Understanding the Gender Performance Gap among Star Performers in STEM Fields

by Young Hun Ji

B.A. in Economics, May 2009, Amherst College
Master of Commerce in Management, February 2012, The University of Auckland

A Dissertation submitted to

The Faculty of

The School of Business
of The George Washington University
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

August 31, 2017

Dissertation directed by

Herman Aguinis
Avram Tucker Distinguished Scholar and Professor of Business

The School of Business of The George Washington University certifies that Young Hun Ji has passed the final examination for the degree of Doctor of Philosophy as of July 26, 2017. This is the final and approved form of the dissertation.

Understanding the Gender Performance Gap among Star Performers in STEM Fields

Young Hun Ji

Dissertation Research Committee:

Herman Aguinis, Avram Tucker Distinguished Scholar and Professor of Business, Dissertation Director

N. Andrew Cohen, Assistant Professor of Management, Committee Member

N. Sharon Hill, Associate Professor of Management, Committee Member

James B. Wade, Avram Tucker Professor of Leadership and Strategy, Committee Member

Acknowledgements

First, I wish to express my deepest regards to my dissertation chair, Dr. Herman Aguinis for all of his guidance and mentorship throughout the time of my dissertation research. This research would not have been possible without his expertise, endless support, patience, and understanding. Dr. Aguinis has been extremely generous with his support and guidance all throughout the time that I have known him, and I will forever be in his debt. I am also extremely grateful to my dissertation committee members: Drs. Andrew Cohen, Sharon Hill, and James Wade for their invaluable guidance and feedback. My dissertation committee has been nothing short of supportive and inspiring, and I cannot thank them enough. Next, I would like to thank my family members, who provided me with invaluable support time after time throughout the course of my program. In particular, I thank my parents and sister for loving and supporting me unconditionally. Finally, I thank the National Science Foundation (Science of Organizations Program)—grant number 1643075 titled "Understanding the Gender Performance Gap among Star Performers in STEM Fields" to Herman Aguinis (PI)—for supporting the work described in this dissertation.

Abstract of Dissertation

Understanding the Gender Performance Gap among Star Performers in STEM Fields

Despite much improvement over the past several decades, women continue to be underrepresented across many STEM fields. In this study, I draw upon past research to theorize that (1) there exist a substantial gender performance gap among STEM researchers and that (2) the gap is disproportionately larger among *star performers*, i.e., individuals who produce output many times greater than others holding the same job (Aguinis & O’Boyle, 2014). I then discuss how a gender performance gap specifically among star performers can be more harmful to the underrepresented group than an equivalent gap among average performers. To investigate the possible existence of such gender performance gaps, I assess the research productivity of all researchers in the fields of mathematics, materials sciences, and genetics who have published in the past decade at least one article in the most influential journals in their fields. Using the process of *distribution pitting* (Joo, Aguinis & Bradley, 2017), I identify the best-fitting theoretical distributions and associated dominant generative mechanisms that shape individual performance across the three STEM fields. Assessment of the shapes of the performance distributions confirms the existence of considerable gender performance gaps in favor of men, although the gap was substantially lower in the field of genetics compared to in the others. In addition, the findings suggest that (1) individual STEM researchers vary in performance predominantly due to differences in their accumulation rates (i.e., average output produced per time period), and (2) women’s research output accumulation rates are lower (on average) and also less variable compared to men’s. Implications for theory and practice based on these findings are discussed.

Table of Contents

List of Tables

List of Figures

CHAPTER I

INTRODUCTION

In this chapter, I articulate the research problem at hand by first discussing the persistent underrepresentation of women in STEM fields and the possible existence of gender performance gaps specifically among star (or elite) performers in these domains. Second, I provide an overview of existing theories on the underlying causes for women's relative absence in STEM fields, followed by an overview of recent research on star performers. I then conclude the statement of my research problem by discussing why gender performance gaps specifically among stars can be more harmful (to the underrepresented gender group) than equivalent gender performance gaps among average performers. Next, I outline my key research questions and articulate the purpose of this study. Finally, I explain the significance of this study for both theory and practice.

Statement of the Problem

Underrepresentation of Women in STEM fields

Scholars and policymakers have long noted the persistent underrepresentation of women in the fields of science, technology, engineering, and mathematics (STEM). According to the 2016 Science and Engineering Indicators by the NSF, although the gender disparity has declined in recent times, women continue to be underrepresented in STEM fields. Research has shown that, in early (K-12) education, boys and girls display similar performance and participation rates in mathematics and science (e.g., Hyde et al., 2008; Xie and Shauman, 2003); however, large gender imbalances in representation can be observed in higher-level academic STEM fields and in the STEM workforce. For example, although women make up half of the college-educated workforce in the U.S., they only make up 29% of the STEM workforce. The underrepresentation is particularly severe for minority women: less than 5% of bachelor's degrees in engineering and in computer sciences and less than 10% of degrees in physical sciences, mathematics, and biological sciences were awarded to minority women in 2012 (NSF, Women, Minorities, and People with Disabilities in Science and Engineering, 2015). Women are also underrepresented in leadership positions across many STEM fields. According to a survey by the Association of American Universities, women chaired only 2.7% of engineering departments, 5.9% of math or physical science departments, and 12.7% of life science departments (Niemeier & Gonzalez, 2004).

An underlying premise of this study is that there are large gender performance gaps in STEM fields that are especially severe among *star performers*. Star performers refer to individuals who produce output and results many times greater than the rest of the individuals holding the same job or position (Aguinis & Bradley, 2015; Aguinis & O'Boyle, 2014; Aguinis, O'Boyle, Gonzalez-Mule, & Joo, 2016). If one was to visualize the performance distribution of all individuals holding the same job, then, star performers would be those at the very right tail of the performance distribution. In this study, I suggest that women's STEM performance is less variable (compared to men's) and that women are thus increasingly underrepresented towards the upper extreme (i.e., right tail) of the performance distribution. More specifically, I suggest that there are certain mechanisms and processes that lead to women's STEM performance being lower (on average) and less variable than men's. This would mean that women are more homogenous than men in terms of their STEM performance, that female representation among stars is even lower than among all STEM performers, and that star women are outperformed by the best of star men. The existence of such gender performance gaps would be a very serious issue, suggesting that women are not only underrepresented but also disproportionately less likely than men to achieve success (and stardom) in STEM fields. Disconcertingly, as I discuss later, findings from prior research on the relative absence of women in STEM fields hint at the existence of such gender performance gaps among stars.

Accordingly, the primary goal of this research is to investigate the existence of gender performance gaps among stars performers in STEM fields and to explore the possible underlying mechanisms and processes that produce those gaps. In the section below, I provide an overview of the main theoretical perspectives generated by prior research into the underrepresentation of women in STEM fields. This discussion serves to outline the major theories on the issue that researchers have proffered over the past several decades, and to illustrate the implied possibility of large gender performance gaps among star performers.

An Overview of Theories on the Gender Disparity in STEM Fields

Over the past several decades, researchers have identified a wide array of factors that contribute to the underrepresentation of women in STEM fields. Currently, there are three broad theoretical frames regarding the cause of the general underrepresentation of women in STEM fields. The first of these theories is that the gender disparity can be explained largely by sex differences in quantitative abilities (Halpern, 2000; Halpern et al., 2007). More broadly, the biological perspective suggests that there are considerable sex differences in aptitudes across cognitive domains. Many researchers have found support for the conventional belief that men outperform women on quantitative tasks and vice versa on verbal tasks (for a review, see Halpern et al., 2007). According to this line of research, biological factors such as exposure to prenatal and postnatal testosterone and brain lateralization enable men to outperform women in mathematical and visuospatial tasks (e.g., Baron-Cohen, 2003). Men have also been found to have greater variability in quantitative abilities compared to women, displaying greater representation in both the left and right tails of the ability distribution (Halpern et al., 2007; Wai, Cacchio, Putallaz, & Makel, 2010). In sum, the 'sex differences' theory suggests that women's underrepresentation in STEM fields is mainly caused by biological factors that enable men to outperform women in quantitative tasks.

The second of these theories stems from a body of research suggesting that women are 'pushed out' of STEM fields due to discrimination. According to this line of research, negative stereotypes about women (regarding their STEM abilities) contribute to prejudice and discrimination against female scientists (Cali, Alawa, Lee, Zhao, & Kim, 2016). Studies suggest that women in STEM fields experience discrimination in the form of limited opportunities for advancement (Xu, 2008), being perceived as less competent than men with comparable achievements (Moss-Racusin, Dovidio, Brescoll, Graham, & Handelsman, 2012), and receiving less credit for collaborative output (Sarsons, 2017). Also, numerous experiments have found support for possible gender discrimination effects in hiring, journal reviewing, and grant funding (e.g., Chesler, Barabino, Bhatia, & Richards-Kortum, 2010; Lortie et al., 2007). This perspective suggests that women are underrepresented in STEM fields mainly because are pushed out of STEM fields due to structural barriers to entry and success.

The third view is that women's underrepresentation in STEM fields is largely the result of the particular career and lifestyle choices made by women. This perspective suggests that the gender imbalance is the result of a wide range of psychological factors that shape women's decision to pursue a STEM career in the first place and to persist in the field once they opt in to a STEM profession (Ceci & Williams, 2010). According to this perspective, the underrepresentation of women in STEM fields can be largely explained by disparate motivations across genders with respect to pursuing and persisting in STEM careers (e.g., Ceci & Williams, 2010, Wang & Degol, 2003). Gender differences in motivational factors such as STEM interest, work goals, occupational preferences, and work/life values result in women being more likely to choose careers outside of STEM. According to this body of research, women are not just being 'pushed out' of STEM fields due to discrimination or lack of quantitative ability but are choosing to opt out of STEM fields at higher rates than men at all stages of their careers due to a complex interaction of a wide range of factors (Kossek, Su, & Wu, 2017).

In sum, the research suggests that the underrepresentation of women in STEM fields could be caused by a wide array of biological, socio-cultural/contextual, psychological, and motivational factors. Each of the three theories offers a unique perspective regarding the most important factors that contribute to the observed gender disparity in STEM fields. In simplified terms, the sex differences theory emphasizes mainly biological impacts on each gender's *ability* to perform; discrimination theories emphasize gendered barriers in individuals' *opportunity* to perform; and the 'women's career and life choices' theory emphasizes gender disparities in individuals' *motivation* to perform in STEM domains.

Each of the three gender perspectives provides some theoretical support to the notion that (1) there exists a gender performance gap in favor of men in STEM fields and that (2) the gap is disproportionately large among star performers than across all performers. First, assuming that quantitative aptitudes are an important driver of STEM performance, there should be a performance gap in favor of men, as men have been found to have higher quantitative aptitudes compared to women on average. The sex differences perspective has also shown that men are more variable in their quantitative aptitudes, which suggests that men may thus be similarly more variable in their STEM performance. This implies that men have increasing representation towards the upper end (i.e., right tail) of the performance distribution and that the performance gap among stars—or any other segment of above-average performers—is greater than among average performers.

Second, there may be a general performance gap in favor of men due to various forms of gender discrimination—such as biases in grant funding and peer reviews—that constrain the productivity of women in STEM fields. In addition, the negative effects of such discrimination on female performance may accumulate at a disproportionately greater rate for star (and other highly productive) women. To illustrate, highly talented STEM researchers who intend to produce more research compared to average STEM performers would need greater access to resources (e.g., grant funding) and would also face more frequent work product evaluations (e.g., peer reviews) in order to turn the research into concrete performance outcomes (e.g., journal publications). Assuming that gender discrimination constrains women's STEM performance at potentially every opportunity to perform, the cumulative negative effects of such discrimination on (women's) performance could be largest among the most productive individuals.

Finally, from the career and lifestyle choices perspective, there may be a gender performance gap in favor of men due to gender differences in various motivational and other psychological factors. For example, women in STEM fields may be more likely than their male colleagues to prioritize family goals and responsibilities over work, meaning there may be gender differences in time and resources devoted to one's work and thus a performance gap in favor of men. Additionally, certain life choices (e.g., the decision to have children early in one's career) may lead to larger productivity losses for women than for men. It is also possible that motivational differences between men and women in these fields are most apparent at the 'upper extreme' in terms of time and effort dedicated to one's work. Theoretically, men and women may devote more-or-less similar amounts of resources to their work on average, but men could have disproportionate representation among the small proportion of individuals who devote extraordinary amounts of resources and time to their work.

In sum, each of the three gender perspectives provides some theoretical support for the worrisome possibility that (1) there is an overall gender performance gap in favor of men in STEM fields, and that (2) women are less variable compared to men in their STEM performance, which would suggest that the gender performance gap is particularly large among star performers. The implied effects of each gender perspective on the individual performance distributions of men versus women in STEM fields—specifically in terms of distribution shape and right tail heaviness—are discussed in more detail in the following chapter. In the sections that follow below, I provide a brief discussion of recent research on star performers and discuss why the existence of large gender performance gaps specifically among star performers in STEM fields would be so detrimental.

Star Performers

Recent research has documented star performers across many jobs, occupations, and industries (Aguinis, O'Boyle, Gonzalez-Mulé, & Joo, 2016; O'Boyle & Aguinis, 2012). The presence of star performers observed in various contexts negates the long-held implicit assumption that individual performance follows a normal distribution. A recent set of studies provides evidence that performance in many contemporary occupations and jobs, measured in terms of cumulative, objective output/results over a substantial period of time (e.g., the total number of scholarly publications in a ten-year timeframe), follows a heavy right tailed distribution (Aguinis & O'Boyle, 2014; Aguinis et al., 2016; O'Boyle & Aguinis, 2012). For such job contexts, then, certain organizational practices (e.g., forced-response performance evaluations) and statistical analyses (e.g., deletion or correction of outliers) based on an (inaccurate) assumption of normally distributed performance output may underestimate the existence and impact of star performers (O'Boyle & Aguinis, 2012). The non-normality of individual performance output distributions is an important finding because the existence of star performers has substantial implications for organizational research and practice, including areas such as personnel recruitment and selection, training and development, remuneration, and so on (Aguinis & O'Boyle, 2014).

As illustrated in Figure 1, there are two important consequences of individual performance following a heavy right tailed distribution as opposed to a normal distribution: (1) the majority of performers are actually below the mean (μ) level of performance, and (2) there is much greater dispersion in performance than previously assumed (e.g., Schmidt, Hunter, McKenzie, & Muldrow, 1979). Thus, star performers are a powerful source of competitive advantage for organizations that employ them, because their output is so vastly superior to the output of others. Star performers are highly valuable, sought after, and also the recipients of idiosyncratic job arrangements such as job crafting (i.e., they are allowed greater freedom and autonomy in how they do their jobs), more and better perks (e.g., flexible work schedules), and additional resources (e.g., compensation, rewards, and perks) (Aguinis & Bradley, 2015).

[Insert Figure 1 about here]

Negative Consequences of Gender Performance Gaps among Star Performers

Understanding the gender performance gap specifically regarding star performers is important several reasons. First, star performers are highly influential to the individuals around them, often serving as role models and mentors (e.g., Javidan, Bemmels, Devine, & Dastmalchian, 1995; Lockwood, 2006; Marx and Roman, 2002; Young, Rudman, Buettner, & McLean, 2013). For example, some researchers have suggested that proximity to high-performers indirectly benefits the career advancement of subordinates through enrichment of the latter's social capital (Malhotra and Singh, 2016). Furthermore, such benefits may be greatest when the mentor and subordinate are of the same gender group. To illustrate, Lockwood (2006) found that women were more inspired by outstanding female than male role models, and concluded that individuals (who perceive themselves to be) in a minority group experience greater benefits from mentors of the same minority group. The presence of a same-gender role model has also been shown to impact individual performance; for example, Marx and Roman (2002) found that women's performance on a math test was greater when administered by a competent female experimenter. There is also research suggesting that supervisors and mentors, in turn, provide more psychosocial support to protégés when they are of the same gender (Koberg, Boss, & Goodman, 1998).

The composition of star performers is also influential in shaping the general attitudes held by all other performers. Social comparison theory, which suggests that individuals are most likely to turn to similar others for information about themselves, is aligned with the notion that a gender performance gap among star performers is detrimental to all members of the disadvantaged gender group. A gender gap among star performers is discouraging for all other performers in the disadvantaged group, as the gap implies to them that they have a lower chance of achieving equivalent success in their fields. A disproportionately large gender imbalance among stars in STEM fields could also have broader societal and even economic implications. For example, it may contribute to the perpetuation of negative stereotypes (e.g., "women perform poorly in mathematics"), thus potentially exacerbating issues of gender discrimination. A disproportionate dearth of female stars may also contribute to reduced STEM interest among girls and women, meaning that, from an economic standpoint, more women may miss out on lucrative STEM professions. To illustrate, Olitsky (2014) found that, in general, the decision to pursue a STEM major is associated with an earnings benefit in the range of 5 to 28%.

Finally, in academia, large gender imbalances in general are potentially detrimental to the progress of those fields. Assuming that great leaps in science require the input of individuals with diverse perspectives and experiences, the underrepresentation of an entire gender group in many STEM fields could be a major barrier to scientific progress. Furthermore, star academics include extraordinary scholars whose achievements encompass paradigm-shifting contributions to their research streams. Given the greater influence of star performers (compared to average performers), a large gender imbalance among stars would be more detrimental to the fields' progress than an equivalent gap among average performers.

For these reasons, a gender gap among star performers is arguably more consequential than a gender gap across the entire range of performers. Star performers, compared to others, have greater access to resources; are more capable of motivating, mentoring, and creating opportunities for their protégés; and are more influential in shaping the attitudes held by all other performers and even society at large. Thus, a gender gap among star performers in favor of men in STEM fields would be more damaging to the success of all women performers—and to scientific progress—than an equivalent gender gap among average performers (Ely, 1995).

Purpose of the Study

This research involves three studies whose goal is *to understand gender performance gaps among star performers in STEM fields*. In this research, I conceptualize star performers as individuals who are exceptionally productive compared to the rest of the population (e.g., the top 1%), consistently displaying superior performance over a prolonged period of time. As discussed previously, I theorize that there is a gender performance gap in favor of men across all STEM performers and that the gap is also disproportionately larger among stars. The core premise here is that, in STEM fields, women perform worse than men on average and are also *less variable* in their performance. A lower variance would suggest that women are increasingly underrepresented towards the upper extremes of STEM performance—and thus among stars. A clear illustration of this would be if, among the top 1% of all STEM performers, 90% of them are male despite that men overall make up, say, only 60% of the population of STEM researchers. Furthermore, a lower variance suggests that star women are less variable than star men in terms of performance and are thus outperformed by the best of star men. In short, the primary purpose of this research is to examine the existence of such gender performance gaps and to better understand their potential causes.

To examine the existence of gender performance gaps among stars, I will compare the shapes of the individual performance distributions of male and female researchers in STEM fields. If women's STEM performance is indeed less variable than men's, the performance distribution for women should have a lighter right-hand tail. For example, if STEM performance is normally distributed for both gender groups, yet there is a lighter right tail for women, it would indicate decreasing female representation towards the right tail and thus a gender performance gap in favor of star men. Another example would be if women's performance is normally distributed, but men's performance follows a power law distribution, which inherently has a heavier right tail than the normal distribution. In essence, the existence of gender performance gaps among stars can be identified by comparing the performance distributions (and their right-tail heaviness) across genders.

In this research, I use a methodological process called *distribution pitting* (Joo et al., 2017) to assess and compare the shapes of the performance distributions for men and women in selected STEM domains (i.e., Mathematics, Material Sciences, and Genetics). This technique is derived from theoretical work and conceptual frameworks developed in other fields such as physics, zoology, biology, and computer science to link particular distribution shapes to the generative mechanisms that lead to these particular shapes. Specifically, distribution pitting compares seven theoretical distributions in terms of their fit with the observed data. Additionally, each of the theoretical distributions is associated with one of four possible 'dominant generative mechanisms,' each of which describes a unique set of major underlying processes that shape the distribution.

Distribution pitting is very compatible with the purposes of this research for the following reasons: First, it can be used to estimate various fit indices that reflect the 'right-tail heaviness' of an observed distribution. A comparison of fit indices across genders should, then, reveal information about the existence (and magnitude) of any gender performance gaps among stars. Second, distribution pitting identifies the best-fitting theoretical distribution with respect to the data, thus also revealing its dominant generative mechanism. As such, the technique can be used to generate insight into the most dominant underlying processes that shape individual differences in STEM performance, possibly differently for men and women. An understanding of the dominant processes and mechanisms that shape STEM performance could then lead to better theorization about why men and women in these fields vary the way they do in terms of their performance outcomes. This would also generate insight into the relative effectiveness of interventions aimed at reducing such gender performance gaps.

A defining characteristic of stars is that they exhibit disproportionately high and *prolonged* performance, which makes them distinct from individuals with "fleeting fame or one-time successes," (Call, Nyberg, & Thatcher, 2015). Accordingly, in this research, I focus on performance in terms of one's accumulated performance outputs over a wide timeframe (i.e., the past decade). Additionally, I adopt a results- or output-based definition of individual performance. Traditionally, individual performance has been defined in terms of behaviors (Beck, Beatty, & Sackett, 2014; Campbell, 1990), results (Bernardin & Beatty, 1984; Minbashian & Luppino, 2014), or both (Viswesvaran & Ones, 2000). One benefit of defining performance as results is that of its greater focus on objective measures of performance. Past research on gender-based performance differences has mostly focused on subjective measures of performance rather than objective measures (e.g., Lyness & Heilman, 2006). However, the use of subjective measures to assess gender differences in performance is often inconclusive because any differences found may be attributed to true differences or, alternatively, biases in the rating process (Oppler, Campbell, Pulakos, & Borman, 1992; Pierce, Aguinis, & Adams, 2000; Stauffer & Buckley, 2005). To reflect the results-based view of performance adopted in this research, I measure STEM performance in terms of individuals' cumulative research productivity (i.e., total number of scholarly publications in top STEM journals) over a period of ten years.

Research Questions

In this study, I investigate possible gender performance gaps among star performers in STEM fields. Specifically, my key research questions are as follows:

R1. *Is there a considerable gender performance gap among STEM researchers in general? Do men and women differ in terms of their average STEM performance?*

R2. *Is there an even larger gender performance gap among star performers than among all other performers? Specifically, do the performance distributions for men and women substantially differ in terms of their right-tail heaviness?*

R3. *What is (are) the dominant generative mechanism(s) that shape individual differences in STEM performance?*

R4. *Given the dominant generative mechanism(s) identified,*

- ***a.*** *What can be inferred about the most important underlying processes that lead to stardom in these fields?*
- ***b.*** *What are the likely underlying causes for any observed gender performance gaps?*
- ***c.*** *If men and women differ in their dominant generative mechanisms, what are the likely causes for such a disparity?*

Significance of the Study

The results of this research will have broad impact and numerous implications for research and practice. First, the findings of this research will help lead to a better understanding of the shapes of the performance distributions for women and men in STEM fields. In particular, the results will reveal the existence of any gender performance gaps and also confirm whether an even larger gap exists among stars performers in these fields. Assessment of the particular shapes and distributional properties as well as gender-based differences will also shed insight into the dominant underlying factors that result in particular distributions.

