social change and well-being for people on the move and the role of psychosocial support."* ICS dissertation, Nijmegen
356. Katrin Müller (2025), *"Shifting perspectives: majority members' perceptions of the prevalence of ethno-racial minority discrimination."* ICS dissertation, Nijmegen
357. Maikel Meijeren (2026), *"Helping those in need: volunteering for humanitarian organizations and for refugees."* ICS dissertation, Nijmegen
358. Jonas Stein (2026), "*Reconciling epistemic and identity diversity: Identifying pathways to better decision-making in social groups."* ICS dissertation, Groningen
359. Mustafa Firat (2026), "*Retirement and life course transitions: Inequality and the welfare state in Europe*." ICS dissertation, Nijmegen
360. Rachel Kollar (2026), *"What's religion got to do with it? Islamic religiosity and the political participation of European Muslims."* ICS dissertation, Nijmegen
361. Zhe Dong (2026), *"Positive and Negative Leadership in Childhood: Emergence, Group Process, and Outcomes."* ICS dissertation, Groningen
362. Tobias Sachs (2026), *"Beyond Bias: Examining Conditions That Shape Ethnic Discrimination in Resume Screening and Employees´ Support for Organizational Inclusion."* ICS dissertation, Amsterdam
363. Viktor Decker (2026), *"(Un)equal Careers: Labor Market Inequalities Between Workers With General and Vocational Education Over the Life Course"* ICS dissertation, Amsterdam
364. Sanjana Singh (2026), *"Structural Gender Inequality at Work: How integration and the accumulation of advantages shape (in)equality*." ICS dissertation, Utrecht

Why does workplace gender inequality persist even as women's representation improves? This book moves beyond headcounts to show how inequality is structurally shaped by two overlooked mechanisms: informal social integration and cumulative (dis)advantage, where early successes or setbacks grow into large career gaps over time.

Drawing on employee social network data from nine countries and six sectors, the research finds gender-diverse networks to benefit both men and women; yet many diversity policies fail to foster them. A large-scale simulation study further reveals that common analytical approaches often miss how inequality accumulates. Applying an improved approach to 2.8 million US patent applications, the book shows that early-career setbacks lead to lasting disadvantages for women, while men are far more likely to recover.

Current policies, like board quotas, often arrive too late. To build a fair workplace, policymakers must shift their focus from headcounts to early-career interventions and meaningful social integration.

Sanjana Singh obtained her master's degree in Economic Policy at University College London. She conducted the present research at the Department of Sociology at Utrecht University, as part of the Interuniversity Center for Social Science Theory and Methodology (ICS), and the transdisciplinary SCOOP research program on sustainable cooperation.