Through comparing the shapes and dominant generative mechanisms of the performance distributions by gender, this research will build on—and possibly refine—existing theoretical perspectives regarding women's general underrepresentation in STEM fields. Currently, there exists much uncertainty regarding the most important cause(s) for the observed gender disparity in STEM fields, and there have been many contradictory findings in the literature (Carli et al., 2016; Ceci, Ginther, Kahn, & Williams, 2014). For example, studies on sex differences in cognitive abilities have found that the gender performance gap on standardized quantitative exams have declined significantly over the past several decades, leading many scholars to question the extent to which sex differences contribute to the observed gender imbalance in high level STEM professions (e.g., Ceci et al., 2014; Ceci & Williams, 2010). Additionally, in contrast to the notion that women are pushed out of STEM fields due to gender discrimination, a recent study suggests that there may now be a 2:1 faculty preference for female (versus male) STEM researchers when hiring (Williams & Ceci, 2015). There is even uncertainty about whether an overall gender performance gap in favor of men exists in these fields. For example, some researchers have suggested that women in STEM fields must be more productive than their male counterparts (i.e., outperform them) in order to survive the winnowing process of less talented women in those fields (e.g., Wenneras and Wold, 1997).

This research aims to make several important contributions to the literature on gender differences in STEM fields. First, this study will generate greater theoretical clarity regarding the direction and magnitude of any gender performance gaps in STEM fields. For example, the confirmation of a gender performance gap in favor of star men would be more aligned with theories on negative (rather than positive) biases toward women in STEM fields. It would also essentially disconfirm the notion that women need to outperform men in order to survive in high level STEM environments. Second, through an examination of dominant generative mechanism(s), this study will generate insights into the most important underlying processes that shape individual differences in performance and stardom in STEM fields, possibly differently for men and women. These new insights could then further extend theory regarding the most important causes for the observed gender performance gap. Third, with its focus on star performers, this study will generate new knowledge regarding the gender disparities at the highest levels of STEM performance.

Additionally, this research may yield unexpected findings that may, in turn, generate important new theories. Hypothetically, the findings may suggest that there is actually a performance gap in favor of star women (rather than star men). Such a finding would suggest the existence of possibly unexplored factors that allow star women to outperform star men, despite the potential constraints on women's STEM performance caused by gender discrimination or gender differences in quantitative aptitudes and/or motivational factors. To illustrate, from the gender discrimination perspective, a performance gap in favor of star women may be an indication that discrimination against women in STEM fields does not actually hurt female performance in STEM fields—at least among stars. It also may reflect positive biases towards star women in STEM fields which benefit their performance. Alternative such a finding may suggest that there is talent gap in favor of women in STEM fields due to, say, men being overvalued (over comparably talented women) in hiring decisions. As such, the findings of this research could lead to substantial extensions to existing theories.

Finally, the results of this research will yield insights of value to not only researchers, but also business practitioners and policy-makers. In particular, a confirmation of disproportionately large gender performance gaps among stars would have powerful implications for practice. The existence of such gaps would indicate that women are not only underrepresented in these fields—they also perform worse than their male counterparts and are less variable in their performance. Given the negative consequences of large gender performance gaps among stars, such findings would call for greater urgency in developing interventions aimed at reducing such gaps. By generating a better understanding of the dominant generative mechanisms that shape men and women's STEM performance, this research will enable recommendations for such interventions. Through diagnosing the underlying mechanisms that lead to possible gender-based differences in the shapes of the performance distributions, this research will generate practical recommendations for reducing the gender performance gap, broadening the participation of women in STEM fields, and advancing the careers of women in these fields.

CHAPTER II

LITERATURE REVIEW

I begin this chapter with a review of the three major perspectives regarding the gender disparity in STEM fields and discuss their implications on a possible gender performance gap among star performers in these fields. Next, I provide an overview of Joo and colleagues' (2017) taxonomy of individual output distributions, which includes seven theoretical distributions grouped into four categories, each with its unique generative mechanism. Discussion of this taxonomy serves to illustrate how an assessment of a distribution's shape can yield meaningful information about the generative processes that create individual differences in output—information that can be compared across multiple distributions. Finally, I discuss the theoretical links between each of the three gender perspectives and the four dominant generative mechanisms in Joo and colleagues' (2017) taxonomy. More specifically, I discuss—from each perspective—how the performance distributions for men and women may differ under each generative mechanism.

Review of the Three Main Gender Perspectives

In my review, I identified three broad explanations for the gender disparity in STEM fields. First, the biological perspective suggests that the underrepresentation of women in STEM fields can be explained mainly by gender differences in quantitative aptitudes. Second, discrimination theories suggest that there are gendered barriers to entry and success in STEM fields. Third, the 'career and lifestyle choices' perspective suggests that the gender disparity is largely a result of the overall choices made women, whether freely made or constrained by social or biological factors. In the sections below, I review the key research findings associated with each of these theoretical explanations.

Sex Differences in Relative Cognitive Aptitudes

Proponents of these theories suggest that men are biologically primed to outperform women in mathematical/spatial tasks, and that male quantitative superiority largely explains the underrepresentation of women in STEM fields (Baron-Cohen, 2003; Halpern et al., 2007). In early research, head/brain size was viewed as an indirect measure of intelligence, and it was theorized that, in general, women have smaller brain sizes and are intellectually inferior. This controversial idea was eventually discarded, however, as it was later revealed that men and women have brains of equal size when controlling for body mass and that many geniuses actually had relatively small brains according to postmortem exanimations (Lynn, 1994). Later research showed that, despite comparable levels of overall intelligence, each gender group is associated with unique strengths and weaknesses in intellectual aptitude across cognitive domains. Specifically, researchers have found consistent support for the conventional belief that boys are quantitatively superior whereas girls are verbally superior (Ceci & Williams, 2010; Halpern et al., 2007; Hedges & Nowell, 1995; Hyde, 1996; Park, Lubinski, & Benbow, 2008; Wai et al., 2010). For example, in their meta-analysis of 286 studies, Voyer, Voyer, and Bryden (1995) concluded that men, on average, significantly outperform women on various measures of visuospatial ability (e.g., mental rotation tasks measuring the ability to imagine how objects would appear when rotated in three-dimensional space). The failsafe value (i.e., number of non-significant and/or unpublished findings that would need to exist to nullify their conclusion) was 178,205.

In more recent studies, researchers have found that, although mean differences between men and women on measures of quantitative performance have closed substantially in the past several decades, women are consistently underrepresented among the top scorers. In particular, researchers have observed greater variance in quantitative abilities among men compared to women, with men displaying greater representation in the extreme left and right tails of the performance distribution. To illustrate, in 1980, Benbow and Stanley showed that the male to female ratio on the quantitative SAT was 2 to 1 the top 0.5% of scorers and 13 to 1 for the top 0.01% of scorers. Subsequent research suggests that while the male to female ratio among top scorers has improved since 1980, there is still an enduring advantage in favor of men. In their examination of SAT and ACT scores of approximately 1.6 million 7th grade students over a period of thirty years (1981-2010), Wai and colleagues (2010) found that the male to female ratio among the top 0.01% of scorers significantly decreased during 1981 to 1990 (from 13.5 to 7.6), but remained relatively stable from 1991 to 2010 and still favors males (ratios ranging from 3.55 to 4.13). Interestingly, these recent ratio estimates (i.e., around 3 to 1 in favor of males) are similar to that of PhD recipients in mathematics: during 2001 to 2010, roughly 30% of mathematics PhDs was awarded to U.S.-born women (Daverman, 2011). In contrast, female students displayed greater representation among top scorers in verbal domains; for example, from 2006 to 2010, the male to female ratio among the top 0.01% of scorers was 0.87 for the SAT verbal and 0.76 for SAT writing. In their follow-up study, Makel, Wai, Peairs, and Putallaz (2016) concluded that, despite shrinkage over the past 20 years, the male to female ratio in the extreme right tail of math ability still favors males. According to Makel and colleagues (2016), the male to female ratio among top verbal performers, in contrast, has been relatively stable during the past two decades and still favors females. They also found similar patterns of male to female ratios using a sample from India.

Researchers have suggested that such sex differences in relative aptitudes across cognitive domains stems from biological differences in brain lateralization (e.g., Baron-Cohen, Knickmeyer, & Belmonte, 2005). Brain lateralization refers to the tendency for certain cognitive and neural functions to be dominant in one hemisphere than the other. This line of research suggests that while women draw upon both hemispheres during spatial tasks, men's spatial abilities mainly reside in the right hemisphere, meaning greater lateralization for the male brain. Greater lateralization is assumed to be more compatible with high spatial performance (Gill & O'Boyle, 1997; Halpern et al., 2007). The different brain lateralization patterns of men and women appear to be the result of differential exposure to certain sex hormones. Specifically, greater exposure to prenatal testosterone and greater activation of postnatal testosterone (i.e., during puberty) has been linked to more optimal brain lateralization (Halpern et al., 2007; Wang & Degol, 2003). In support of this theory, several studies have found that androgen administration for female-to-male transsexuals led to improved performance on visuospatial tasks (see Halpern et al., 2007). To recap, hormonal differences between men and women create differences in cerebral organization which, in turn, contribute to an innate quantitative advantage among men.

Despite these findings, some researchers are suggesting that research on male quantitative superiority is largely inconclusive and that biological differences in aptitude alone do not account for the level of gender disparity observed in STEM fields. For example, Ceci and Williams (2010) explains that, if scoring among the 1% is required for success in math-intensive careers, the ratio of women to men should be around 1 to 2 in these fields. Yet, women's representation in these fields is much lower than what is implied by the 1 to 2 ratio. In the fields of geoscience, engineering, economics, mathematics/computer science, and the physical sciences, for example, only 7% to 16% of full professors were women (Ceci et al., 2014). In addition, despite gender differences in standardized examinations of quantitative ability, female students have been shown to get better grades in mathematics compared to their male peers (Ceci et al., 2014), suggesting that the foregoing tests may be inherently biased towards males. As such, some researchers have concluded that, while sex differences in cognitive abilities remain an important factor, various socio-cultural factors play a significant role in producing the observed underrepresentation of women in STEM fields (Wai et al., 2010). Recently, the debate has shifted more towards the question of whether the gender disparity is primarily caused by gender discrimination or by the career and lifestyle choices made by women (Ceci Williams, 2010; Eagly & Miller, 2016).

Gender Discrimination

Some researchers suggest that discrimination against women is the primary cause of the gender disparity in STEM fields. According to many social psychologists, overt sexism that existed several decades ago in the U.S. has been replaced by unconscious, covert, sexism (see Halpern et al., 2007). This perspective suggests that women experience constant discrimination with respect to STEM participation starting from early childhood and all the way to higher education and professional settings. The bottom line is that prejudice and discrimination directly hinders the entry and success of women in

STEM domains at all stages of their development and also indirectly by creating a dissatisfying environment—often referred to as a 'chilly climate'—towards women that reduces their STEM interest and STEM self-efficacy beliefs.

First, in childhood, girls are often the target of negative bias held by teachers, parents, and peers who believe that girls are not as capable as boys in STEM subjects (Brown & Stone, 2016). According to prior research, many parents and teachers perceive boys to be more logical and mathematically inclined and believe that science is less interesting and more challenging for girls than for boys (Tenenbaum & Leaper, 2003). Such attitudes held by teachers and parents are expressed through subtle behaviors that endorse the stereotype that girls are not as competent as boys in math and science subjects. For example, Brown and Stone (2016) found that mothers are about twice as likely to speak in numerical terms (e.g., "How many feet do you have?") to their sons as to their daughters. Researchers have also found that female teachers' own math anxiety is linked to girls' lower math performance, suggesting that female teachers model these stereotypes through subtle behaviors (Beilock, Gunderson, Ramirez, & Levine, 2010). Accordingly, based on their surveys with over 600 adolescent girls, Leaper and Brown (2008) found that half of the sample reported hearing biased statements about girls in STEM subjects from their peers (both male and female), teachers, and parents. Similarly, Robnett (2015) found that 38% of girls reported hearing disparaging statements about girls' STEM abilities and that 39% of girls felt that they had to work harder than their male peers to be taken seriously in classroom settings.

Given that adults' beliefs about children's abilities, in turn, predict children's self-efficacy beliefs (Tenenbaum & Leaper, 2003), such negative stereotypes about girls can be highly damaging to their STEM self efficacy and STEM interest (Brown & Leaper, 2010). Such stereotypes communicate to girls that "girls are less competent and less interested in math and science subjects," which consequently shifts their academic self-concepts away from STEM subjects (Steffens & Jelenec, 2011) and reduces their STEM interest (Grossman & Porche, 2014). To illustrate, Perez-Felkner, McDonald, Schneider, and Grogan (2012) found that gender differences in STEM orientations emerged as early as by the age of 5, with girls perceiving that "math is for boys." Additionally, Brown and Leaper (2010) found that girls who perceived negative comments about their STEM abilities had lower STEM self-efficacy compared to girls who did not perceive such comments, even when controlling for their grades in math and science subjects.

It has also been suggested that, in some contexts, negative stereotypes about girls/women in academic STEM domains can impair girls' performance on STEM tasks, even in the absence of discriminatory behaviors. Specifically, the theory of *stereotype threats* (Steele & Aronson, 1995) suggests that the activation of negative stereotypes regarding women's STEM abilities (e.g., "Women perform more poorly in mathematics") leads to performance anxiety and ultimately reduced performance of women (Danaher & Crandall, 2008; Picho, Rodriguez, & Finnie, 2013; Spencer, Steele, Quinn, 1999; Stoet & Geary, 2012). For example, Rudman and Phelan (2010) found that priming women with traditional gender roles (e.g., male surgeon and female nurse), increases their implicit gender stereotype beliefs, which in turn, leads to reduced interest in pursuing traditionally masculine occupations. Additionally, Danaher and Crandall (2008) found that girls' pass rates on the AP calculus exam was 6% higher when they had to report their gender at the end of the exam as opposed to the beginning. In practical terms, this means that, every year, roughly 4,763 more girls would enter college with calculus credit if test administrators had students report their gender at the end of the exam as opposed to the before the exam. In sum, perceptions of society's negative beliefs about girls' STEM abilities among girls during their childhood and adolescent years can significantly reduce their interest, self-efficacy beliefs, and performance in STEM domains.

Beyond educational settings, discrimination may impose on women more direct barriers to entry and success in STEM fields. Specifically, findings from many experiments suggest discrimination against women in hiring, journal reviewing, and grant funding (e.g., Trix & Psenka, 2003; Wenneras & Wold, 1997). On the issue of hiring discrimination, in one study where biology, chemistry, and physics professors evaluated applications for a lab manager position, candidates with a female name were less likely to be hired, offered a lower starting salary, offered less mentoring opportunities, and generally perceived as being less competent than candidates with identical application materials but with a male name (Moss-Racusin, Dovidio, Brescoll, Graham, & Handelsman, 2012). Similarly, an experiment by Reuben, Sapienza, and Zingales (2014) showed that, in the absence of information about candidates other than their gender, women were twice as less likely to be hired for a mathematical task (i.e., women were chosen 33.9% of the time). Women fared slightly better (i.e., chosen 39.1% of the time) when employers received objective information about the candidates' past performance, but slightly worse (i.e., 32.0%) when employers received subjective information about the candidates' past performance. According to Reuben and colleagues (2014), employers are also less likely to take into account that men are also more likely than women to 'boast' about their future performance, thus leading to suboptimal hiring choices where men are overvalued while women (with equivalent or even superior abilities) are undervalued.

Of course, being hired onto a tenure track is only the beginning of an academic career, and, researchers have suggested that gender discrimination does not end at the hiring stage. Specifically, research on possible gender discrimination in journal reviewing, grant funding, and remuneration suggest that women are disadvantaged with respect to work-product evaluations. Many researchers have found that women in most science fields tend to publish fewer papers than men overall, and that their publications receive fewer citations (Ceci, Ginther, Kahn, & Williams, 2014; Eagly & miller, 2016). In one study where scholars rated abstracts submitted for a conference, it was found that abstracts with a female name were rated as lower in quality and less appealing for possible collaboration (Knobloch-Westerwick, Glynn, & Huge, 2013). Similarly, in a study of data using economists' CVs, Sarsons (2017) found that, with respect to tenure-ship, women incur penalties for co-authorship that men do not experience, and that these penalties were most pronounced when coauthoring with men. With respect to grant funding, a frequently cited experiment by Wenneras and Wold (1997) showed that women had to be significantly more productive than men in order to receive the same peer review score. According to their study, in order to receive grant funding, women either had to publish at least three more papers in a prestigious science journal or an additional 20 papers in lesser-known specialty journals compared to male competitors. Finally, it has been suggested that women's underrepresentation in STEM fields is exacerbated by gender differences in salaries in these fields. In support of this notion, Ceci and colleagues (2014) found that, specifically among assistant and full professors in economics, assistant professors in life sciences, associate and full professors in engineering and the physical sciences, and full professors in geosciences, men receive significantly greater salaries than their female counterparts.

Interestingly, some researchers suggest that there is a ‘queen bee effect’ that further contributes to the underrepresentation of women in the sciences. This theory suggests that women in male-dominated professions are more likely than their male counterparts to hold negative perceptions of other women in their discipline and are less likely to provide help to female colleagues. For example, Ellemers, Heuvel, Gilder, Maass, and Bonvini (2004) found that the faculty members of two universities perceived female students to be less committed to their work and that the female faculty members held these perceptions most strongly. Similarly, workplace studies have demonstrated that, compared to men, women are less supportive of the advancement of other women (especially as they advance in the organization) and are more biased in their perceptions of female colleagues (Ellemers et al., 2004; Garcia-Retamero & Lopez-Zafra, 2006; Ng & Chiu, 2001). Overall, such biases against women held by other women—and heterophily effects among women in general— could be severely damaging for the participation of women and the emergence of female star performers in STEM fields, especially given research showing that young women are more likely to persist in a STEM field when mentored by other women (e.g., Price, 2010).

In sum, the foregoing studies suggest that gender discrimination in hiring, pay, and work-product evaluations impose significant ‘floor’ and ‘ceiling’ effects that hinder the entry and success of women in STEM fields. However, the gender discrimination perspective has received some criticism in recent research. Based on their review of research on gender discrimination in these fields, Ceci and Williams (2011) concluded that many of these studies have considerable methodological flaws. They explained that, after controlling for several important confounding variables, the effects of discrimination on peer reviews, hiring, and grant funding were minimal or non-existent. In particular, there have been many contradictory findings regarding gender discrimination in journal reviewing. For example, Borsuk and colleagues (2009) found evidence of gender neutrality in acceptance rates when identical manuscripts with either female or male names were sent to the same reviewers. In addition, Gilbert, Williams and Lundberg found that, in 3,400 publication recommendations submitted to the Journal of the American Medical Association, there was no association between author gender and acceptance.

With respect to hiring decisions, Ceci and colleagues (2014) argue that lab experiments demonstrating negative female bias in hiring may not be reflective of actual hiring patterns of tenure-track professors in STEM fields. In particular, they explain that while academic hiring decisions involve large committees of researchers evaluating other researchers, a large proportion of the foregoing lab experiments dealt with either (1) individual biases against female undergraduates in STEM positions such as lab managers or (2) individual biases among undergraduates evaluating the work of researchers, thus limiting the external validity of their findings. Accordingly, Ceci and colleagues (2014) concluded that, while they do not oppose the view that stereotyping and gender biases can affect hiring decisions, the extent to which they can be generalized to real-life hiring decisions for professorial jobs in STEM fields is questionable.

Furthermore, Ceci and colleagues (2014) found that female tenure-track applicants are equally, or in some cases, even more successful than their male competitors in heavily male-dominated STEM fields. For example, they found that, in the field of mathematics, “out of the 96 hires at the assistant-professor tenure-track level from 1995 through 2003, only 20% of applicants for these positions were female, but 28% of those invited to interview were female, as were 32% of those offered tenure-track positions,” (Ceci et al., 2014, p. 101). In a follow-up study, Willams and Ceci (2015) conducted randomized experiments on over 800 tenure-track faculty in biology, engineering, economics, and psychology at 371 U.S. universities, and concluded that there is, in fact, a 2:1 preference for women. They also found that that the same preference for women was held by faculty of both genders.

The notion preference women in terms of hiring may sound bizarre, given women’s underrepresentation in STEM fields. However, it is consistent with past research suggesting that people often display positive/leniency bias toward minority members (e.g., Harber, 1998). It is also consistent with the increasing emergence of EEOC/affirmative action issues as if late. As noted by Ceci and colleagues (2014), the observed preference for women in hiring decisions does not necessarily reflect a female advantage in these fields. For example, it is possible that women are seemingly preferred in tenure-track hiring decisions due to female applicants being superior to male applicants. On average, female applicants may be more qualified than their male competitors due to having survived a prior winnowing of less qualified women in these domains. Additional research is thus required to better understand the ways in which gender discrimination and biases impact the performance of women (versus men) in contemporary professional STEM settings.

Women's Career and Lifestyle Choices

Scholars have long noted that women tend to opt out of STEM fields at higher rates than men during all stages of their careers. According Ceci and Williams (2010), the primary cause of the gender disparity in STEM fields has to do with the 'choices made by women,' whether freely made or constrained by biology or society. In this body of research, differences in motivational/psychological factors, such as occupational preferences, lifestyle values, goals and interest in STEM work have been theorized to be key contributors to the gender disparity (Ceci & Williams, 2010; Ferriman, Lubinski, & Benbow, 2009; Morgan, Gelbgiser, & Weeden, 2013; Schwartz & Rubel, 2005). These theories suggest that men and women differ in their career and lifestyle choices, which ultimately result in fewer women choosing to enter and persist in STEM fields. With regards to explaining the observed gender disparity, this view suggests that women's overall choices outweigh the impacts of women's possible disadvantages in STEM abilities (due to sex differences) and of any structural barriers to success (due to gender discrimination).

According to Ceci and Williams (2010), women's underrepresentation in STEM fields is largely a reflection of their particular choices, including those that are constrained by socio-cultural factors. To illustrate, one motivational domain where men and women tend to significantly diverge is that of their STEM interest and attitudes during their student years. For example, it has been found that by middle school, boys are twice as likely as girls to expect a career in STEM-related fields (9.5% vs. 4.1%; Legewie

& Diprete, 2012). Similarly, Xie and Shauman (2003) found that, among high school seniors, the proportion of female students expecting to major in a STEM subject in college was less than a third of that of male students. Such gender differences in early STEM interest and attitudes are thought to be at least partly the result of various socio-cultural and contextual factors, such as early classroom experiences and interactions, the relative lack of female scientists and engineers as role models, and certain societal and cultural norms that, together, promote perceptions of misalignment between STEM work and traditional gender roles (Blickenstaff, 2015; Ceci et al., 2014; Kossek et al., 2017; Wang & Degol, 2013). In particular, female students who hold negative stereotypes about women's STEM abilities and/or about STEM work in general (e.g., "science is a masculine field") develop reduced STEM interest and consequently are more likely to choose careers outside of STEM fields (Wang & Degol, 2013).

Women's relatively greater tendency to choose non STEM-related careers is also thought to be at least partly informed by biological factors. For example, research on sex differences in relative cognitive aptitudes have found that women are more likely than men to have both high math and high verbal abilities (Halpern et al., 2007). Specifically, it has been found that, among men and women with comparably high quantitative abilities, women are more likely to have greater verbal ability and are thus more balanced in their ability profiles (Park et al., 2008; Wang, Eccles, & Kenny, 2013). It has thus been theorized that women's greater balance in cognitive abilities provide them with more career options (Wang & Degol, 2013). Furthermore, given the widespread perception that STEM fields are characterized as having a 'chilly climate' towards female scientists, women who are equally talented in quantitative and verbal domains are more likely to be drawn to careers outside of STEM, expecting greater success in domains with reputably more positive reception towards women. To use an economic metaphor, many women perceive themselves as having a ‘comparative advantage’ in non-STEM related domains.

It has also been suggested that women opt out of STEM fields at disproportionate rates due to perceived misalignment between STEM work and their occupational preferences (Kossek et al., 2017; Wang & Degol, 2013). According to prior research, women are more likely to prefer ‘working with people rather than objects,’ which clashes with traditional STEM environments that are often described as impersonal and individualistic (Diekman, Brown, Johnston, & Clark, 2010; Su, Rounds, Armstrong, 2009). There is also research suggesting that, compared to men, women tend to be more oriented towards communal and altruistic goals that place emphasis on helping others and benefiting society (Diekman, Weisgram, & Belanger, 2015; Schwartz & Rubel, 2005). In contrast, men tend to have a more agentic/power-based orientation and may be more likely to pursue jobs that allow them to make lots of money, power, and fame. Relative to non-STEM careers, many STEM careers are thought to be more compatible with agentic rather than communal goals (Diekman, Clark, Johnston, Brown, & Steinberg, 2011). As such, women have been found to perceive less fit between communal goals and most STEM occupations compared to with occupations such as law, medicine, nursing, or teaching (Diekman et al., 2010). Even within STEM fields, more women are drawn to domains such as biomedical and environmental engineering than, say, mechanical or nuclear engineering (Gibbons, 2009).

Similarly, work-life and work-family balance has been shown to be an important factor in shaping the career decisions of women (Hill, Corbett, & St. Rose, 2010).

Women tend to desire a job that allows them to spend more time with family, and, relative to men, are more likely to make occupational sacrifices in pursuit of family goals and responsibilities (Diekman et al., 2015). In one study of highly talented men and women, it was also found that women were less likely than men to want to work long hours, i.e., 50+ hours in their current job or 60+ hours in their ideal job (Robertson, Smeets, Lubinski, & Benbow, 2010). It has thus been suggested that many STEM occupations (especially those that are highly research-intensive and competitive) clash with the work/life and work/family values of women. Furthermore, maternity makes it especially difficult for women in STEM to work long hours and achieve the same level of productivity as their male colleagues (Ceci & Williams 2011; Diekman et al., 2015; Wang & Degol, 2013). In academia, especially, a large amount of one's time is dedicated to graduate, postdoctoral, and junior faculty work, and these years tend to coincide with women's years of greatest fertility. It has thus been suggested that the decision to have children early in one's career creates productivity losses that are disproportionately greater for women than form men (Diekman et al., 2015). For these reasons, it has been suggested that relatively fewer women choose to pursue STEM careers, and those who do may work (or desire to work) fewer hours than their male colleagues (Ferriman, Lubinski, & Benbow, 2009; Lubinski & Benbow, 2006). As such, women's work-family and fertility decisions may also explain much of the observed gender differences in the number of publications by professors in many STEM fields (Ceci et al., 2014).

In summary, research on the underrepresentation of women in STEM fields has identified a myriad of biological, socio-cultural/contextual, and psychological factors as potential causes. In the literature, three broad mechanisms have been theorized as being major drivers of the observed gender disparity in these domains. First, the ‘sex differences’ perspective suggests that men are more variable than women in their quantitative abilities. This means that men are better represented at the upper tail of the innate quantitative ability distribution and thus have greater success in STEM domains. The second explanation is that women are pushed out of STEM fields due to discrimination and negative stereotypes. It has been theorized that negative stereotypes about women’s STEM abilities reduces female students’ STEM interest and STEM self efficacy. Moreover, women may face gendered barriers to entry and success in STEM fields due to possible discrimination in hiring and work-product evaluations. Finally, the ‘career and lifestyle choices’ perspective emphasizes that women *choose* to opt out of STEM fields rather than being pushed out due to discrimination or innate differences in quantitative abilities.

Despite the wealth of findings in this body of research, however, there remain uncertainties regarding the extent to which the observed gender imbalance in STEM fields is driven by the different factors representing each gender perspective. In particular, there have been contradictory findings regarding the severity of gender discrimination in hiring and journal reviewing. Similarly, while many studies have found persistent sex differences in the upper tail of the performance distribution for quantitative tasks, there are disagreements in the literature about the explanatory value of the ‘sex differences’ perspective. As such, while sex differences in cognitive aptitudes and gender discrimination issues continue to remain important factors in the debate, there has been increased emphasis on women’s particular career and lifestyle choices as being the primary explanation.

The existence of gender performance gaps in STEM fields—particularly among stars—has not been systematically examined to date. Nonetheless, each of these perspectives alludes to the existence of large gender performance gaps in STEM fields in favor of men. First, assuming that quantitative aptitudes is an important driver of STEM performance, the notion of sex differences in quantitative aptitudes strongly suggests that there may be a considerable gender performance gap among STEM researchers. Second, despite some contradictory findings, much of the research on gender discrimination suggests that women's STEM performance is constrained due to various forms of discrimination that hinder their success once they have opted in to a STEM career. For example, gender discrimination in promotions, grant funding, and journal reviews, may all contribute to lower overall STEM performance of women. Third, the career and lifestyle choices perspective suggests that there is a greater 'leakage' of talented women compared to similarly talented men, which in turn suggests the existence of a gender performance gap in favor of men.

In addition, each of these perspectives alludes to the existence of disproportionately large gender performance gaps among star performers. First, research on sex differences has shown that men are more variable in their quantitative aptitudes and, accordingly, have greater representation (compared to women) at the upper extremes of the ability distribution. Assuming, again, that quantitative aptitudes is an important driver of STEM performance, men may thus be similarly more variable in their STEM performance, suggesting that the gender performance gap would be greater among star performers than among average performers. In other words, due to sex differences in the variability of quantitative aptitudes, women may be underrepresented among stars even after accounting for their overall underrepresentation, and, the best star women may be outperformed by the best star men.

Second, from the gender discrimination perspective, it is possible that the performance-constraining effects of gender discrimination may be most pronounced among stars than among average performers. For researchers in STEM fields, maintaining high levels of research output requires gaining access more resources (e.g., funding) as well as more frequent work evaluations (e.g., peer reviews)—both of these steps represent areas where significant gender discrimination may occur. Accordingly, the cumulative negative effects of gender discrimination may be greatest (in absolute terms) among stars than among average performers. It should be noted, however, that this argument lies in the assumption that gender discrimination consistently undermines women's performance at potentially every opportunity to perform, a notion that has not been examined in prior research. It is possible that negative biases and stereotypes about women in STEM fields operate in complex ways such that, contrary to the foregoing assumption, gender discrimination is somehow minimal or nonexistent among star performers specifically.

Finally, from the perspective of women's career and lifestyle choices, women in STEM fields may choose to devote less time and energy to their work compared to men due to gender differences in work goals, work/family values, and other motivational factors. In particular, men and women may devote relatively similar amounts of resources to their work on average, but there may be a significant gender gap among those who devote the most amounts of time and other resources to their work. In other words, the small proportion of individuals who are committed to putting in extraordinary amounts of effort and the kind of long hours necessary for star performance may be lower among women compared to men. Essentially, women in STEM fields may be less variable with regard to performance inputs, thus producing a gender performance gap in performance outputs that is disproportionately larger among star (and other above average) performers.

Output Distributions and Generative Mechanisms

As explained above, there are numerous factors that may contribute to a gender gap among star performers. To investigate the differences in the shape of performance distributions of men versus women, this research will use a methodological approach that is novel in organizational research, but has been used in the field of physics (Clauset, Shalizi, & Newman, 2009). The first step in the analysis will include determining for each gender-based distribution of the best-fitting model among the following theoretical distributions: (a) normal, (a) exponential, (c) lognormal, (d) pure power law, (e) power law with exponential cutoff, (f) Weibull and (g) Poisson. Table 1 offers a summary of the theoretical framework involving the seven types of distributions and each of the underlying generative mechanisms. Table 1 also includes technical information about each distribution and illustrative empirical research from fields outside of organizational research that has documented the presence of each distribution and the generative mechanism for each.

Self-Organized Criticality

Empirical research conducted in fields other than organizational science such as biology, zoology, computer science, and physics has revealed that each of the distributions included in Table 1 is produced by a one of four different generative mechanisms (Joo et al., 2017). First, presence of a power law distribution (which has the longest tail) is indicative of a complex generative mechanism which Joo and colleagues (2017) refer to as *self-organized criticality*, i.e., complex interactions involving various components which results in *output shocks* that lead to large and unpredictable increases future output. For example, the result that firm growth follows a power law distribution was explained by sustained and complex interactions among numerous resources and variables, including a firm's technology, human capital, social capital, financial capital, self-efficacy, motivation, and environmental characteristics (Axtell, 2001). Compared to other distributions with different dominant generative mechanism, the pure power law distribution has the heaviest right tail, meaning that there is great—potentially infinite—variability among top performers.

In the context of individual performance, the self-organized criticality mechanism suggests that a small proportion of individuals experience unpredictable and potentially very large output shocks after reaching a critical state. According to Joo and colleagues (2017), critical states occur when components accumulated by an individual interconnect. Then, even a seemingly trivial set of events affecting one component may in turn impact all other interconnected components. For example, a scientist's single breakthrough on one project may lead to more breakthroughs in other intricately related projects and thus lead to an explosive growth in his/her subsequent productivity. In this sense, self-organized criticality involves a significant element of randomness and luck in addition to extraordinary ability and persistence. Thus, it generally takes a long time to reach a critical state, and most individuals never reach it in their lifetime (Joo et al., 2017). However, after individuals reach such critical states, even a trivial event may cause unpredictably large output shocks.

Proportionate Differentiation

Second, lognormal distributions result from the generative process of *proportionate differentiation*, which is a process where an individual's future output is a distinct percentage of his/her prior output (Joo et al., 2017). This generative mechanism implies that one's initial (or prior) output is positively linked to future output through interactions with his/her accumulation rate. To clarify, 'accumulation rate' refers to the average amount of a variable that an individual produces per time period (e.g., sales generated per month), whereas 'initial value' refers to "the amount of a variable that each individual has accumulated during a relatively short period of time (e.g., 1 year) since the beginning of a common baseline (e.g., since the first date of employment for all individuals hired in the same year)" (Joo et al., 2017, p. 9). Proportionate differentiation suggests that individuals' initial value on an outcome has a multiplicative relationship with their accumulation rates in determining their future amounts on the same outcome. Joo and colleagues (2017) use an example involving the number of signatures collected by community organizers to illustrate the proportionate differentiation mechanism. To illustrate, one group of organizers may initially acquire a larger number of signatures than the others (i.e., larger initial output), and, as a result, find it easier to obtain additional signatures from people who are more willing endorse something that others already have. If these organizers are simultaneously able to obtain signatures more quickly (i.e., higher accumulation rate), then they would be able to differentiate their accumulation of signatures by increasingly larger amounts—this suggests a positive feedback loop between the number of signatures obtained so far and additional signatures obtained. Out of the seven distributions in Joo and colleagues (2017) taxonomy, the lognormal distribution has the second heaviest right tail, next to the pure power law distribution.

The proportionate differentiation mechanism suggests that individual variation in cumulative outputs is caused by some individuals being able to achieve large *output loops* based on a positive feedback mechanism. In other words, individuals diverge in their cumulative outputs due to significant differences in their initial outputs and/or accumulation rates. As another example, assume there are two researchers, John and Sally, who are comparably talented, but John starts his/her tenure-track career with five publications in top tier journals whereas Sally starts with one. Here, John may find it easier to produce subsequent publications due to, say, greater researcher visibility and more opportunities for collaboration as a result of starting his academic career with more publications. Thus, even if both researchers are comparably talented and put in more-or-less the same amount of work, Sally would not be able to catch up John in terms of their number of publications, unless she is able to increase her accumulation rate (through investing more time, effort, resources, etc) enough to eventually offset the impacts of John's greater initial value.

Incremental differentiation

Third, the exponential distribution and the pure power law with an exponential cutoff distribution result from the generative mechanism referred to as *incremental differentiation*, which suggests that each individual's output increases at an approximately linear rate based on his/her accumulation rate (Joo et al., 2017). Here, unlike proportionate differentiation, prior output is not linked to future output through a positive feedback mechanism, and future value is simply a function of individuals' accumulation rates. For example, research in economics has documented that people's wages accumulate at different linear rates as a result of heterogeneity in labor productivity across individuals, leading to an exponential distribution of cumulative wages (Nirei & Souma, 2007). Also, with the incremental differentiation mechanism, individuals with the highest accumulation rates may experience diminishing returns over time—this potential attribute is reflected in the power law with exponential cutoff distribution. To illustrate this point, Joo and colleagues (2017) use a network science example involving nodes (e.g., airports) and the acquisition of additional links with other nodes (e.g., new departures to other airports). They explain that, as an airport reaches full capacity, the costs associated with acquiring additional links can increase dramatically, meaning that airports that accumulate links more quickly than others (i.e., those with a higher accumulation rate) will eventually be subject to diminishing returns, thus resulting in a power law with an exponential cutoff distribution of links per node (Amaral et al., 2000).

In terms of individual performance, the incremental differentiation mechanism would suggest that individual differences in performance are largely shaped by differences in accumulation rates (i.e., output achieved per opportunity to perform). Top performers with the highest accumulation rates enjoy larger *output increments* (i.e., linear increases in cumulative output) than others; however, depending on the particular case, these individuals may be subject to diminishing returns on their outputs. As another example, research on individual performance has noted that cognitive ability is linearly related to individual performance; additionally, the personality trait of conscientiousness is subject to diminishing returns in terms of their positive effects on individual performance (Pierce & Aguinis, 2013; Whetzel, McDaniel, Yost, & Kim, 2010). Hypothetically, if research productivity was strictly a function of cognitive ability and/or conscientiousness, it would be very likely that individual variation in research productivity is shaped by incremental differentiation.

Homogenization

Finally, the three symmetric (or nearly symmetric) distributions, i.e., normal, Weibull, and Poisson distributions, result from the generative mechanism of *homogenization*. Unlike the other generative mechanisms (i.e., self-organized criticality, proportionate differentiation, and incremental differentiation), homogenization *reduces* individual variability in outputs over time. In zoology, for example, the homogenization processes of various species are characterized by a normal distribution (Spear & Chown, 2008). In the context of individual performance, this generative mechanism suggests that individuals are subject to *output homogenization* processes that act as a 'floor' and 'ceiling' to future output differences (Joo et al., 2017). For example, uniform expectations of production or service tend to homogenize individual workers' outputs (e.g., Groshen, 1991). As another example, promotion policies involving the termination of tenure-track researchers who fail to produce a certain high number of publications act as a 'floor' that limits researchers' variability in the number of publications (Joo et al., 2017).

As discussed so far, each of the four dominant generative mechanisms in Joo and colleagues' (2017) taxonomy describes unique underlying processes that may shape individual differences in cumulative outputs. In this research, I will examine the individual performance distributions of male and female researchers in STEM fields, specifically the shapes of the distributions and their associated dominant generative mechanisms. A by-gender comparison of dominant generative mechanisms could reveal useful new insights into the most important underlying mechanisms and processes that shape individual research performance/productivity in STEM fields, possibly differentially across gender groups. This assessment could thus lead to an enhanced understanding of the most critical issues that create gender performance gaps among in STEM fields, especially in the upper tail of the performance distribution. These insights should then help to inform practical interventions aimed to resolve those issues. In the section below, I discuss the connections between the main theoretical perspectives regarding the cause of the gender disparity in STEM fields and each of the four generative mechanisms. In particular, I discuss the theoretical implications for each the possible gender-to- generative mechanism combinations (e.g., self-organized criticality for men versus homogenization for women) that could be found in this study.

[Insert Table 1 about here]

The Implications of Each Gender Perspective on the Four Generative Mechanisms

In this section, I build theory on the connections between each of the three gender perspectives and the four possible dominant generative mechanisms (i.e., self-organized criticality, proportionate differentiation, incremental differentiation, and homogenization). In particular, I discuss implications of each gender perspective with respect to how the performance distributions for men and women may differ under a given dominant generative mechanism. In other words, I discuss how the STEM performance of men and women may differ—particularly for stars—under each generative mechanism, given what is known about sex differences quantitative aptitudes, gender discrimination in STEM fields, and women's career and lifestyle choices. Additionally, I speculate about the implications of an unexpected absence of gender performance gaps (in favor of men) under each generative mechanism.

Sex Differences and the Four Generative Mechanisms

To recap, the sex differences perspective suggests that men and women differ on their relative cognitive aptitudes. Comparatively, women tend to be more balanced in their quantitative and verbal abilities. Women with high quantitative abilities are thus likely more balanced in their ability profiles and thus have more career options outside of STEM fields. On the other hand, men have been found to be more variable than women, specifically in terms of their quantitative aptitudes and are thus disproportionately represented towards the upper extreme (i.e., right tail) of the quantitative ability distribution. In short, this perspective suggests that women are underrepresented in STEM fields because (1) relatively fewer women possess high quantitative aptitudes compatible with success in STEM domains, and (2) women who do are more likely than their male counterparts to pursue careers outside of STEM fields.

The implications of the sex differences perspective on each of the four generative mechanisms are summarized in Table 2. First, if self-organized criticality is the dominant generative mechanism driving individual differences in STEM performance, the distribution for women may have a lighter right tail due to a strong connection between one's quantitative aptitudes and his/her capacity to achieve self-organized criticality in STEM fields. Reaching a critical state involves the interconnection of many performance components (e.g., interrelated projects), and the interaction of many events and variables—innate quantitative aptitudes could certainly be one of these factors. Thus, given women's lower variability in quantitative aptitudes, it is possible that they are disproportionately underrepresented among STEM researchers who achieve self-organized criticality. If it turns out that no such performance gaps (in favor of star men) exist, it may be an indication that individual quantitative aptitudes do not strongly reflect their capacity to achieve self-organized criticality. It is possible that, beyond a certain requisite level, slight variations in quantitative aptitudes have little impact on whether STEM researchers are able to reach a critical state (and subsequent output shocks), while other abilities (e.g., multi-tasking), behaviors (e.g., choosing the 'right projects' and collaborators), and randomness (i.e., luck) become increasingly important in enabling self-organized criticality.

Second, if proportionate differentiation is the dominant generative mechanism driving individual differences in STEM performance, the distribution for women may similarly have a lighter right tail. Proportionate differentiation occurs when individuals' prior outputs are linked to future outputs through positive feedback (i.e., a multiplicative interaction) with their accumulation rates. In short, proportionate differentiation enables individuals with high initial outputs and accumulations rates to enjoy large output loops. In academia, such output loops may occur due to the presence of various benefits associated with having a prolific pipeline of prior research outputs. For example, a high level of accumulated prior research outputs would lead to greater researcher visibility and greater access to various research-related resources, thus leading to even greater future accumulation of research outputs. Additionally, certain behaviors focused on 'leveraging' prior outputs (e.g., self-promotion behaviors, forming new ties with past collaborators' network connections) may also contribute to the presence of such positive feedback mechanisms. From a sex differences perspective, there may be gender disparities in STEM researchers' accumulation rates (e.g., average research productivity per year) and initial outputs (e.g., number of publications at the start of one's research career). Specifically, if quantitative aptitudes are an important driver of STEM researchers' accumulation rates and/or initial outputs, men would be more variable in terms of their STEM performance, suggesting the existence of a large gender performance gap in favor of star men. If it turns out that no such performance gaps exist, it would suggest that STEM researchers' accumulation rates and/or initial outputs are shaped largely by factors other than quantitative aptitudes.

Third, if incremental differentiation is dominant generative mechanism driving individual differences in STEM performance, the distribution for women may, again, have a lighter right tail from the sex differences perspective. Incremental differentiation suggests that the overriding driver of individual variation in performance has to do with individual differences in accumulation rates. Unlike proportionate differentiation which relies on a positive interaction between accumulation rates and initial/prior outputs, incremental differentiation is determined solely through accumulation rates. So, assuming, again, that quantitative aptitudes are linked to STEM researchers' accumulation rates, performance should be less variable for women and there should be gender performance gap in favor of star men. If it turns out that no such performance gaps exist, it would suggest that STEM researchers' accumulation rates—particularly among stars—may be shaped largely by factors other than quantitative aptitudes.

Finally, the homogenization mechanism suggests that individuals undergo output homogenization overt time, and that variability in past individual output would be followed by lower variability in future individual output. Accordingly, this dominant generative mechanism would suggest the existence of processes that homogenize the research productivity of STEM researchers over time. The sex differences perspective suggests that men and women vary in terms of their innate quantitative aptitudes, but it does not make inferences about such output homogenization processes. Women may be more homogenized than men in terms of their innate STEM aptitudes (due to lower variability in aptitudes in the upper tail), but this does not imply that they are subject to output homogenization processes that do not simultaneously apply to men. In academia, situational factors can act as a 'floors' or 'ceilings' with respect to individuals' future output. One example would be that of promotion policies that require researchers to achieve a requisite number of publications in order to receive tenure, thus limiting the variance in individuals' research productivity levels over time. Such situational output homogenization mechanisms should be faced by both male and female researchers regardless of the differences in their quantitative aptitudes. Nonetheless, if individual differences in quantitative aptitudes significantly drive the variation in individuals' STEM performance (i.e. any variation that exists in spite of the output homogenization processes), the performance distribution for women would have a lighter right tail, and there should be a gender performance gap in favor of star men. In other words, even if women are not subject to output homogenization processes to a greater (or lesser) extent compared to men, women's STEM performance may be more homogenized at any given point in time due to sex differences in quantitative aptitudes.

[Insert Table 2 about here]

Gender Discrimination and the Four Generative Mechanisms

The gender discrimination perspective suggests that women are pushed out of STEM fields due to barriers imposed by gender discrimination and negative stereotypes about women's STEM abilities. The implications of gender discrimination on the four dominant generative mechanisms are summarized in Table 3. First, the gender discrimination perspective suggests that the self-criticality mechanism may be relatively weaker for women than for men. Again, reaching a critical state involves a complex interaction of many events and variables. It is possible gender discrimination constrains women on many variables important for enabling self-organized criticality. For researchers, peer evaluations of one's past work, access to research-related resources, and availability of collaboration opportunities are all likely important variables with respect to enabling self-organized criticality. It is thus possible that gender discrimination in work-product evaluations (i.e., undervaluing women's past achievements, giving women less credit for co-authored work, etc.), journal reviews, and grant funding substantially constrain talented female researchers' potential to achieve large output shocks. In short, female STEM researchers may find it more difficult to achieve critical states due to the various barriers imposed by gender discrimination in these fields. Accordingly, the performance distribution for women would have a lighter right tail, and there would be a gender performance gap favor of star men. If it turns out that there is actually a gender performance gap in favor of women, it may suggest the possibility of positive biases toward star women that benefit their performance.

Second, assume that incremental differentiation is the dominant generative mechanism driving individual differences in STEM performance. From the gender discrimination perspective, it is possible that, women's accumulation rates are lower than men’s, due to the effects of various forms discrimination that constrain women’s performance. Gender discrimination in areas such as peer reviews and research support, may lead to lower accumulation rates for star women compared to star men in these fields. Additionally, the initial outputs for women (e.g., number of publications upon entering a tenure-track position) may be smaller due to, say, the relative dearth of female mentors in academic STEM fields. As such, women may have lower accumulation rates and/or initial outputs compared to men in these fields. Moreover, substantial gender differences in accumulation rates and/or initial output values would suggest that, over time, it would become increasingly difficult for women to ‘catch up’ to their male colleagues in terms of cumulative performance, as proportionate differentiation is based on the positive feedback between accumulation rates and initial/prior outputs. Thus, assuming that gender discrimination constrains the productivity of all female performers (including stars), there should be a gender performance gap among in favor of star men. It is even possible that there be a less powerful feedback mechanism between women’s prior outputs and accumulation rates. For example, talented female researchers may experience smaller output loops than their comparably talented male colleagues due to receiving less credit/recognition for co-authored work, resulting in fewer opportunities for advancement. On the other hand, an absence of such gender performance gaps would suggest that, despite any gender discrimination that exists in those fields, men do not have an advantage in terms of their initial outputs, accumulation rates, and/or the strength of the feedback mechanism between the two.

Third, assume that incremental differentiation is the dominant generative mechanism driving individual differences in STEM performance. If gender discrimination constrains women’s accumulation rates in STEM fields, there should be a gender performance gap in favor of men. Furthermore, it possible that gender discrimination has a greater (negative) net effect on the performance of star women than among average female performers. To wit, the negative effects of gender discrimination on (women’s) performance may accumulate at a greater rate among the most productive individuals. In simplest terms, assume that, given research papers of comparable quality, journal reviewers are twice as likely to accept papers that were authored by men. Now, take male researcher W and female researcher X, both of whom produce 10 manuscripts (of comparable quality) per year. W should then have double the acceptance rate compared to X, say, 8 versus 4. In comparison, there would be a greater gap between male researcher Y and female researcher Z, both of whom produces 100 manuscripts per year, with an average acceptance rate of 80 versus 40, respectively. Accordingly, in a case where gender discrimination constrains the individual STEM performance of all women more-or-less consistently at every opportunity to perform, there may be a disproportionately large gender performance gap among stars in favor of men. If it turns out that there are no such performance gaps, it suggests that gender discrimination in STEM fields does not significantly constrain the performance of women, at least among star performers.

Finally, the gender discrimination perspective provides several implications regarding possible differences in men and women’s performance distributions under the homogenization mechanism. One possibility is that discrimination in the form of undervaluing women (or overvaluing men) in hiring decisions may contribute to a stronger performance floor for women (or a lower floor for men). Accordingly, there may be a larger representation of men towards the lower end (i.e., left tail) of the performance distribution, and women may tend to outperform men, at least among average and/or below-average performers. Alternatively, it is possible that the cumulative negative effects of gender discrimination on women's STEM performance are greatest among the most productive individuals. Accordingly, there may be a lower STEM performance ceiling for women compared to men, suggesting that there would be a disproportionate gender gap among stars in favor of men. If it turns out that there is actually a gender performance gap in favor of star women, the finding may be an indication of unique positive biases and forms of preferential treatment that lead to greater STEM performance and less output homogenization specifically among star women relative to star men.

[Insert Table 3 about here]

Women's Career and Lifestyle Choices and the Four Generative Mechanisms

Researchers have also suggested that women's underrepresentation in STEM fields is largely a reflection of their overall career and lifestyle choices. According to this perspective, due to diverging STEM interests, occupational preferences, and work-life/work-family values, women are disproportionately more likely than men to opt out of STEM fields at all stages of their careers and are also more likely to make career/productivity sacrifices in pursuit of other goals. The implications of women's career and lifestyle choices on the four generative mechanisms are summarized in Table 4.

First, the self-organized criticality mechanism may be weaker for women due to gender disparities in various motivational aspects. As explained previously, reaching a critical state generally takes a long time and involves a complex interaction of many variables and events. In particular, motivational factors such as the amount of time, effort, and other resources devoted to one's projects could be critical in terms of enabling self-organized criticality. So, given that women are more likely to opt out of STEM fields and to make career sacrifices in pursuit of other goals, it is possible that fewer female researchers end up reaching critical states in their STEM careers compared to male researchers. Additionally, certain life decisions (e.g., the decision to have children early in one's career) may make it more challenging for women to reach critical states. Also, compared to star men, star women may also take longer to reach critical states and experience relatively smaller output shocks over the course of their STEM career. Accordingly, the performance distribution for women could have a lighter right tail. If it turns out that there is no such performance gap in favor of star men, it would suggest that the pattern of work and various lifestyle choices made by men and women in these fields may not be so different, at least in terms of their impacts on one's capacity to achieve self-organized criticality.

Second, the career and lifestyle choices perspective suggests that the performance distribution for women may have a lighter right tail under the dominant generative mechanism of proportionate differentiation. To illustrate, it is possible that gender differences in various motivational factors (e.g., work preferences and goals, work/family values, etc) result in significant gender disparities in accumulation rates (e.g., number publications produced per year) and/or initial output values (e.g., number of publications at the beginning of their tenure-track career). For example, women may be less likely than men to devote extremely long hours to their work, despite putting in more-or-less the same amount on average. Thus, from a motivational perspective, it is possible that women have less variability in their accumulation rates and/or initial output values. Even if the gender groups do not differ much in terms of their mean accumulation rates and initial outputs, a disproportionate concentration of men among the small proportion of individuals with the greatest accumulation rates (and initial outputs) would suggest the existence of a gender performance gap among star performers in favor of men.

Third, for the same reasons, the performance distribution for women may have a lighter right tail under the dominant generative mechanism of incremental differentiation As explained above, women may have less variation than men in their accumulation rates due to, say, gender differences in the amount of resources devoted to one's work. One additional cause for women's possibly smaller variance in accumulation rates has to do with the 'leakage' of talented women in STEM fields: talented women may be more likely to opt out of STEM fields than their comparably talented male counterparts due to being more balanced in their cognitive ability profiles and hence having more career options. Another possibility is that star women experience disproportionate losses in research productivity due to fertility and other life decisions.

Finally, the homogenization mechanism may be stronger for women than for men as a result of the particular choices made by women throughout the course of their STEM careers. For example, female researchers may undergo additional output homogenization in the form of prolonged productivity losses stemming from fertility decisions and increasing prioritization of family needs over career goals. In particular, such gendered constraints on performance may be most pronounced among stars and other above-average performers. Over time, the particular choices made by women may contribute to greater output homogenization among women, specifically via reductions in variability at the right tail of the distribution.

[Insert Table 4 about here]

Summary

As discussed thus far, each of the three gender perspectives has unique implications regarding how performance may differ across genders under each dominant generative mechanism. Thus, depending on the particular distribution shapes and dominant generative mechanisms identified for each gender group, the findings of this study could yield (varying levels of) support to the numerous and possibly competing theoretical notions within each of the three gender perspectives. For example, a gender performance gap in favor of men would be more aligned with the notion that women are given less research support (and other resources) compared to, say, the notion that women must be more productive than men in order to ‘survive’ in high-level STEM environments.

This study may also yield unexpected findings that conflict with one or more of the existing gender perspectives. For example, the analysis may reveal that the performance distribution for women actually has a substantially heavier right tail, suggesting that there is greater variability in the productivity of top female researchers compared to top male researchers and that there is a gender performance gap among star performers in favor of women. Such a finding could indicate that the impacts of, say, sex differences in quantitative aptitudes are *either* not substantial enough to create a gender performance gap in favor of men (at least among top STEM researchers) *or* that those impacts are overridden by other competing forces, such positive biases towards star women over star men.

It may even turn out that the gender groups have entirely different dominant generative mechanisms. Such a finding would call for additional theorization with respect to what factors and processes set men and women in STEM fields apart so much that that they would differ in terms of the fundamental underlying mechanisms driving the variation in their STEM performance. For example, from a gender discrimination perspective, if proportionate differentiation is revealed to be the dominant generative mechanism for men but not for women, the finding may be indicative of gender discrimination that constrains women's ability to achieve large output loops. In particular, gender biases embodied by fellow researchers and potential collaborators (e.g., giving women less credit for co-authored work, greater preference for collaborating with star men compared to star women, etc). Alternatively, the finding may be suggest the effects previously unexplored factors/mechanisms at play, such as gender differences in certain behaviors that enable proportionate differentiation (e.g., tendency to engage in networking behaviors with prior collaborators' network ties, self-promotion strategies, etc).

CHAPTER III

METHOD

I conducted three separate studies involving objective measures of performance in different domains within STEM fields. Study 1 included data regarding research productivity in the field of mathematics. Study 2 included data regarding research productivity in the field of material sciences. Study 3 included data regarding data regarding research productivity in the field of genetics. Conducting these studies across multiple domains should help to observe differences across STEM fields and to generalize the study results to different types of work contexts and domains characterized by similar gender imbalances.

Data and Measures

Measuring Research Performance

Number of papers published in top-tier journals. Prior to data collection, I identified the ten most influential journals in the field of Mathematics (for Study 1), and the five most influential journals each in the fields of Material Sciences (for Study 2) and Genetics (for Study 3). For this step, I gathered impact factor data of the top journals in each fields from 2011 to 2015, included in the Thomson Reuters Journal Citation Reports database. The impact factor of a journal is the average number of citations received per paper published in that journal during the two preceding years (Aguinis, Suarez-González, Lannelongue, & Joo, 2012). For example, if a journal has an impact factor of 4 in 2015, then its papers published in 2014 and 2013 received 4 citations each on average in 2015. I selected the ten journals with the highest mean impact factor over the past five years (i.e., from 2011 to 2015) for Study 1 and the top five journals for Studies 2 and 3.

Due to subjective bias, psychometric deficiency, and contamination problems associated with performance measurement (collectively known as the "criterion problem"), it is challenging to obtain accurate subjective ratings of performance (Austin & Villanova, 1992; Campbell, McCloy, Oppler & Sager, 1993; Cascio & Aguinis, 2011). As such, this research utilizes objective, rather than subjective, performance measures. In this study, research performance was assessed as a count of the total number of articles published by each author in all of the top journals (top 10 for Study 1, and top 5 for Studies 2 and 3) during the 10-year period from January 2006 to December 2015. I focus on performance as being assessed by the number of article publications in top-tier journals, because this is the most important antecedent of numerous meaningful outcomes for faculty, including salary and tenure status (Gomez-Mejia & Balkin, 1992). In this stage of the data collection, I used the Web of Science database to identify all articles and their authors in the selected journals in the past decade. Using the metadata associated with each of the articles, I recorded the names of all authors, disambiguating author information using ORCID to identify unique authors as needed. After creating a complete list of the researchers, their total publication counts were recorded.

The question of author order on co-authored papers. One weakness of measuring research performance as the total count of one's publications is that the measure does not differentiate between sole- versus co-authorship. As such, the total count measure is prone to inflating the performance of researchers with relatively greater proportions of multi-authored publications. This is a potentially significant limitation given that the number first author papers in particular is one of the key metrics that academics rely on when judging the work of prospective graduate students, prospective junior faculty, etc. Because author ordering may more-or-less be a proxy for differential contributions, I considered utilizing an additional measure of research performance that reflects both the frequency and order of authorship. Specifically, I employed a procedure developed by Howard, Cole, and Maxwell (1987) that, for each publication, assigns the authors *rank-weighted author credits* that are proportional to their ordinal position. The formula used to compute the credits is as follows:

$$\text{credit} = (1.5^{n-i}) \;/\; \left(\sum_{i=1}^{n} 1.5\right)^{i-1}$$

where n is the total number of authors and *i* is the focal author's ordinal position. In essence, each publication is worth 1 full point, which, in articles with multiple authors, is shared among the authors. This procedure ensures that an author in ordinal position $i + 1$ will receive at least half of the authorship credits assigned to the author in ordinal position *i*. For example, in an article with three authors, the first, second, and third authors would be given 0.4737, 0.3158, and 0.2105 credit units, respectively.

Although I initially considered author credits as an additional measure of research performance, I utilized the inputs of field experts to deduce that this would not be a suitable measure. This measure assumes that author order reflects each author's relative contribution, with the first author having contributed the most, then the second, and so on in descending order. While this is true for publications in the management field, this tradition does not seem to exist in the fields of mathematics, material Sciences, and genetics, according to field experts. For example, in the fields of Materials Sciences and Genetics, the first and last (sets of) authors are usually the most important in terms of contribution, although there may sometimes be exceptions. A more detailed discussion of author ordering across the STEM fields is provided in Appendix B.

Gender

The gender of each of the researchers was recorded based on first names. In cases where the gender associated with a first name is ambiguous (e.g., by the use of initials only or gender neutral first names), I visited the author's webpage (personal, faculty, profile on ResearchGate and/or Google Scholar, etc.) to ascertain their gender. In rare cases where the first name was ambiguous *and* I could not find a webpage, photo, or other information that would reveal that author's gender, I used the website Namepedia.org as a last resort to find the gender that is most strongly associated with his/her name. When a first name entered into the Namepedia database, it generates entries for that name (by country/language), including miscellaneous information such as the regions and languages to which that name can be traced back to and whether the gender associated with that name in that particular region is male, female, mostly male, or mostly female, etc. For example, although Jean is more likely a female first name in English-speaking countries, in France, it is usually a male first name. In such cases where a first name can be associated with different genders depending on the region, I first deduced the author's ethnicity and/or geographic region via their surname or other information such as the location of their current workplace, alma mater etc. Then, I coded their gender that best matches their name according to Namepedia, given his/her geographic/ethnic information.

Study 1: Mathematics Research Performance

In Study 1, I examine the performance of *all researchers* in the field of Mathematics who have published in the past decade (i.e., from the start of 2006 to the end of 2015) at least one article in one of the ten most influential Mathematics journals.

The field of mathematics is one of the most masculine and male-dominated disciplines within STEM and represents a domain where some of the most severe gender performance gaps might be observed. For example, according to a recent estimate, only 7.3% of full professor positions in the field of mathematics are occupied by women (Ceci & Williams, 2010). It has also been suggested that certain forms of gender discrimination (e.g., with respect to peer reviews) may be confined to "culturally masculine topics or male-dominated research areas" (Eagly& Miller, 2016, p. 901). These reasons make the field of mathematics a compelling context within which to explore possible gender performance gaps among stars.

The sample is representative of the population of star researchers in the field of Mathematics because it includes the entire population of individuals who have published at least once in the most influential journals. The ten most influential journals (in terms of their mean impact factor from 2011 to 2015) were as follows: Acta Numerica, Journal of the American Mathematical Society, Communications on Pure and Applied Mathematics, Acta Mathematica, Annals of Mathematics, Fractional Calculus and Applied Analaysis, Foundations of Computational Mathematics, Publications of Mathematiques de I IHES, Inventiones Mathematicae, and Bulletin of the American Mathematical Society. The sample consists of N = 3,853 unique researchers of whom 360 (9.3%) are women. The total number of article published by these researchers in the selected mathematics journals was 3,161.

At first, I had gathered data on all mathematics researchers who had published a paper in one of the top five journals as oppose to the top ten. However, this resulted in a sample size that was much smaller than anticipated (i.e., a total of 2,169 unique researchers of whom 158 are women). To acquire a greater sample size, I gathered more data to encompass all researchers who had published in the top ten mathematics journals. Even so, the sample size—and the total number of publications—appears somewhat small, considering that that it includes every researcher who has published at least one paper in the top ten mathematics journals over a period of ten years.

I later discovered that this has to do with unique publication practices in mathematics that differ from those of other disciplines. In 2006, the American Mathematical Society (AMS) made a statement that "mathematicians tend to publish at rates that are modest compared to some other sciences" and that even exceptionally productive mathematicians (e.g., Sloan Fellowship winners) publish an average of two or fewer articles per year. According to the statement, in the field of mathematics, the majority of research is published in refereed journals as opposed to conference proceedings or books. Moreover, articles typically represent "considerable advances on a mathematical question," and as a result, publication tends to take longer, and research is generally considered less time-sensitive in this field. According to the AMS, a key metric for judging the work of mathematicians is that of "the quality of publications" in addition to the rate. They also added that these facts are familiar to mathematicians, but are generally unknown to other researchers.

As illustrated in the next sections below, I was able to acquire larger sample sizes and capture a greater of range of top performers (with respect to the total count of publications) by conducting the studies in the fields of materials sciences and genetics, where publication rates among top performers appear to be greater and more varied than in the field of mathematics.

Study 2: Materials Sciences Research Performance

In Study 2, I examine the performance of *all researchers* in the field of Material Sciences who have published at least one article in one of the five most influential journals in the past decade. The process for identifying the five most influential journals and star performers was identical to Study 1. One unique issue regarding this field is that the Thomson Reuters database includes eight different entries for material sciences (e.g., ceramics, paper and wood, composites). So, I identified authors across all material sciences so that authors publishing in more than one area would have all their publications included. All other data collection procedures and measures were identical to Study 1. The top five journals in the field of material sciences, based on their mean impact factor from 2011 to 2015, were as follows: Nature Materials, Nature Nanotechnology, progress in Materials Science, Materials Science & Engineering R-Reports, and Advanced Materials. The sample for this study is representative of the population of star researchers in the field of Material Sciences because it includes the entire population of those who have published at least once in the five most influential journals. The sample for Study 2 consists of N = 35,642 unique researchers of whom 5,086 (14.2%) are women. The total number of articles published by these researchers in the selected journals was 13,303.

As suggested by these statistics, publishing rates in top material sciences journals appear higher than in the field of mathematics. The sample size (N = 35,642) is also considerably larger than that of Study 1 (N = 3,853) despite the fact that I gathered data on authors publishing in only the top five journals as opposed to ten. Also, the representation of women in the sample size is greater (14.2% versus 9.3%), as consistent with the observation that mathematics is one f the most male-dominated fields within STEM.

Study 3: Genetics Research Performance.

In Study 3, I examine the performance of *all researchers* in the field of Genetics who have published at least one article in one of the five most influential journals in the past decade. The field of genetics is also a very compelling context within which to examine the gender performance gap, as the field of biological and life sciences have the one of the greatest concentration of women across the sciences. According to the NSF, women comprise approximately 48% scientists in biological, agricultural, and environmental life sciences (NSF, Science & Engineering Indicators, 2016), which is substantially larger than their representation in other fields (e.g., 25% in computer science and mathematics). It would be thus be insightful to examine whether gender disparities exist among star performers even in a field with relatively high levels of female participation.

The process for identifying the five most influential journals and star performers was identical to Study 1 and Study 2. This sample is representative of the population of star researchers in the field of Genetics because it will include the entire population of those who have published at least once in the five most influential journals. The top five journals in the field of genetics, based on their mean impact factor, were as follows: Nature Reviews Genetics, Nature Genetics, Annual Review of Genetics, Trends in Ecology & Evolution, and Genome Research. The sample for Study 3 consists of N = 45,007 unique researchers of whom 14,685 (32.6%) are women. The total number of articles published by these researchers in the selected journals was 7,746.

Analytic Technique

Distribution Pitting

Data analysis for this research will be comprised of two main stages. The first step is that of identifying the theoretical distributions that best model the performance of male versus female researchers. To this end, I conducted a methodological approach referred to by Joo and colleagues (2017) as *distribution pitting*. Distribution pitting is a falsification approach-based procedure (Gray & Cooper, 2010; Lakatos, 1976; Leavitt, Mitchell, & Peterson, 2010; Popper, 1959) that involves pitting theoretical distributions against one another with respect to the observed distribution. Specifically, it involves pitting the seven distributions (described in Table 1) against one another to identify those that do not accurately reflect the data, and, in doing so, determine the dominant (i.e., surviving) distribution.

The distribution pitting procedure involves three decision rules used to ultimately identify the likely dominant distribution (and associated generative mechanism) for each of the six samples (i.e., the male versus female samples in Studies 1, 2, and 3). The first decision rule involves generating distribution pitting statistics using Dpit. Dpit is a novel R package that allows pairwise comparisons of distribution fit with respect to the observed distribution. Dpit was used to conduct 21 pairwise fit comparisons for each of the seven theoretical distributions. The Dpit package yields distribution pitting statistics that form the basis of the first decision rule. Specifically, the package provides the log likelihood ratio (LR) and its associated *p* value (Aguinis, Gottfredson, & Culpepper, 2013) for each pairwise comparison. LR is calculated by subtracting the log likelihood fit of the second distribution from that of the first distribution. So, positive LR values indicate greater empirical support for the first distribution, whereas negative LR values indicate greater empirical support for the second distribution. The *p* value associated with each LR value was used to rule out whether or not the non-zero LR value is due to random fluctuations alone (Clauset et al., 2009). Because the null hypothesis is set to LR = 0, the lower the *p* value, the less likely that the LR value is just due to chance. In this study, a *p* value cutoff of 0.10 (Clauset et al., 2009) was used. These statistics were used in the first decision rule to 'disconfirm' distributions that can be ruled out by any of the other distributions. In short, the only the distribution(s) that has not been ruled out (i.e., the surviving distribution) is considered as having a plausible fit to the data

If the first step did not result in only one surviving distribution, I applied the principle of parsimony to rule out one of the surviving distributions. Specifically, if two mathematically nested distributions survived the first step of the distribution pitting process, I subsequently chose the distribution with fewer parameters as being the better explanation for the observed distribution. Distributions with more parameters are guaranteed to have equivalent or superior fit to the observed distribution; however, they are associated with reduced patrimony and risks reduced generalizability beyond the specific sample being used. Out of the 21 distribution comparisons, three are nested: (1) pure power law (one parameter) is nested within power law with exponential cutoff (two parameters); (2) exponential distribution (one parameter) is nested within power law with exponential cutoff (two parameters); and (3) exponential distribution (one parameter) is nested within the Weibull distribution (two parameters). So, for example, if the exponential and Weibull distributions equally fit a sample, I identified the former as being the better explanation for the observed distribution.

Third, if the first and second steps did not result in only one surviving distribution for a particular sample, I relied on the principle of triangulation to rule out one or more of the surviving distributions. Specifically, over certain parameter values, three distributions (i.e., lognormal, Poisson, and Weibull) are "flexible" in that each can look similar to the other four "inflexible" distributions (i.e., normal, exponential, pure power law, and power law with exponential cutoff). The converse is not necessarily true. So, if a flexible distribution and an inflexible distribution remain, then the flexible distribution serves as evidence for the presence of both distributions, whereas the inflexible distribution can only serve as evidence of itself. In other words, because the inflexible distribution was confirmed twice, whereas the flexible distribution confirmed only once, the flexible distribution can be ruled out. In short, if one or more of the three flexible distributions along with one or more of the four other distributions still remain survivors, the appropriate decision is to rule out the flexible distribution(s) while keeping the inflexible distribution(s) (Edwards & Berry, 2010). Using these three I identified the best-fitting distribution associated with the each of the samples. Identifying best-fitting distributions should reveal the implied differences, if any, in the generative mechanisms of male versus female star performers in STEM fields.

Comparison of Fit Parameters

Upon completing the first step of the analysis, i.e., identifying the best-fitting theoretical distributions (and associated generative mechanisms) for each gender group, I estimated the specific parameters associated with those distributions, e.g., the mean (μ) and standard deviation (σ) of a normal distribution or the pure power law's rate of decay (α). These parameters provide further details on the specific shapes of the performance distributions. In particular, this assessment is necessary to compare differences in the 'right-rail heaviness' of the performance distributions across genders. The presence of different generative mechanisms across genders, by alone, could be a 'smoking gun' for significant gender gaps among star performers. To illustrate, the pure power law distribution has a heavier tail than all the other distributions and implies the most variability among top performers. Hypothetically, if the pure power law distribution had the best fit for men, but not for women, then this finding alone is enough to conclude that the performance distribution for men has a heavier right tail and that there is performance gap in favor of star men over star women. However, if both gender groups share the generative mechanisms, a comparison of fit parameters is necessary to compare the heaviness of the distributions' right tails.

First, the power law's parameter alpha, which can be used as a proxy for 'heavy tailed-ness' in general (Joo et al., 2017), was computed for all performance distributions. Lower alpha values (i.e., closer to 1) suggest high variability in performance and that productivity is dominated by a small group of elites. In addition, I computed distribution-specific parameters for groups with best-fitting distributions other than the pure power law. So, for example, if the performance of one group was best described by the lognormal distribution, the lognormal standard deviation (σ) associated with that sample was estimated. In addition to the pure power law's parameter alpha, distribution-specific parameters can provide a more precise way to assess the heaviness of the distribution's right hand tail and useful for comparing two groups sharing same best-fitting theoretical distribution. All of the fit parameters were computed using functions included in the Dpit package.

To supplement this information, I also computed the productivity of top performers (top 10, 5, and 1%) relative to total output of their gender group and include visual comparisons of the distributions across genders. This information serves to better illustrate the practical significance of the fit parameters associated with each distribution, specifically in terms of gender differences in the right tails. Together, this information will allow me to better assess the extent of any gender performance gaps that exist at the distributions' right tails. The findings of this analysis should then yield insight into how and why those distributions—and differences between distributions—emerge. Precise estimates of the size of the gender performance gap generated by these results could then be used as baseline data for future studies.

CHAPTER IV

RESULTS

Results for Study 1: Mathematics Research Performance

Distribution Pitting Results

Table 5 summarizes the distribution pitting results for Study 1. The analyses revealed that, in both the female and male samples, the power law with exponential cutoff distribution had the best fit with the data. For women, the normal, pure power law, Poisson, and exponential distributions were disconfirmed via the first decision rule. The normal distribution had significantly worse fit than all of the other distributions (as indicated by negative LR values with a p-value less than 0.10); the pure power law had significantly worse fit than the power law with exponential cutoff, Weibull, and lognormal distributions; the Poisson distribution had worse fit than all other distributions except for the normal distribution; and the exponential distribution had worse fit than the power law with exponential cutoff and Weibull distributions. Thus, the pure power law with exponential cutoff and Weibull distributions remained after the implementation of the first decision rule. The second decision rule (i.e., that among nested distributions, the distribution with more parameters is ruled out) did not further rule out any of the remaining distributions as none of the remaining distributions (power law with exponential cutoff, Weibull, and lognormal) are nested within another. Finally, the third decision rule was used to rule out the Weibull and lognormal distributions—these are ‘flexible’ distributions as opposed to the power law with exponential cutoff distribution which is ‘inflexible’ and thus serves only as evidence of itself. To summarize, for women, the power law with exponential cutoff, Weibull, and lognormal distributions remained after the implementation of the first and second decision rules, and only the former remained after implementing the third decision rule.

[Insert Table 5 about here]

For men, all but the pure power law with cutoff and Weibull distributions remained after the implementation of the first decision rule. The normal distribution had significantly worse fit than the other six distributions; the pure power law distribution had worse fit than the power law with exponential cutoff, Weibull, and lognormal distributions; the lognormal distribution had worse fit than the Weibull distribution; the Poisson distribution had worse fit than all other distributions except for the normal distribution; and the exponential distribution had worse fit than the power law with exponential cutoff, Weibull, and lognormal distributions. The second decision rule did not further rule out either of the remaining distributions, as the power law with exponential cutoff and Weibull distributions are not nested within one another. Finally, the third decision rule was used to disconfirm the Weibull distribution (flexible). To summarize, for men, the power law with exponential cutoff and Weibull distributions remained after the implementation of the first and second decision rules, and only the former remained after implementing the third decision rule.

Fit parameters and Visual Inspection

In this study, the pure power law with exponential cutoff distribution was found to be the best fitting model for both gender groups. This suggests that both genders share the same dominant generative mechanism of incremental differentiation. To further assess the shape of the distributions, I computed the pure power law's parameter alpha and exponential distribution's parameter lambda using Dpit. For women, the alpha and lambda parameters were $\alpha = 2.94$ and $\lambda = 0.57$, respectively. For men, the alpha and lambda parameters were $\alpha = 2.39$ and $\lambda = 0.47$ (i.e., smaller for both), respectively. These parameter estimates imply that, despite both genders sharing the same dominant generative mechanism, the distribution for men has a heavier right tail and that there is a gender performance gap in favor of men. To illustrate the practical significance of this gap, the top 10%, 5%, and 1% of female performers accounted for 23.7%, 15.2%, and 4.9% of the total output (i.e., the total number of authorships) of all women, whereas the top 10%, 5%, and 1% of male performers accounted for 29.7%, 18.9%, and 6.3% of the total output of all men.

Figure 2 depicts histograms and kernel density plots of the research productivity of female versus male researchers in mathematics. The binwidth for all histograms and kernel density plots are set to 1—this decision makes sense given that the performance measure utilized in this study (i.e., total count of publications) is a discrete measure. For women, the number of publications ranged in total from 1 to 7. Additionally, the top 10% of female performers published papers within the range of 2 to 7, the top 5% published within the range of 3 to 7 papers, and the top 1% published within the range of 5 to 7. For men, the number of publications ranged in total from 1 to 20. The top 10% of male performers published within the range of 3 to 20 papers, the top 5% published within the range of 4 to 20 papers, and the top 1% published within the range of 8 to 20 papers. This means that all of the top 1% of male researchers, individually, outperformed all of the female researchers in the study in terms of their number of publications. In addition, the data shows that among all mathematics researchers in the study (i.e., men and women combined), the top 1% of performers are composed entirely of male researchers with least 8 top-tier journal publications. To recap, the findings thus suggest that, despite both genders sharing the same dominant generative mechanism, the performance distribution for men has a heavier right tail and that there is a significant gender performance gap among star performers in the field of mathematics in favor of men.

[Insert Figure 2 about here]

Results for Study 2: Materials Sciences Research Performance

Distribution Pitting Results

Table 6 summarizes the distribution pitting results for Study 2. For women, the power law with exponential cutoff distribution was identified as the best-fitting distribution, and, for men, the lognormal distribution was identified as the best-fitting distribution. For women, the normal, pure power law, Poisson, Weibull, and exponential distributions were disconfirmed via the first decision rule. The normal distribution had significantly worse fit than all of the other distributions; the pure power law had significantly worse fit than the power law with exponential cutoff, Weibull, and lognormal distributions; the Poisson distribution had worse fit than all other distributions except for normal; the Weibull distribution had worse fit than the lognormal and the power law with exponential cutoff distributions; and the exponential distribution had worse fit than the pure power law, power law with exponential cutoff, lognormal and Weibull distributions. Thus, the pure power law with exponential cutoff and lognormal distributions remained after the implementation of the first decision rule. The second decision rule did not further rule out either of the remaining distributions. Finally, the third decision rule was used to rule out the lognormal distribution (i.e., flexible). In summary, for women in Study 2, the power law with exponential cutoff and lognormal distributions remained after the implementation of the first and second decision rules, and only the former remained after implementing the third decision rule.

[Insert Table 6 about here]

For men, all but the lognormal distribution were disconfirmed after implementing the first decision rule. The normal distribution had worse fit than all of the other distributions; the Poisson distribution had worse fit than all other distributions except for the normal distribution; the pure power law distribution had worse fit than the power law with exponential cutoff, Weibull, and lognormal distributions; the power law with exponential cutoff had worse fit than the lognormal distribution; the Weibull distribution had worse fit than the lognormal distribution; and the exponential distribution had worse fit than the pure power law, power law with exponential cutoff, Weibull, and lognormal distributions.

Fit parameters and Visual Inspection

The distribution pitting results imply the presence of different dominant generative mechanisms across gender groups (i.e., incremental differentiation for women and proportionate differentiation for men). Given that the lognormal distribution is associated with the second heaviest right tail next to the pure power law distribution (Joo et al., 2017), the results suggest that there is greater performance variation among top male performers compared to the top female performers. Accordingly, the pure power law's alpha parameter was lower for men (α = 2.36) than for women (α = 2.59). Similarly, the exponential distribution's lambda parameter was equal to greater for women (λ = 0.49) than for men (λ = 0.44). In addition, estimates of the lognormal standard deviation was lower for women (SD = 0.57) than for men (SD = 0.65). These fit parameters suggest that the performance distribution for men has a heavier right tail and that there is a performance gap in favor of men. To illustrate the practical significance of this gap, the top 10%, 5%, and 1% of female performers accounted for 32.1%, 21.8%, and 8.1% of the total output of all women, whereas the top 10%, 5%, and 1% of male performers accounted for 35.9%, 25.1%, and 10.1% of the total output of all men.

Figure 3 depicts histograms and kernel density plots of the research productivity of female versus male researchers in materials sciences. For women, the number of publications ranged in total from 1 to 36. The top 10% of female performers published within the range of 3 to 36 papers, the top 5% published within the range of 4 to 36 papers, and the top 1% published within the range of 8 to 36 papers. For men, the number of publications ranged in total from 1 to 91. The top 10% of male performers published within the range of 3 to 91 papers, the top 5% published within the range of 5 to 91 papers, and the top 1% published within the range of 11 to 91 papers. The data also shows that among the top 1% of all materials sciences researchers in the study (i.e., men and women combined), only 7% were women in despite the fact that women comprised of 14% of the total combined sample. Furthermore, the top 0.1% of all performers is composed entirely of male researchers with least 37 top-tier journal publications. In summary, the findings imply that there are different dominant generative mechanisms across gender groups (i.e., incremental differentiation for women and proportionate differentiation for men) and that there is a considerable gender performance gap among star performers in the field of materials sciences in favor of men.

[Insert Figure 3 about here]

Results for Study 3: Genetics Research Performance

Distribution Pitting Results

Table 7 summarizes the distribution pitting results for Study 3. The analyses revealed that, for both gender groups, the power law with exponential cutoff distribution had the best fit with the data. For women, the normal, pure power law, Poisson, Weibull, and exponential distributions were disconfirmed via the first decision rule. The normal distribution had significantly worse than fit than all of the other distributions; the pure power law had worse fit than the power law with exponential cutoff and lognormal distributions; the Poisson distribution had worse fit than all of the other distributions except for the normal distribution; the Weibull distribution had worse fit than the lognormal, pure power law, and power law with exponential cutoff distributions; and the exponential distribution had worse fit than the pure power law, power law with exponential cutoff, lognormal and Weibull distributions. Thus, the pure power law with exponential cutoff and lognormal distributions remained after implementing the first decision rule. The second decision rule did not further rule out any of the remaining distributions as the remaining distributions are not nested within another. Finally, the third decision rule was used to rule out the lognormal distribution (i.e., flexible), To summarize, for women, the power law with exponential cutoff and lognormal distributions remained after the implementation of the first and second decision rules, and only the former remained after implementing the third decision rule.

[Insert Table 7 about here]

For men, the normal, pure power law, Poisson, Weibull, and exponential distributions were disconfirmed via the first decision rule. The normal distribution had significantly worse fit than all of the other distributions; the pure power law had significantly worse fit than the power law with exponential cutoff and lognormal distributions; the Poisson distribution had worse fit than all of the other distributions except for the normal distribution; the Weibull distribution had worse fit than the lognormal, pure power law, and power law with exponential cutoff distributions; and the exponential distribution had worse fit than the pure power law, power law with exponential cutoff, lognormal and Weibull distributions. Thus, the power law with exponential cutoff and lognormal distributions remained after implementing the first decision rule. The second decision rule did not further rule out either of the remaining distributions. The third decision rule was used to rule out the lognormal distribution (flexible). In summary, for men, the power law with exponential cutoff and lognormal distributions remained after the implementation of the first and second decision rules, and only the former remained after implementing the third decision rule.

Fit parameters and Visual Inspection

As with Study 1, the pure power law with exponential cutoff distribution was found to be the best fitting model for both gender groups. This suggests that both genders share the same dominant generative mechanism of incremental differentiation. To further assess the shape of the distributions, I estimated the alpha and lambda parameters (i.e., rates of decay) for each gender group. Both parameters were slightly greater for women (α = 2.43 and λ = 0.44 for women; α = 2.30 and λ = 0.40 for men). These parameters suggest that the right for women may be lighter than for men, but this difference appears to be very small relative to the size of the performance gaps identified in the first two studies. To illustrate, the top 10%, 5%, and 1% of female performers accounted for 38.9%, 28.3%, and 12.4% of the total output of all women. These values are quite similar to those of men—the top 10%, 5%, and 1% of male performers accounted for 40.8%, 29.6%, and 12.6% of the total output of all men.

Figure 4 depicts histograms and kernel density plots of the research productivity of female versus male researchers in genetics. Women actually exceeded men in terms of their total range of publications (i.e., 1 to 123 for women versus 1 to 102 for men). Accordingly, the top 10% of female performers published papers within the range of 3 to 123 papers, the top 5% within the range of 5 to 123 papers, and the top 1% within the range of 12 to 123 papers. The top 10% of male performers published within the range of 4 to 102, the top 5% within the range of 6 to 102 papers, and the top 1% within the range of 15 to 102 papers. In addition, among the top 1.2 % of all genetics researchers in the study (i.e., men and women combined), 24.9% were women; and, among the top 0.1% of all researchers, 24% were women. Given that women comprised of 32.6% of the combined sample of genetics researchers, these findings suggest that women are relatively underrepresented at the upper tail of the combined performance distribution. However, this the size of this gender performance gap is very small compared to those identified in the previous studies. The findings across all three studies are summarized in Table8.

[Insert Figure 4 about here]

[Insert Table 8 about here]

CHAPTER V

DISCUSSION

In this chapter, I discuss the results of my findings and provide an interpretation of the results. I then offer the implications of this study in terms of both theory and practice. Next, I describe the limitations of the study, and discuss possible directions for future research.

Observed Gender Performance Gaps among Star Performers

In Study 1, I found that the power law with exponential cutoff distribution best fit the performance distributions of both male and female researchers in the field of mathematics. Given the dominant generative mechanism associated with this shape (i.e., incremental differentiation), the findings suggest that, for both genders, individual variation in cumulative performance is driven by differences in accumulation rates (i.e., output produced per opportunity to perform). Although both gender groups share the same dominant generative mechanism, the date revealed significant gender disparities in the heaviness of the distributions' right tails. According to the results, there is less variability in research productivity among top-performing women than among top-performing men. To illustrate, while the top 1% of female performers accounted for 4.9% of the total output of females and published in the range of 5 to 7 papers in the ten most influential mathematics journals from 2005 to 2015, the top 1% of male performers accounted for 6.3% of the total output of males and published in the range of 8 to 20 papers in those same journals. Moreover, the top 1% of all performers—in terms of their total count of top-tier journal publications—was comprised entirely of men. Thus, the findings suggest that, in the field of mathematics, there is a significant gender performance gap among star performers in favor of men.

In Study 2, I found that the power law with exponential cutoff distribution best fit the performance distribution of female researchers in the field of materials sciences, while the lognormal distribution best fit the performance distribution of male researchers. This suggests a gender disparity in dominant generative mechanisms (i.e., proportionate differentiation for men versus incremental differentiation for women) in this field. Specifically, the findings suggest that, for women, individual performance differences are shaped by individual differences in accumulation rates. For men, individual performance differences are shaped by individual differences in initial output values and/or accumulation rates. This suggests that, for men in this field, prior outputs contribute to subsequent outputs through positive interaction with men's accumulation rates. In contrast, for women, cumulative performance is determined largely by their linear output increments based on their accumulation rates, and prior outputs are not (as strongly) linked to future output. Visual inspection and comparison of fit parameters also suggest that the performance distribution for men has a heavier right tail. To illustrate, the top 1% of female performers accounted for 8.1% of the total output of females and published in the rage of 8 to 26 papers in the five most impactful journals in their field. The top 1% of male performers accounted for 10.1% of the total output of males, however, and published in the rage of 11 to 91 papers in the same top-tier journals. Among the top 1% of all performers (men and women combined), 7% were women, despite the fact that women comprised of 14% of the total sample; additionally, the top 0.1% of performers were composed entirely of men. Thus, the findings suggest that, in the field of materials sciences, there is a considerable gender performance gap among in favor of men.

In Study 3, I found that the power law with exponential cutoff distribution best fit the performance distributions of both male and female researchers in the field of genetics. This suggests that, for both genders, individual variation in cumulative performance is driven by differences in accumulation rates (i.e., output produced per opportunity to perform). In this study, there were relatively smaller gender differences in the performance distributions' right tail heaviness. To illustrate, the top 1% of female performers accounted for 12.4% of the total output of females and published in the rage of 12 to 123 papers in the top tier journals in their field. The top 1% of male performers accounted for 12.6% of the total output of males and published in the rage of 15 to 102 papers in the top tier journals in their field. Additionally, while women comprised of roughly one third of the combined sample, they represented roughly one fourth of both the top 1% and the top 0.1% of all researchers in the study (women and women combined). The findings thus suggest that there may be a slight gender performance gap among star performers in favor of men in the field of genetics.

Theoretical Implications

In this section, I provide my interpretation of each study's findings and discuss their theoretical implications, particularly with regard to the three gender perspectives (i.e., sex differences, gender discrimination, and women's choices). First, in the field of mathematics, there was a substantial gender performance gap among star performers in favor of men. The performance distribution for women had a significantly lighter right tail, meaning that there is an even larger gender gap among stars than among all performers. In addition, incremental differentiation was identified as being the dominant generative mechanism for men and women in the field of mathematics. Given the dominant generative mechanism, it can be inferred that mathematicians differ in terms of their total research outputs mainly due to individual differences in accumulation rates. Moreover, the findings suggest that men are more variable in their accumulation rates, and are thus disproportionately represented among stars than among average performers. The highest-performing male researchers should thus experience greater increments in their cumulative performance outputs compared to the highest-performing female researchers.

The observation that women are less variable in their accumulation rates provides support to many theoretical notions within each of the three gender perspectives. First, with regards to sex differences perspective, the findings of Study 1 provide support to the notion that quantitative aptitudes are indeed an important driver of performance— in the field of mathematics at least—and that sex differences in the variability of quantitative aptitudes may contribute to the disproportionate underrepresentation of women among stars. It is possible that women are underrepresented among stars due to their having more balanced cognitive ability profiles and thus more career options outside of STEM fields. To illustrate, women who possess high quantitative aptitudes may be more likely than men with comparably high quantitative aptitudes to also possess high verbal abilities. As such, there may a greater leakage of talented women compared to comparably talented men (and compared to less talented women) in these fields.

Given the large size of the performance gap identified in this study, however, it is likely that there are additional factors—other than sex differences—that constrain the accumulation rates of star women. Recent studies have estimated a female-to-male ratio of roughly 1 to 2 in the top 1% of the quantitative aptitude distribution (Ceci & Williams 2010). In this study, however, the top 1% of all mathematics performers was composed entirely of men. Moreover, all of the top 1% of male researchers outperformed all of the top 1% of female researchers in terms of their total count of top-tier publications (i.e., between 5 to 7 for women and 8 to 20 for men). Thus, given the size of the performance gap compared to what the 1 to 2 ratio would suggest, the finding likely reflects the impacts of other mechanisms at play. Indeed, it is likely that sex differences in quantitative aptitudes alone do not fully explain the gender disparity STEM fields, especially among stars (e.g., Ceci et al., 2014; Ceci & Williams 2010, 2011; Eagly & Miller, 2016; Wang & Degol, 2017).

With regards to the gender discrimination perspective, the finding provides support to the notion that various forms of gender discrimination constrain women's performance in STEM fields. Gender biases embodied by fellow researchers and 'higher-ups' may contribute to, say, female faculty being assigned more teaching roles, receiving less research support, and being perceived as less desirable for collaboration. In short, gender discrimination may act as a ceiling on star women's accumulation rates in this field. On the other hand, the finding does not suggest that women in STEM fields need to outperform men in order to be hired for the same position, receive the same peer review score, or simply to survive the winnowing of less productive women (Wenneras & Wold, 1997).

Finally, the finding is aligned with the notion that there are certain gender differences in motivational/psychological factors that, in turn, lead to disparate performance outcomes for men and women in STEM fields. In particular, the observed gender performance may be an indication that high-performing female mathematicians are disproportionately more likely than high-performing male mathematicians to opt out the field and/or sacrifice research productivity in pursuit of family needs and other goals.

In Study 2, proportionate differentiation was identified as the dominant generative mechanism for men, whereas incremental differentiation was identified as the dominant generative mechanism for women. In addition, the performance distribution for women was found to have a lighter right tail. This finding suggests that male researchers in the materials sciences field differ in their total research outputs due to individual differences in accumulation rates and initial output values. Top-performing male researchers thus enjoy greater output loops based on positive/multiplicative interactions between accumulation rates and prior outputs. This suggests that, in terms of increases in future research output, there is a distinct advantage associated with having a high number of accumulated prior outputs. So, between two male researchers with comparable accumulation rates (i.e., average productivity per time period), the one with higher initial outputs (i.e., a more prolific pipeline of past publications) would enjoy greater subsequent increases in research outputs. For women in the materials sciences field, however, incremental differentiation was identified as the dominant generative mechanism, suggesting that differences in their performance outputs are shaped primarily by individual variation in accumulation rates with minimal regard to initial/prior outputs.

The finding that proportionate differentiation is the dominant generative mechanism for men but not for women, suggest that there are gender differences in the fundamental underlying process that drive individual differences in cumulative performance. This particular gender disparity in the dominant generative mechanisms is unlikely to be the result of any sex differences in quantitative aptitudes. Quantitative aptitudes may be very important in shaping STEM researchers' accumulation rates and initial output values. It is even theoretically plausible that men in these fields begin their tenure-track careers with a greater number of publications and display greater productivity than their female colleagues due to an innate quantitative advantage. Thus, under the same generative mechanism of proportionate differentiation, it makes sense that there would be a lighter tail for women. However, differences in quantitative aptitudes should not affect the strength of any positive interactions that occur between prior outputs and accumulation rates. It also seems unlikely that women's particular career and lifestyle choices negate the presence of proportionate differentiation.

I suggest that the observed gender disparity in dominant generative mechanisms in Study 2 may be the result of certain forms of gender discrimination and biases. It is theoretically possible that, for female researchers in materials sciences, the proportionate differentiation mechanism is relatively constrained due to various gender biases embodied by academic departments and fellow researchers. For example, female researchers in this field may incur unique penalties for co-authorship similar to those that have been observed in the field of economics (Sarsons, 2017). Additionally, minimization of women's past research outputs in decisions affecting women's access to research support (e.g., grant funding) and career advancement opportunities (e.g., promotion decisions) may similarly constrain the link between their prior and future outputs. Alternatively the finding may be an indication of other gender differences in individual attributes (such as competitiveness, work-goal orientations etc.) and or/behavioral tendencies (e.g., establishing network ties with influential individuals in the field) that influence the strength of the proportionate differentiation mechanism disparately across genders.

Finally, in Study 3, incremental differentiation was revealed to be the dominant generative mechanism for both male and female researchers in the field of genetics. In addition, the study revealed much smaller differences in right-tail-heaviness of the performance distributions across genders. The findings thus suggest that (1) both male and female researchers in the field of genetics differ in their total research outputs due to individual differences in accumulation rates and that (2) there is a slight gender performance gap among star performers in favor of men in the field of genetics.

Together, the study findings illustrate the presence of larger gender performance gaps in favor of men in the fields of mathematics and materials sciences compared to genetics. The finding that there the size of the performance gaps vary across fields yields further implications with regards to each of the gender perspectives. First, from the sex differences perspective, it is possible that innate quantitative are more important in shaping performance in some STEM fields over others. There is a possibility that, due to differences in the fundamental nature of work in each discipline—and in how research is conducted across disciplines—quantitative aptitudes are more strongly linked to research productivity in, say, the field of mathematics than in genetics. The other possibility is that sex differences in quantitative aptitudes is not strongly tied to performance in any of the three domains, and that there are simply other factors responsible for the larger gender performance gaps observed in the fields of mathematics and materials sciences.

Similarly, from the gender discrimination perspective, it is possible that the negative effects of gender discrimination women's performance may be more pronounced in certain STEM fields (i.e., mathematics and materials sciences) than in others (i.e., genetics). In particular, it is possible that women's greater relative presence in the field of genetics means that there is less gender bias and discrimination in the field compared to more male-dominated STEM fields. The greater presence of female mentors and fellow female colleagues, for example, may contribute to a climate that is more conducive to the success of female researchers in the field of genetics. In comparison, negative stereotypes about women may also be more severe in STEM fields such as mathematics that are more male-dominated and traditionally viewed as being more masculine.

From the perspective of women's career and lifestyle choices, the findings suggest that women with the highest accumulation rates may be more drawn to certain STEM disciplines over others. Talented women (i.e., with high productivity potential) may be more likely to enter and persist in certain STEM fields (i.e., genetics) over others due to, say, better perceived fit with their occupational preferences and work-goal orientations. To this point, it has been shown that women tend to be more drawn to careers that endorse altruistic goals that benefit members of society, which is arguably a better description of work in genetics and other life science fields compared to work in mathematics or materials sciences. The finding that there is a relatively small gender performance gap in one of the three fields, to some extent, disconfirms the theoretical notion that the performance gaps are caused largely by women's disproportionate productivity losses stemming from fertility decisions. To illustrate, the performance gaps observed in the fields of mathematics and materials sciences are caused largely by women's fertility decisions, a similar performance gap should be observed in the field of genetics. Accordingly, the gender performance gaps across the three STEM fields are likely shaped more strongly by other processes and mechanisms at play.

Practical Implications

The findings of this study confirm the existence of significant gender performance gaps among star performers in favor of men in (at least some) STEM fields. This suggests that, compared to men, women in these fields have even lower representation among stars than among average performers, and that they are outperformed by star men. This is an important finding that highlights the seriousness of the issue of gender disparities in STEM fields. While the performance at the average level illustrates how well people are performing in general, star performance signifies 'potential,' or, the known upper limits of performance and success in that particular domain. As such, the finding that there exists an even larger gender performance gap among stars than among average performers speaks to the urgency of the problem and the need to develop interventions aimed at reducing such gender disparities. In terms of practice, the important question thus far has been: "Why are women underrepresented in STEM fields, and what can be done to increase their participation in these fields?" Given the findings of this study, a more urgent, possibly more important, question has to do with "What can be done to reduce the observed gender performance gaps in STEM fields, particularly among stars?"

The findings of this research suggest that top female STEM researchers may have less variability in their accumulation rates compared to their male counterparts. Additionally, women's potential to achieve output loops, i.e., positive interactions between initial outputs and accumulations rates, may be constrained. It seems highly unlikely that there is a singly culprit for the gender performance gaps identified in this study. For example, the performance gaps may certainly be driven, in part, by sex differences in quantitative aptitudes, but, given the large size of the gender performance. Similarly, women may indeed incur larger productivity losses stemming from fertility decisions, but given that the size of the gender performance gaps differ significantly across STEM fields, it seems unlikely it is the main culprit, as fertility choices should affect women similarly across the STEM fields. It is possible that various forms of gender discrimination in STEM fields constrain women's performance once they have opted in to a STEM profession, possibly more so for stars. The findings suggest, however, that women do not need to outperform men in order to survive the 'winnowing' of less talented women. In fact, it is possible that there may a leniency bias towards women in hiring decisions, resulting in a relatively larger concentration of women whose performance is average or below average. This notion is consistent with Ceci and colleagues' (2014) finding that there may be a 2 to 1 preference for women in tenure track STEM professions.

Given the findings of this study, I suggest that academic STEM departments and STEM organizations should focus on strategies to minimize harmful biases and stereotypes against women. To promote gender equality in career access and advancement, organizations could, say, implement training for recruitment and hiring committees aimed at removing implicit gender biases. Organizations should also consider implementing similar bias training in areas such as performance appraisals. Cluster hiring, which refers to the practice of hiring of multiple scholars into a department based on shared research interests, may also be effective in terms of improving gender diversity and reducing tokenism and stigmatization (Kossek et al., 2017). Such initiatives aimed at removing gendered barriers in these fields should, over time, mitigate the observed gender disparities in representation and performance.

STEM departments and organizations should also focus on attracting and retaining talented women. To this end, they should implement strategies aimed at making STEM jobs and work environments more appealing for women. For example, redesigning jobs and changing organizational cultures to include more relational elements may help to promote better fit between STEM careers and women's occupational preferences (Kossek et al., 2017). Additionally, they could implement policies aimed at improving workplace social support systems. STEM departments and organizations should also ensure that women have sufficient opportunities to advance in their careers while engaged in family life. For example, organizations could implement policies allowing greater flexibility and control over job hours for women who have recently had children, in order to minimize productivity losses due overload and burnout (Kossek et al., 2017). These changes should help to enable a greater influx of talented women over time and create work environments that are more compatible with women and their success.

Limitations of the Study and Directions for Future Research

There are three main limitations associated with this study. First, the presence of a 'dominant' generative mechanism does not imply that it is the *only* mechanism at play. Distribution pitting relies on disconfirming (as opposed to confirming) a set of theoretical distributions (and their associated dominant generative mechanisms) with respect to their fit with the data. Thus, even if one mechanism (e.g., incremental differentiation) was identified as being the dominant mechanism, this does not preclude the possibility that other mechanisms also play a role—albeit a lesser one—in shaping individuals' performance outputs. For example, in Study 2, incremental differentiation was identified as the dominant generative mechanism for women after the application of all three decision rules; however, the lognormal distribution (which is associated with the proportionate differentiation mechanism) remained after the application of the first and second decision rules and was subsequently ruled out by the third decision rule. Based on the logic of distribution pitting, this finding does not suggest that proportionate differentiation mechanisms are absent, but rather that the mechanism is likely not as powerful as the incremental differentiation mechanism in terms of shaping the performance of women in this field. In sum, despite revealing the dominant generative mechanisms, this study is limited in the sense that it does not yield precise conclusions about the relative importance of each generative mechanism with respect to individual performance. Accordingly, future research could focus on examining the precise extents to which certain generative mechanisms 'dominate' over others, across genders and across various STEM disciplines.

Second, this study does not yield conclusions about the precise relative weights of each of the various processes within each gender perspective in terms of explaining the observed gender performance gaps. For example, given what we know about the size of sex differences in quantitative abilities, the findings of this study strongly suggest that sex differences alone do not fully explain the observed gender performance gaps. However, additional investigation would be necessary to determine precisely how much (or how little) sex differences contribute to those gender gaps. As such, follow-up studies could focus on estimating the explanatory value of each of the gender mechanisms with respect to the performance gaps revealed in this study. Specifically, future research should explore the effects of each gender mechanism on star women and men's accumulation rates, initial output values, and possible interactions between the two.

Finally, the performance measure used in this study, i.e., total count of publications in top-tier journals, is limited in two ways. First, the measure does not differentiate sole versus co-authorships or take into account author order on (co-authored publications). As such, the measure is prone to inflating the performance of researchers with relatively greater proportions of multi-authored publications. Second, the extent to which the one's quantity of top-tier publications reflects his/her research performance may be different across disciplines. For example, in the field of mathematics where publishing rates are lower than many other STEM fields, the quality of one's publications—as opposed to the rate—may be the most important performance metric. Accordingly, follow-up studies could employ alternative/additional measures of individual performance, such as the total/average number citations received, and supervisory/departmental ratings on key research-related behaviors and outcomes.

Conclusion

This dissertation has examined gender performance gaps among star performers in STEM fields. The theory and empirical results confirm the existence of considerable gender performance gaps in favor of men. In addition, the gender performance gap was larger among stars than among all performers. The findings of this study suggest that STEM researchers vary in their performance largely due to individual differences in accumulation rates (i.e., average output per opportunity to produce). I hope that, in addition to the specific contributions described above, this research spurs further exploration of the underlying dominant mechanisms and processes that shape individual performance (and stardom) across particular domains, possibly differently for men and women.

References

Aguinis, H., & Adams, S. K. R. (1998). Social-role versus structural models of gender and influence use in organizations: A strong inference approach. *Group and Organization Management*, *23*, 414-446.

Aguinis, H, & Bradley, K. J. (2015). The secret sauce for organizational success: Managing and producing star performers. *Organizational Dynamics*, *44*, 161-168.

Aguinis, H., Gottfredson, R. K., & Culpepper, S. A. (2013). Best-practice recommendations for estimating interaction effects using multilevel modeling. *Journal of Management*, 39, 1490-1528.

Aguinis, H., & O'Boyle, E. (2014). Star performers in twenty-first-century organizations. *Personnel Psychology*, *67*, 313-350.

Aguinis, H., O'Boyle, E., Gonzalez-Mulé, E., & Joo, H. (2016). Cumulative Advantage: Conductors and Insulators of Heavy-Tailed Productivity Distributions and Productivity Stars. *Personnel Psychology*, *69*, 3-66.

Aguinis, H., Suarez-González, I., Lannelongue, G., & Joo, H. (2012). Scholarly impact revisited. *Academy of Management Perspectives*, *26*, 105-132.

Amaral, L. A. N., Scala, A., Barthélémy, M., & Stanley, H. E. (2000). Classes of small-world networks. *Proceedings of the National Academy of Sciences, 97*, 11149-11152.

Austin, J. T., & Villanova, P. (1992). The criterion problem: 1917–1992. *Journal of Applied Psychology, 77,* 836-874.

Axtell, R. L. (2001). Zipf distribution of U.S. firm sizes. *Science, 293,* 1818-1820.

Bak, P. (1996). *How nature works: The science of self-organized criticality*. Copernicus, New York.

Baron-Cohen, S. *The essential difference*. (2003). New York: Basic Books.

Baron-Cohen, S., Knickmeyer, R. C., & Belmonte, M. K. (2005). Sex differences in the brain: Implications for explaining autism. *Science, 310*, 819 – 823

Beck, J. W., Beatty, A. S., & Sackett, P. R. (2014). On the distribution of job performance: The role of measurement characteristics in observed departures from normality. *Personnel Psychology*, *67,* 531-566.

Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences*, *107*, 1860-1863.

Benbow, C. P., & Stanley, J. C. (1980). Sex differences in mathematical ability: Fact or artifact? *Science*, *210*, 1262-1264.

Bernardin, H. J., & Beatty, R. (1984). *Performance appraisal: Assessing human behavior at work*. Boston, MA: Kent.

Blickenstaff, J. C. (2005). Women and science careers: leaky pipeline or gender filter? *Gender and education*, *17*, 369-386.

Borsuk, R. M., Aarssen, L. W., Budden, A. E., Koricheva, J., Leimu, R., Tregenza, T., & Lortie, C. J. (2009). To name or not to name: The effect of changing author gender on peer review. *BioScience*, *59*, 985-989.

Brown, C. S., & Leaper, C. (2010). Latina and European American girls' experiences with academic sexism and their self-concepts in mathematics and science during adolescence. *Sex Roles*, *63*, 860-870.

Brown, C. S., & Stone, E. A. (2016). Chapter Four-Gender Stereotypes and Discrimination: How Sexism Impacts Development. *Advances in child development and behavior*, *50*, 105-133.

Call, M. L., Nyberg, A. J., & Thatcher, S. (2015). Stargazing: An integrative conceptual review, theoretical reconciliation, and extension for star employee research. *Journal of Applied Psychology*, *100*, 623-640.

Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1993). *A theory of performance*. In N. Schmitt & W. C. Borman (Eds.), Personnel selection in organizations (pp. 35-70). San Francisco: Jossey-Bass.Carmeli, A., Gelbard, R., & Gefen, D. (2010). The importance of innovation leadership in cultivating strategic fit and enhancing firm performance. *The Leadership Quarterly 21*, 339-349.

Carli, L. L., Alawa, L., Lee, Y., Zhao, B., & Kim, E. (2016). Stereotypes about gender and science: Women≠ scientists. *Psychology of Women Quarterly*, *40*, 244-260.

Cascio, W. F., & Aguinis, H. (2011). *Applied psychology in human resource management*, 7th edition. Upper Saddle River, NJ: Pearson Prentice Hall.

Ceci, S. J., & Williams, W. M. (2011). Understanding current causes of women's underrepresentation in science. *Proceedings of the National Academy of Sciences*, *108*, 3157-3162.

Ceci, S. J., & Williams, W. M. (2010). Sex differences in math-intensive fields. *Current Directions in Psychological Science, 19*, 275-279

Ceci, S. J., Ginther, D. K., Kahn, S., & Williams, W. M. (2014). Women in academic science: A changing landscape. *Psychological Science in the Public Interest*, *15*, 75-141.

Chesler, N. C., Barabino, G., Bhatia, S. N., & Richards-Kortum, R. (2010). The pipeline still leaks and more than you think: a status report on gender diversity in biomedical engineering. *Annals of biomedical engineering*, *38,* 1928-1935.

Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Review, 51,* 661–703.

Danaher, K., & Crandall, C. S. (2008). Stereotype threat in applied settings re-examined. *Journal of Applied Social Psychology*, *38*, 1639-1655.

Daverman, R. J. (2011). Statistics on women mathematicians compiled by the American Mathematical Society. *Notices of the American Mathematical Society, 58,* 1310.

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

Diekman, A. B., Clark, E. K., Johnston, A. M., Brown, E. R., & Steinberg, M. (2011). Malleability in communal goals and beliefs influences attraction to stem careers: evidence for a goal congruity perspective. *Journal of personality and social psychology*, *101*, 902-918.

Diekman, A. B., Weisgram, E. S., & Belanger, A. L. (2015). New routes to recruiting and retaining women in STEM: Policy implications of a communal goal congruity perspective. *Social Issues and Policy Review*, *9*, 52-88.

Downey, A. B. (2005). Lognormal and Pareto distributions in the Internet. *Computer Communications*, *28*, 790-801.

Downey, A. B. (2001). The structural cause of file size distributions. In *Modeling, Analysis and Simulation of Computer and Telecommunication Systems. Proceedings of the Ninth International Symposium, IEEE, Cincinnati, OH* (pp. 361-370). doi:10.1109/MASCOT.2001.948888

Eagly, A. H., & Miller, D. I. (2016). Scientific eminence: Where are the women? *Perspectives on Psychological Science, 11*, 899-904.

Edwards, J. R., & Berry, J. W. (2010). The presence of something or the absence of nothing: Increasing theoretical precision in management research. *Organizational Research Methods, 13*, 668-689.

Ellemers, N., Heuvel, H., Gilder, D., Maass, A., & Bonvini, A. (2004). The underrepresentation of women in science: differential commitment or the queen bee syndrome? *British Journal of Social Psychology*, *43*, 315-338.

Ely, R. J. (1995). The power in demography: Women's social constructions of gender identity at work. *Academy of Management Journal, 38*, 589-634.

Ferriman, K., Lubinski, D., & Benbow, C. P. (2009). Work preferences, life values, and personal views of top math/science graduate students and the profoundly gifted: Developmental changes and sex differences during emerging adulthood and parenthood. *Journal of Personality and Social Psychology, 97*, 517–532

Fletcher, W. K. (1981). *Analytical methods in geochemical prospecting* (vol. 1). Amsterdam: Elsevier.

Garcia-Retamero, R., & López-Zafra, E. (2006). Prejudice against women in male-congenial environments: Perceptions of gender role congruity in leadership. *Sex roles*, *55*, 51-61.

Gibbons, M. T. (2009). Engineering by the numbers. American Society for Engineering Education. Retrieved from *http://www.columbia.edu/~tad61/paper_pathway_07172011.pdf*

Gill, H. S. & O'Boyle, M. W. (1997). Sex differences in matching circles and arcs: A preliminary EEG investigation. *Laterality: Asymmetries of Body, Brain and Cognition, 2*, 33-48.

Gomez-Mejia, L. R., & Balkin, D. B. (1992). Determinants of faculty pay: An agency theory perspective. *Academy of Management journal, 35*, 921-955.

Gray, P. H., & Cooper, W. H. (2010). Pursuing failure. *Organizational Research Methods, 13*, 620-643.

Groshen, E. L. (1991). Five reasons why wages vary among employers. *Industrial Relations: A Journal of Economy and Society*, *30*, 350-381.

Grossman, J. M., & Porche, M. V. (2014). Perceived gender and racial/ethnic barriers to STEM success. *Urban Education*, *49*, 698-727.

Halpern, D. F. (2000). *Sex differences in cognitive abilities (3rd ed.)*. Mahwah, NJ: Erlbaum

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological science in the public interest*, *8*, 1-51.

Harber, K. D. (1998). Feedback to minorities: evidence of a positive bias. *Journal of personality and social psychology*, *74*, 622.

Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*, *269*, 41-45.

Hill, C., Corbett, C., & St. Rose, A. (2010). *Why so few? Women in science, technology, engineering, and mathematics*. Washington, DC: American Association of University Women.

Howard, G. S., Cole, D. A., & Maxwell, S. E. (1987). Research productivity in psychology based on publication in the journals of the American Psychological Association. *American Psychologist*, *42*, 975.

Hubbell, S. P. (1979). Tree dispersion, abundance, and diversity in a tropical dry forest. *Science*, *203*, 1299-1309.

Hyde, J. S. (1990). Meta-analysis and the psychology of gender differences. *Signs: Journal of Women in Culture and Society*, *16*, 55-73.

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender similarities characterize math performance. *Science*, *321*, 494-495.

Javidan, M., Bemmels, B., Devine, K. S., & Dastmalchian, A. (1995). Superior and subordinate gender and the acceptance of superiors as role models. *Human Relations*, *48*, 1271-1284.

Joo, H., Aguinis, H., & Bradley, K. (2017). Not all non-normal distributions are created equal: Improved theoretical and measurement precision. *Journal of Applied Psychology, 102,* 1022-1053.

Knobloch-Westerwick, S., Glynn, C. J., & Huge, M. (2013). The Matilda effect in sciencecommunication: an experiment on gender bias in publication quality perceptions and collaboration interest. *Science Communication*, *35,* 603-625.

Koberg, C. S., Boss, R. W., & Goodman, E. (1998). Factors and outcomes associated with mentoring among health-care professionals. *Journal of Vocational Behavior, 53*, 58-72.

Kossek, E. E., Su, R., & Wu, L. (2017). "Opting Out" or "Pushed Out"? Integrating Perspectives on Women's Career Equality for Gender Inclusion and Interventions. *Journal of Management*, *43,* 228-254.

Lakatos, I. (1976). Falsification and the methodology of scientific research programmes. In S. G.Harding (Ed.), *Can theories be refuted? Essays on the Duhem-Quine thesis (pp. 205-29)*. UK: Cambridge University Press.

Leavitt, K., Mitchell, T. R., & Peterson, J. (2010). Theory pruning: Strategies to reduce our dense theoretical landscape. *Organizational Research Methods, 13,* 644-667.

Legewie, J. & DiPrete, T. A. (2012). High School Environments, STEM Orientations, and the Gender Gap in Science and Engineering Degrees. Available at SSRN: https://ssrn.com/abstract=2008733 or http://dx.doi.org/10.2139/ssrn.2008733

Lockwood, P. (2006). "Someone like me can be successful": Do college students need same gender role models?. *Psychology of Women Quarterly*, *30,* 36-46.

Lortie, C. J., Aarssen, L. W., Budden, A. E., Koricheva, J. K., Leimu, R., & Tregenza, T. (2007). Publication bias and merit in ecology. *Oikos*, *116*, 1247-1253.

Lubinski, D., & Benbow, C. P. (2006). Study of mathematically precocious youth after 35 years: Uncovering antecedents for the development of math-science expertise. *Perspectives on psychological science, 1*, 316-345.

Lyness, K. S., & Heilman, M. E. (2006). When fit is fundamental: performance evaluations and promotions of upper-level female and male managers. *Journal of Applied Psychology*, *91,* 777-785.

Lynn, R. (1994). Sex differences in intelligence and brain size: A paradox resolved. *Personality and individual differences*, *17*, 257-271.

Makel, M. C., Wai, J., Peairs, K., & Putallaz, M. (2016). Sex differences in the right tail of cognitive abilities: An update and cross cultural extension. *Intelligence*, *59,* 8-15.

Malhotra, P., & Singh, M. (2016). Indirect impact of high performers on the career advancement of their subordinates. *Human Resource Management Review*, *26*, 209-226.

Marx, D. M., & Roman, J. S. (2002). Female role models: Protecting women’s math test performance. *Personality and Social Psychology Bulletin, 28*, 1183-1193.

Minbashian, A., & Luppino, D. (2014). Short-term and long-term within-person variability in performance: An integrative model. *Journal of Applied Psychology, 99*, 898-914.

Morgan, S. L., Gelbgiser, D., & Weeden, K. A. (2013). Feeding the pipeline: Gender, occupational plans, and college major selection. *Social Science Research*, *42*, 989-1005.

Moss-Racusin, C. A., van der Toorn, J., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2016). A “Scientific Diversity” Intervention to Reduce Gender Bias in a Sample of Life Scientists. *CBE-Life Sciences Education*, *15*, 1-11.

Ng, C. W., & Chiu, W. C. (2001). Managing equal opportunities for women: Sorting the friends from the foes. *Human Resource Management Journal*, *11*, 75-88.

Nirei, M., & Souma, W. (2007). A two factor model of income distribution dynamics. *Review of Income and Wealth, 53*, 440-459.

O’Boyle, E., & Aguinis, H. (2012). The best and the rest: Revisiting the norm of normality of individual performance. *Personnel Psychology, 65*, 79-119.

Olitsky, N. H. (2014). How do academic achievement and gender affect the earnings of STEM majors? A propensity score matching approach. *Research in Higher Education*, *55,* 245-271.

Oppler, S. H., Campbell, J. P., Pulakos, E. D., & Borman, W. C. (1992). Three approaches to the investigation of subgroup bias in performance measurement: Review, results, and conclusions. *Journal of Applied Psychology, 77*, 201-217.

Park, G., Lubinski, D., & Benbow, C. P. (2008). Ability differences among people who have commensurate degrees matter for scientific creativity. *Psychological Science*, *19*, 957 -961.

Perez-Felkner, L., McDonald, S. K., Schneider, B., & Grogan, E. (2012). Female and male adolescents' subjective orientations to mathematics and the influence of those orientations on postsecondary majors. *Developmental psychology*, *48*, 1658.

Picho, K., Rodriguez, A., & Finnie, L. (2013). Exploring the moderating role of context on the mathematics performance of females under stereotype threat: A meta-analysis. *The Journal of social psychology, 153*, 299-333.

Pierce, J. R., & Aguinis, H. (2013). The too-much-of-a-good-thing effect in management. *Journal of Management, 39*, 313-338.

Pierce, C. A., Aguinis, H., & Adams, S. K. R. (2000). Effects of a dissolved workplace romance and rater characteristics on responses to a sexual harassment accusation. *Academy of Management Journal, 43*, 869-880.

Popper, K. R. (1959). The logic of scientific discovery. London: Hutchinson.

Price, J. (2010). The effect of instructor race and gender on student persistence in STEM fields. *Economics of Education Review*, *29*, 901-910.

Reuben, E., Sapienza, P., & Zingales, L. (2014). How stereotypes impair women’s careers in science. *Proceedings of the National Academy of Sciences*, *111*, 4403-4408.

Rinne, H. (2008). *The Weibull distribution: A handbook*. Boca Raton, FL: CRC Press.

Robertson, K. F., Smeets, S., Lubinski, D., & Benbow, C. P. (2010). Beyond the threshold hypothesis: Even among the gifted and top math/science graduate students, cognitive abilities, vocational interests, and lifestyle preferences matter for career choice, performance, and persistence. *Current Directions in Psychological Science, 19*, 346-351.

Robnett, R. D., Chemers, M. M., & Zurbriggen, E. L. (2015). Longitudinal associations among undergraduates' research experience, self-efficacy, and identity. *Journal of Research in Science Teaching*, *52*, 847-867.

Rudman, L. A., & Phelan, J. E. (2010). The effect of priming gender roles on women’s implicit gender beliefs and career aspirations. *Social Psychology*, *41*, 192-202.

Sarsons, H. (2017). *Recognition for group work* (Working Paper). Retrieved from Harvard University website *https://scholar.harvard.edu/files/sarsons/files/groupwork_full_v6.pdf*

Schwartz, S. H., & Rubel, T. (2005). Sex differences in value priorities: cross-cultural and multimethod studies. *Journal of personality and social psychology*, *89*, 1010-1028.

Spear, D., & Chown, S. L. (2008). Taxonomic homogenization in ungulates: Patterns and mechanisms at local and global scales. *Journal of Biogeography*, *35*, 1962-1975.

Speliotes, E. K., Willer, C. J., Berndt, S. I., Monda, K. L., Thorleifsson, G., Jackson, A. U.,...& Randall, J. C. (2010). Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature genetics*, *42*, 937-948.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of experimental social psychology*, *35*, 4-28.

Stauffer, J. M., & Buckley, M. R. (2005). The existence and nature of racial bias in supervisory ratings. *Journal of Applied Psychology*, *90*, 586-591.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of personality and social psychology*, *69*, 797.

Steffens, M. C., & Jelenec, P. (2011). Separating implicit gender stereotypes regarding math and language: Implicit ability stereotypes are self-serving for boys and men, but not for girls and women. *Sex Roles*, *64*, 324-335.

Stoet, G., & Geary, D. C. (2012). Can stereotype threat explain the gender gap in mathematics performance and achievement?. *Review of General Psychology*, *16*, 93-102.

Su, R., Rounds, J., & Armstrong, P. I. (2009). Men and things, women and people: a meta analysis of sex differences in interests. *Psychological Bulletin*, *135*, 859-884.

Tenenbaum, H. R., & Leaper, C. (2003). Parent-child conversations about science: The socialization of gender inequities? *Developmental Psychology*, *39*, 34-47.

Trix, F., & Psenka, C. (2003). Exploring the color of glass: Letters of recommendation for female and male medical faculty. *Discourse & Society*, *14*, 191-220.

Viswesvaran, C., & Ones, D. S. (2000). Perspectives on models of job performance. *International Journal of Selection and Assessment, 8,* 216-226.

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. *Psychological Bulletin, 117,* 250 -270.

Wai, J., Cacchio, M., Putallaz, M., & Makel, M. C. (2010). Sex differences in the right tail of cognitive abilities: A 30year examination. *Intelligence, 38,* 412-423.

Wang, M. T., & Degol, J. (2013). Motivational pathways to STEM career choices: Using expectancy–value perspective to understand individual and gender differences in STEM fields. *Developmental Review, 33,* 304-340.

Wang, M. T., Eccles, J. S., & Kenny, S. (2013). Not lack of ability but more choice individual and gender differences in choice of careers in science, technology, engineering, and mathematics. *Psychological Science, 24,* 770-775.

Watkins, J. M., Cool, V. A., Usner, D., Stehbens, J. A. Nichols, S. Loveland, K. A., Bordeaux, J.D., Donfield, S., Asarnow, R.F., & Nuechterlein, K.H. (2000). Attention in HIV infected children: Results from the Hemophilia Growth and Development Study. *Journal of the International Neuropsychological Society, 6,* 443-454.

Wenneras, C. & Wold, A. (1997) Nepotism and sexism in peer-review. *Nature, 387*, 341–343.

Whetzel, D. L., McDaniel, M. A., Yost, A. P., & Kim, N. (2010). Linearity of Personality Performance Relationships: A large-scale examination. *International Journal of Selection and Assessment, 18,* 310-320.

Williams, W. M., & Ceci, S. J. (2015). National hiring experiments reveal 2: 1 faculty preference for women on STEM tenure track. *Proceedings of the National Academy of Sciences, 112,* 5360-5365.

Xie, Y., Shauman, K. A., & Shauman, K. A. (2003). *Women in science: Career processes and outcomes*. Cambridge, MA: Harvard University Press.

Xu, Y. J. (2008). Gender disparity in STEM disciplines: A study of faculty attrition and turnover intentions. *Research in Higher Education, 49,* 607-624.

Young, D. M., Rudman, L. A., Buettner, H. M., & McLean, M. C. (2013). The influence of female role models on women's implicit science cognitions. *Psychology of Women Quarterly, 37*, 283-29.

Table 1

Technical Description, Generative Mechanisms, and Examples of Empirical Research Illustrating the Seven Distributions

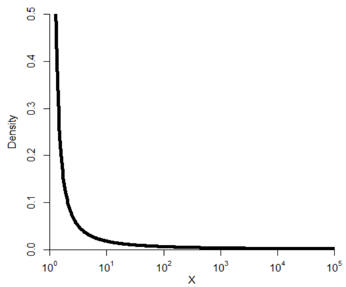
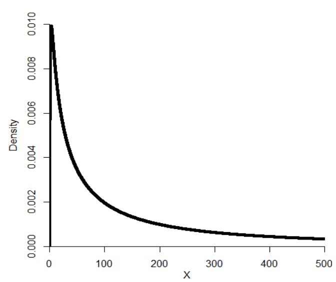
Distribution & Visual Representation	**Technical Description**	**Shape, Generative Mechanism, and Examples**
Pure power law 	A set of values from a variable, or x, follows a pure power law if: $$p(x) \propto x^{-\alpha}$$ where alpha (α) (> 1) is the rate of decay, or how quickly the distribution's right tail falls. So, the lower the value of α (closer to 1), the heavier is the distribution's right tail. For example, a distribution where α = 2 has a heavier right tail compared to a distribution where α = 3.	*Shape*: A long head and often a very heavy (i.e., infinite) right tail. Out of the seven distributions, the pure power law has the heaviest right tail. *Generative mechanism*: Self-organized criticality, a process where some individuals reach 'critical states' triggered by an interaction of events that can subsequently produce rapid emergence of extreme outcomes. In other words, individuals differ in terms of total output because some individuals experience 'output shocks,' i.e., unpredictable and extremely large output increases. *Examples*: In physics, the distribution of size of sand avalanches has been shown to follow a pure power law if there is enough tension (i.e., gravity) and connectivity (i.e., irregularity in the shapes of sand grains) (Bak, 1996). As another example, a scientist may discover a set of findings, which subsequently helps the scientist rapidly discover a much larger set of findings.
Lognormal 	A set of values from a variable, or x, follows a lognormal distribution if: $$p(x) \propto e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}$$ where Euler's number $e \approx 2.718$. $\ln(x)$ is the natural log of x and is normally distributed. Mu (μ) (> 0) is the mean. Sigma (σ) (> 0) is the standard deviation. μ does not affect the heaviness of the distribution's right tail. In contrast, σ does. The higher the value of σ (further away from 0), the heavier is the distribution's right tail	*Shape*: Bell-shaped head and a heavy but ultimately finite right tail. The lognormal distribution has a bell-shaped head (not simply a long one) and a heavier right tail compared to the exponential distribution. *Generative mechanism*: Proportionate differentiation, which is a process where an individual's future output is a distinct percentage of prior output. In other words, future output equals the product between a growth rate and prior output. Because individuals initially differ on the growth rate and also prior output, output increases at an increasing rate for some individuals. As such, a small proportion of individuals experience output loops (i.e., increasingly larger output increases based on positive feedback between past and future output). *Examples*: Research in biology has provided evidence that proportionate growth leads to a lognormal distribution of populations per species in a community (e.g., a forest) (Hubbell, 1979). In computer science, proportionate growth has been shown to result in a lognormal distribution of file sizes on the Internet (Downey, 2001, 2005).

Table 1 (continued)

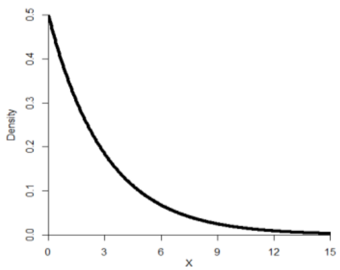
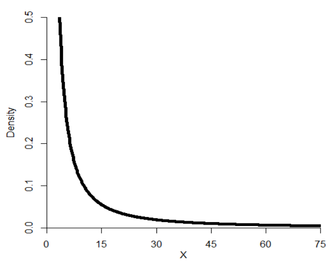
Distribution & Visual Representation	**Technical Description**	**Shape, Generative Mechanism, and Examples**
Exponential 	A set of values from a variable, or x, follows an exponential distribution if: $$p(x) \propto e^{-\lambda x}$$ where Euler's number $e \approx 2.718$. Lambda (λ) (> 0) is the rate of decay, or how quickly the distribution's right tail falls. So, the lower the value of λ (closer to 0), the heavier is the distribution's right tail.	*Shape*: Long head and a somewhat heavy (i.e., fat) right tail. Also, the top performers tend to be similar in terms of output (i.e., low variability in the right tail). *Generative mechanism*: Incremental differentiation, which is a process where each individual's value on an outcome variable increases at a distinct and relatively linear (i.e., constant) rate. In the simplest case, each individual's performance is a linear equation containing one predictor variable with a distinct slope value. As such, individuals differ in terms of total value on an outcome because of their differences with respect to the accumulation rate on the outcome, i.e., output generated per opportunity to produce, which has linear effects on their output levels. *Examples:* Research in economics has documented that people's wages accumulate at different linear rates as a result of heterogeneity in labor productivity across individuals, leading to an exponential distribution of cumulative wages (Nirei & Souma, 2007).
Power law with exponential cutoff 	A set of values from a variable, or x, follows a power law with exponential cutoff if: $$p(x) \propto x^{-\alpha} e^{-\lambda x}$$ where Euler's number $e \approx 2.718$. Both alpha α (> 1) and lambda λ (> 0) are rates of decay, or how quickly the distribution's right tail "falls". So, the lower the values of α (i.e., closer to 1) and λ (i.e., closer to 0), the heavier is the distribution's right tail. Between the two rates of decay, λ is "stronger" in terms of making the distribution's right tail fall.	*Shape*: A long head and a heavy yet increasingly decaying right tail. Depending on its parameter values, the distribution's right tail may range from being as heavy as that of a lognormal distribution to being even lighter than that of an exponential distribution *Generative mechanism* Incremental differentiation. In this case however, individuals with the highest accumulation rates are subject to diminishing returns after reaching a certain high output level. *Examples*: Research in network science has documented that nodes (e.g., airports) incur various costs when they make additional links with other nodes (e.g., new departures to other airports). Thus, after a certain point, nodes that accumulate links more quickly than other vertices would nonetheless be subject to diminishing returns, giving rise to a distribution of links per nodes that follows a power law with an exponential cutoff (Amaral et al., 2000).

Table 1 (continued)

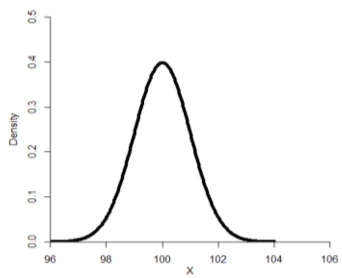
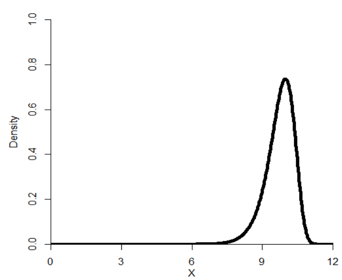
Distribution & Visual Representation	Technical Description	Shape, Generative Mechanism, and Examples
Normal (Gaussian) 	A set of values from a variable, or x, follows a normal distribution if: $$p(x) \propto e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$ where Euler's number $e \approx 2.718$. Mu μ (> 0) is the mean. Sigma σ (> 0) is the standard deviation. μ does not affect the lightness (i.e., thinness) of the symmetric tails. In contrast, σ does. The lower the value of σ (i.e., closer to 0), the lighter are the distribution's symmetric tails.	*Shape*: Bell-shaped body and symmetric tails, which quickly become light (i.e., thin) along values further from the mean. Only the normal distribution does not have a noticeable skew compared to the other distributions. *Generative mechanism*: Homogenization, defined as a process that reduces differences among individuals (e.g., individuals) in terms of their values on a variable X (e.g., performance) over time. In other words, individuals undergo output homogenization that reduces future differences in individual output. *Examples*: In zoology, the homogenization processes of various species are characterized by a normal distribution (Spear & Chown, 2008). Also, in any occupation where workers are subject to situational constraints that act as output floors and/or ceilings would homogenize future output.
Weibull 	A set of values from a variable, or x, follows a Weibull distribution if: $$p(x) \propto \left({x}/{\lambda}\right)^{\beta-1} e^{-\left({x}/{\lambda}\right)^{\beta}}$$ where Euler's number $e \approx 2.718$. Beta β (> 0) is the extent to which the distribution is "pulled" up and to the right. So, the lower the value of β (i.e., closer to 0), the lower is the height of the bell-shaped head and heavier is the right tail. Lambda λ (> 0) is the extent to which the distribution is "pushed" down and to the right. So, the lower the value of lambda λ (i.e., closer to 0), the higher is the height of the bell-shaped head and lighter is the right tail.	*Shape*: Bell-shaped head and a slight right- or left-skew. The Weibull distribution has a wider bell-shaped head but a lighter right tail than the lognormal distribution. *Generative mechanism*: Homogenization. This distribution is generated by a type of homogenization process in where an individual (e.g., a machine) consisting of many components increases in the outcome variable (e.g., longevity) until the failure of a component, or the "weakest link." Upon such failure, further increases in the outcome variable will slow down drastically or simply halt, reducing future variability in output. *Examples*: The lifetime of roller bearings, time to death after exposure to carcinogenesis, and duration of industrial stoppages (Rinne, 2008).

Table 1 (continued)

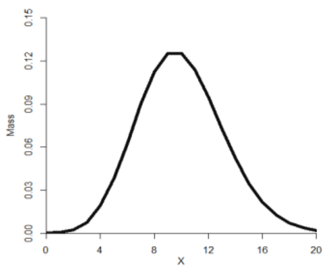
Distribution & Visual Representation	Technical Description	Shape, Generative Mechanism, and Examples
Poisson 	A set of values from a variable, or x, follows a Poisson distribution if: $$p(x) \propto \mu^x/x!$$ where mu μ (> 0) is the mean, which also equals the variance of the distribution. The lower the value of μ (i.e., closer to 0), the heavier is the distribution's right tail.	*Shape*: A bell-shaped head, and possibly a somewhat heavy but ultimately finite right tail consisting of relatively low (i.e., infrequent) counts. Unlike the other distributions, the Poisson distribution can only model discrete (not continuous) variables and, therefore, has a "jagged" curve. *Generative mechanism*: Homogenization. In this case, infrequent events (i.e., non-negative integers) accumulate within a limited time period. In doing so, random fluctuations around a low mean number of events produce a heavy right tail. *Examples*: The Poisson distribution has been used to model various phenomena characterized by infrequent events, such as the number of correct answers to a cognitive impairment test by HIV-infected children (Watkins et al., 2000) and gold content (measured by parts per million) in a soil or sediment sample (Fletcher, 1981).

Table 2

Implications of Sex Differences in Quantitative Aptitudes on the Four Possible Dominant Generative Mechanisms of Individual Performance

Self-organized criticality	**Proportionate differentiation**	**Incremental differentiation**	**Homogenization**
Explanation of the mechanism: Individuals differ in their STEM performance largely due to a small proportion of individuals who experience large output shocks after reaching self-organized criticality. *Possible gender differences:* The performance distribution for women may have a lighter right tail. Accordingly, there may be a disproportionately large gender performance gap among star performers in favor of men. *Reasoning:* Reaching a critical state involves the interconnection of many performance components, and the interaction of many events and variables. This should include innate quantitative aptitudes. Women have been shown to be less variable in their quantitative aptitudes. Thus, assuming that quantitative aptitudes are an important driver of self-organized criticality, there should be a gender performance gap in favor of star men.	*Explanation of the mechanism:* Individuals differ in their STEM performance largely due to some individuals enjoying larger output loops than others. Output loops occur when an individual's prior output value is linked to future output through positive feedback with his/her accumulation rate. *Possible gender differences:* The performance distribution for women likely has a lighter right tail due to less variability in women's accumulation rates and/or initial output values. Accordingly, there may be a disproportionately large gender performance gap among star performers in favor of men. *Reasoning*: It is possible that sex differences in STEM aptitudes contribute to a smaller variance in women's accumulation rates (e.g., average number of articles published per year) and/or their initial output values (e.g., number of articles published during the first year of one's tenure track career), thus producing performance distribution with a lighter right tail for women.	*Explanation of the mechanism:* Individuals differ in their STEM performance largely due to some individuals enjoying larger output increments than others. Output increments refer to linear increments in one's total outputs based on his/her accumulation rate. *Possible gender differences:* The performance distribution for women likely has a lighter right tail, due to less variability in women's accumulation rates. Accordingly, there may be a disproportionately large gender performance gap among star performers in favor of men. *Reasoning*: Again, it is possible that sex differences in quantitative aptitudes contribute to a smaller variance in women's accumulation rates (e.g., average number of articles published per year). Assuming that quantitative aptitudes are an important driver of accumulation rates (i.e., average productivity per time period), there should less variance in women's accumulation rates, and the performance distribution for women should have a lighter right tail.	*Explanation of the mechanism:* Individuals undergo output homogenization over time, and variability in past individual outputs would be followed by lower variability in future. *Possible gender differences:* Sex differences in quantitative aptitudes do not suggest that output homogenization would operate differently for men versus women. Nonetheless, women's STEM performance could be more homogenized than men's, and there may be a disproportionately large gender performance gap among star performers in favor of men. *Reasoning*: Women may be more homogenized than men in terms of their innate STEM aptitudes (due to lower variability), but this does not imply that they are subject to output homogenization processes that do not simultaneously apply to men (e.g., promotion practices that act as a 'floor' with regard to future output). However, assuming that quantitative aptitudes remain an important driver of STEM performance, women's performance may be more homogenized than men's at any given point in time.

Table 3
Implications of Gender Discrimination on the Four Possible Dominant Generative Mechanisms of Individual Performance

Self-organized criticality	**Proportionate differentiation**	**Incremental differentiation**	**Homogenization**
Explanation of the mechanism: Individuals differ in their STEM performance largely due to a small proportion of individuals who experience large output shocks after reaching self-organized criticality. *Possible gender differences:* The performance distribution for women could have a lighter right tail. Accordingly, there may be a disproportionately large gender performance gap among star performers in favor of men. *Reasoning*: Achieving a critical state involves a complex interaction of many events and variables. For researchers, peer evaluations of one's work, access to research-related resources, and availability of collaboration opportunities are all likely important variables with respect to enabling self-organized criticality. It is thus possible that gender discrimination in areas such as work-product evaluations, journal reviews, and grant funding substantially constrain/cripple talented female researchers' potential to achieve self-organized criticality.	*Explanation of the mechanism:* Individuals differ in their STEM performance largely due to some individuals enjoying larger output loops than others. Output loops occur when an individual's prior output value is linked to future output through positive feedback with his/her accumulation rate. *Possible gender differences:* The performance distribution for women may have a lighter right tail due to specific gender differences in accumulation rates and/or initial output values caused by discrimination. *Reasoning:* Gender discrimination may constrain women's accumulation rates and/or initial output values, possibly with a greater (negative) cumulative effect for star women. It is also possible that the strength of the positive interactions between prior outputs and accumulation rates are weaker for women compared to men due to certain forms of discrimination (e.g., undervaluing women's past achievements when choosing potential collaborators).	*Explanation of the mechanism:* Individuals differ in their STEM performance largely due to some individuals enjoying larger output increments than others. Output increments refer to linear increments in one's total outputs based on his/her accumulation rate. *Possible gender differences*: The performance distribution for women may have a lighter right tail due to specific gender differences in accumulation rates caused by discrimination. *Reasoning:* Again, various forms of gender discrimination may constrain women's accumulation (e.g., biases in peer reviews, grant funding, etc). Assuming that gender discrimination affects the accumulation rates of all women performers at potentially every opportunity to perform, it is possible that the gender performance gap is greatest—in absolute terms—among stars.	*Explanation of the mechanism:* Individuals undergo output homogenization over time, and variability in past individual outputs would be followed by lower variability in future. *Possible gender differences*: The performance distribution for women may have a lighter right tail due to certain forms of discrimination. Accordingly, there may be a disproportionately large gender performance gap among star performers in favor of men. *Reasoning*: The gender discrimination perspective suggests that STEM fields possibly impose stronger output homogenization processes on female researchers compared to their male colleagues (e.g., overvaluing star men and/or undervaluing star women). If gender discrimination occurs in such a way that 'ceiling' effects are stronger for star women than for star men, there would likely be a lighter right hand tail for women.

Table 4

Implications of Women's Career and Lifestyle Choices on the Four Possible Dominant Generative Mechanisms

Self-organized criticality	**Proportionate differentiation**	**Incremental differentiation**	**Homogenization**
Explanation of the mechanism: Individuals differ in their STEM performance largely due to a small proportion of individuals who experience large output shocks after reaching self-organized criticality. *Possible gender differences:* The performance distribution for women could have a lighter right tail. Accordingly, there may be a disproportionately large gender performance gap among star performers in favor of men. *Reasoning*: It has been suggested that, due to women's particular occupational preferences and work/family values, women tend to opt out of STEM fields at higher rates than men. They are also more likely to make career sacrifices in pursuit of other goals. Thus, assuming that career persistence and time/effort dedicated to one's job are important factors for enabling self-organized criticality, it is possible that fewer female researchers end up reaching critical states in their careers compared to male researchers.	*Explanation of the mechanism:* Individuals differ in their STEM performance largely due to some individuals enjoying larger output loops than others. Output loops occur when an individual's prior output value is linked to future output through positive feedback with his/her accumulation rate. *Possible gender differences:* The performance distribution for women may have a lighter right tail due to smaller variance in women's accumulation rates and/or initial output values. *Reasoning*: Women's higher emphasis on work/life balance and family values suggests that there may less variance in their accumulation rates (and/or initial output values) compared to among men. Specifically, women may be disproportionately less likely to devote extreme amounts of time and other resources to their work compared to men in these fields. As such, it is likely that the distribution for women a lighter right tail, suggesting a significant gender performance gap among star performers in favor of men.	*Explanation of the mechanism:* Individuals differ in their STEM performance largely due to some individuals enjoying larger output increments than others. Output increments refer to linear increments in one's total outputs based on his/her accumulation rate. *Possible gender differences*: The performance distribution for women may have a lighter right tail due to smaller variance in women's accumulation rates. *Reasoning*: Again, there may a smaller variance for women in terms of their accumulation rates. Due to gender differences in various motivational and other psychological factors, women may be disproportionately underrepresented among the small proportion of individuals with the greatest accumulation rates. Accordingly, there may be a disproportionately large gender performance gap among star performers in favor of men.	*Explanation of the mechanism:* Individuals undergo output homogenization over time, and variability in past individual outputs would be followed by lower variability in future. *Possible gender differences*: The performance distribution for women may have a lighter right tail. Accordingly, there may be a disproportionately large gender performance gap among star performers in favor of men. *Reasoning*: It is possible that the particular choices made by women (e.g., talented women opting out of STEM fields at higher rates, productivity losses stemming from fertility decisions, increasing prioritization of family needs, etc.) result in a stronger 'ceiling' with respect to women's future outputs compared to men's. Thus, over time, there may be greater homogenization in the cumulative performance of female researchers compared to that of male researchers.

Table 5

Individual Research Productivity of Female and Male Researchers in Mathematics: Distribution Pitting Statistics

Gender	***N***	**NormvPL**	**NormvCut PLvCut**	**NormvWeib PLvWeib CutvWeib**	**NormvLogN PLvLogN CutvLogN WeibvLogN**	**NormvExp PLvExp CutvExp WeibvExp LogNvExp**	**NormvPois PLvPois CutvPois WeibvPois LogNvPois ExpvPois**
Women	360	-5.69 (0)	-6.10 (0) -3.46 (0.01)	-6.14 (0) -1.66 (0.10) 0.54 (0.59)	-6.09 (0) -1.55 (0.12) 1.08 (0.28) 0.03 (0.97)	-6.77 (0) 0.63 (0.53) 1.76 (0.08) 1.80 (0.07) 1.67 (0.10)	-7.71 (0) 1.73 (0.08) 2.42 (0.02) 2.40 (0.02) 2.40 (0.02) 2.87 (0)
Men	3,493	-16.55 (0)	-18.28 (0) -83.30 (0)	-18.18 (0) -7.72 (0) 0.31 (0.75)	-18.09 (0) -7.76 (0) 0.88 (0.38) 1.71 (0.09)	-20.25 (0) 1.49 (0.14) 6.16 (0) 6.17 (0) 5.88 (0)	-26.48 (0) 6.53 (0) 8.46 (0) 8.46 (0) 8.34 (0) 9.21 (0)

Note. N = sample size; LR = loglikelihood ratio. Distribution pitting results are presented in the final six columns of the table. For each instance of distribution pitting, the LR value is presented followed by its *p*-value in parentheses. In the first row of the table, distribution names are abbreviated: Norm = Normal, PL = Pure power law, Cut = Power law with exponential cutoff, Weib = Weibull, LogN = Lognormal, and Exp = Exponential. Distribution pitting titles are presented such that the first distribution is pitted against the second distribution (e.g., NormvPL = Normal distribution versus pure power law). Positive LR = superior fit for first distribution as listed in the distribution pitting title. Negative LR = superior fit for second distribution as listed in the distribution pitting title.

Table 6

Individual Research Productivity of Female and Male Researchers in Materials Sciences: Distribution Pitting Statistics

Gender	***N***	**NormvPL**	**NormvCut** **PLvCut**	**NormvWeib** **PLvWeib** **CutvWeib**	**NormvLogN** **PLvLogN** **CutvLogN** **WeibvLogN**	**NormvExp** **PLvExp** **CutvExp** **WeibvExp** **LogNvExp**	**NormvPois** **PLvPois** **CutvPois** **WeibvPois** **LogNvPois** **ExpvPois**
Women	5,086	-14.72 (0)	-15.00 (0) -22.96 (0)	-15.11 (0) -1.47 (0.14) 2.26 (0.02)	-14.97 (0) -4.07 (0) -0.10 (0.92) -2.01 (0.04)	-15.98 (0) 7.15 (0) 8.26 (0) 8.76 (0) 8.25 (0)	-19.44 (0) 8.06 (0) 8.43 (0) 8.57(0) 8.41 (0) 8.37 (0)
Men	30,556	-29.47 (0)	-30.08 (0) -232.87 (0)	-30.13 (0) -6.26 (0) 1.45 (0.15)	-29.98 (0) -12.85 (0) -3.35 (0) -4.16 (0)	-30.93 (0) 17.06 (0) 20.19 (0) 20.62 (0) 20.14 (0)	-35.82 (0) 18.63 (0) 19.38 (0) 19.48 (0) 19.34 (0) 18.90 (0)

Note. N = sample size; LR = loglikelihood ratio. Distribution pitting results are presented in the final six columns of the table. For each instance of distribution pitting, the LR value is presented followed by its *p*-value in parentheses. In the first row of the table, distribution names are abbreviated: Norm = Normal, PL = Pure power law, Cut = Power law with exponential cutoff, Weib = Weibull, LogN = Lognormal, and Exp = Exponential. Distribution pitting titles are presented such that the first distribution is pitted against the second distribution (e.g., NormvPL = Normal distribution versus pure power law). Positive LR = superior fit for first distribution as listed in the distribution pitting title. Negative LR = superior fit for second distribution as listed in the distribution pitting title.

Table 7

Individual Research Productivity of Female and Male Researchers in Genetics: Distribution Pitting Statistics

Gender	***N***	**NormvPL**	**NormvCut PLvCut**	**NormvWeib PLvWeib CutvWeib**	**NormvLogN PLvLogN CutvLogN WeibvLogN**	**NormvExp PLvExp CutvExp WeibvExp LogNvExp**	**NormvPois PLvPois CutvPois WeibvPois LogNvPois ExpvPois**
Women	14,685	-17.69 (0)	-17.78 (0) -23.78 (0)	-17.81 (0) 5.25 (0) 7.88 (0)	-17.76 (0) -4.29 (0) -0.41 (0.68) -7.89 (0)	-17.97 (0) 13.60 (0) 14.11 (0) 14.40 (0) 14.07 (0)	-21.97 (0) 11.41 (0) 11.50 (0) 11.52 (0) 11.49 (0) 10.66 (0)
Men	30,322	-40.29 (0)	-40.83 (0) -126.91 (0)	-41.17 (0) 3.33 (0) 8.37 (0)	-40.78 (0) -9.33 (0) -0.38 (0.70) -10.36 (0)	-43.59 (0) 22.91 (0) 24.62 (0) 26.02 (0) 24.68 (0)	-43.64 (0) 21.60 (0) 21.93 (0) 22.11 (0) 21.93 (0) 20.98 (0)

Note. *N* = sample size; LR = loglikelihood ratio. Distribution pitting results are presented in the final six columns of the table. For each instance of distribution pitting, the LR value is presented followed by its *p*-value in parentheses. In the first row of the table, distribution names are abbreviated: Norm = Normal, PL = Pure power law, Cut = Power law with exponential cutoff, Weib = Weibull, LogN = Lognormal, and Exp = Exponential. Distribution pitting titles are presented such that the first distribution is pitted against the second distribution (e.g., NormvPL = Normal distribution versus pure power law). Positive LR = superior fit for first distribution as listed in the distribution pitting title. Negative LR = superior fit for second distribution as listed in the distribution pitting title.

Table 8
Summary of Results

STEM domain	Gender	Distribution pitting results					Other Indicators		
		Distributions remaining after decision rule #1	Distributions remaining after decision rule #2	Distributions remaining after decision rule #3	Dominant generative mechanism	Fit parameters	Total range in individual's cumulative output (i.e. total # of articles)	Percentage of total output accounted for by the top 10%, 5%, and 1% of performers, respectively	Total range of publications among the top 10%, 5%, and 1% of performers, respectively
Mathematics	Female (N = 360)	PL w/cutoff, Weibull, and Lognormal	PL w/cutoff, Weibull, and Lognormal	PL w/cutoff	Incremental differentiation	α = 2.94, λ = 0.57	1 to 7	23.7%, 15.2%, and 4.9%	1 to 7, 3 to 7, and 5 to 7
	Male (N = 3,493)	PL w/cutoff and Weibull	PL w/cutoff and Weibull	PL w/cutoff	Incremental differentiation	α = 2.39, λ = 0.47	1 to 20	29.7%, 18.9%, and 6.3%,	3 to 20, 4 to 20, and 8 to 20
Material sciences	Female (N = 5,086)	PL w/cutoff and Lognormal	PL w/cutoff and Lognormal	PL w/cutoff	Incremental differentiation	α = 2.59, λ = 0.49, lognormal SD = 0.57	1 to 36	32.1%, 21.8%, and 8.1%	3 to 36, 4 to 36, and 8 to 26
	Male (N = 30,556)	Lognormal	Lognormal	Lognormal	Proportionate differentiation	α = 2.36, λ = 0.44, lognormal SD = 0.65	1 to 91	35.9%, 25.1%, and 10.1%	3 to 91, 5 to 91, and 11 to 91
Genetics	Female (N = 14,685)	PL w/cutoff and Lognormal	PL w/cutoff and Lognormal	PL w/ cutoff	Incremental differentiation	α = 2.43, λ = 0.44	1 to 123	38.9%, 28.3%, and 12.4%	3 to 123, 5 to 123, and 12 to 123
	Male (N = 30,322)	PL w/cutoff and Lognormal	PL w/cutoff and Lognormal	PL w /cutoff	Incremental differentiation	α = 2.30, λ = 0.40	1 to 102	40.8, 29.6%, and 12.6%	4 to 102, 6 to 102, and 15 to 102

Note. N = sample size; PL w/cutoff = Power law with exponential cutoff. Alpha (α) and lambda (λ) values represent the rates of decay and reflect the distributions' right tail heaviness. The lower the values of α (i.e., closer to 1) and λ (i.e., closer to 0), the heavier is the distribution's right tail.

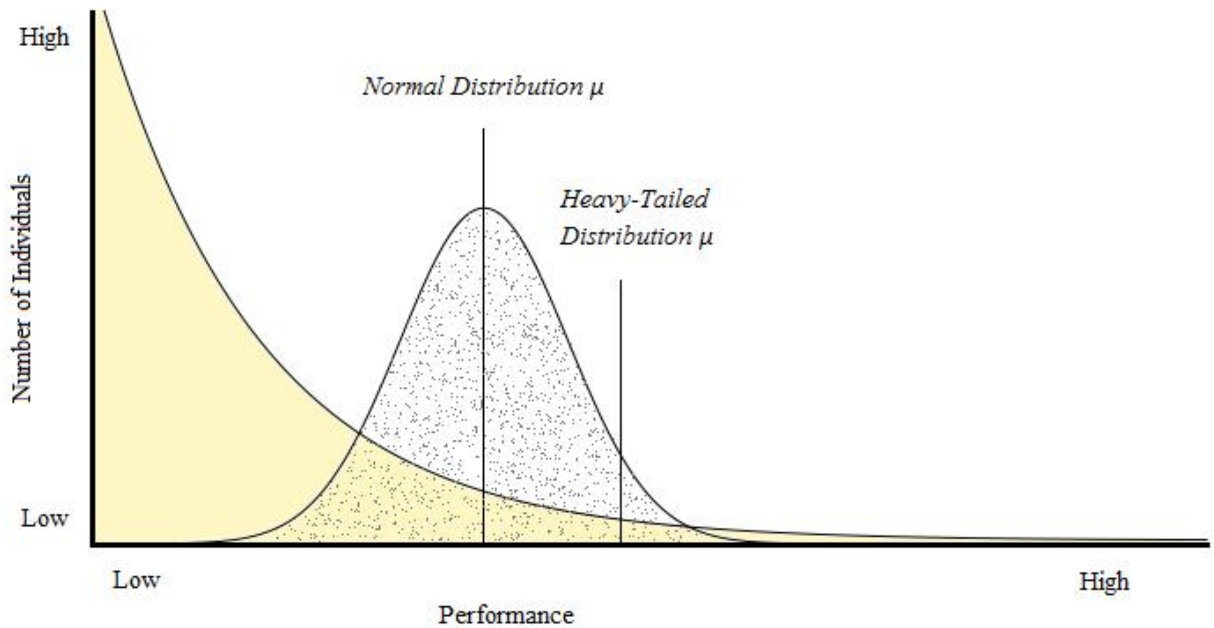


Figure 1. Generic Normal Distribution (dotted grey) overlaying a heavy-tailed distribution (yellow).

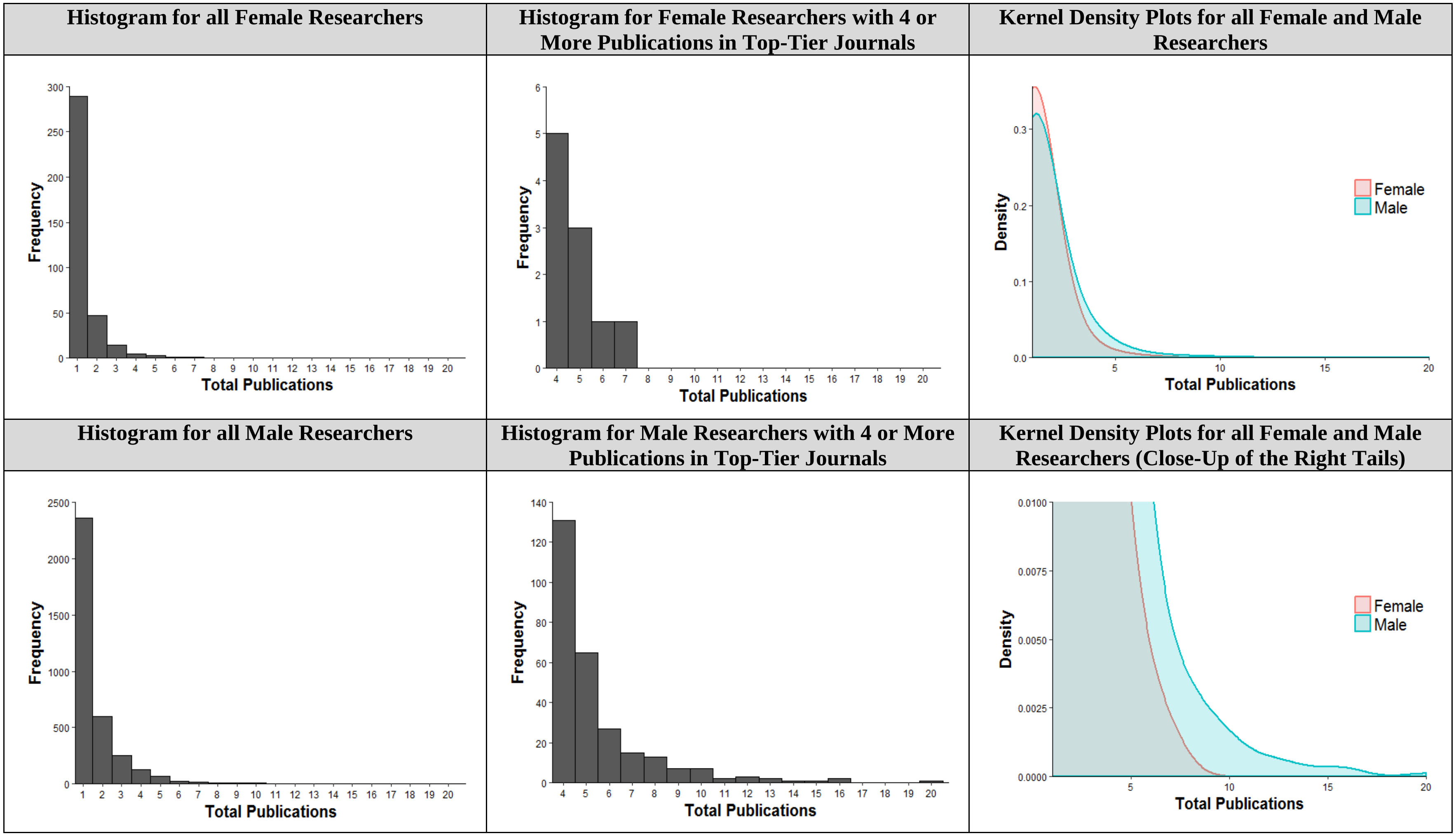


Figure 2. Histograms and Kernel Density Plots of the Research Productivity of Female and Male Researchers in Mathematics

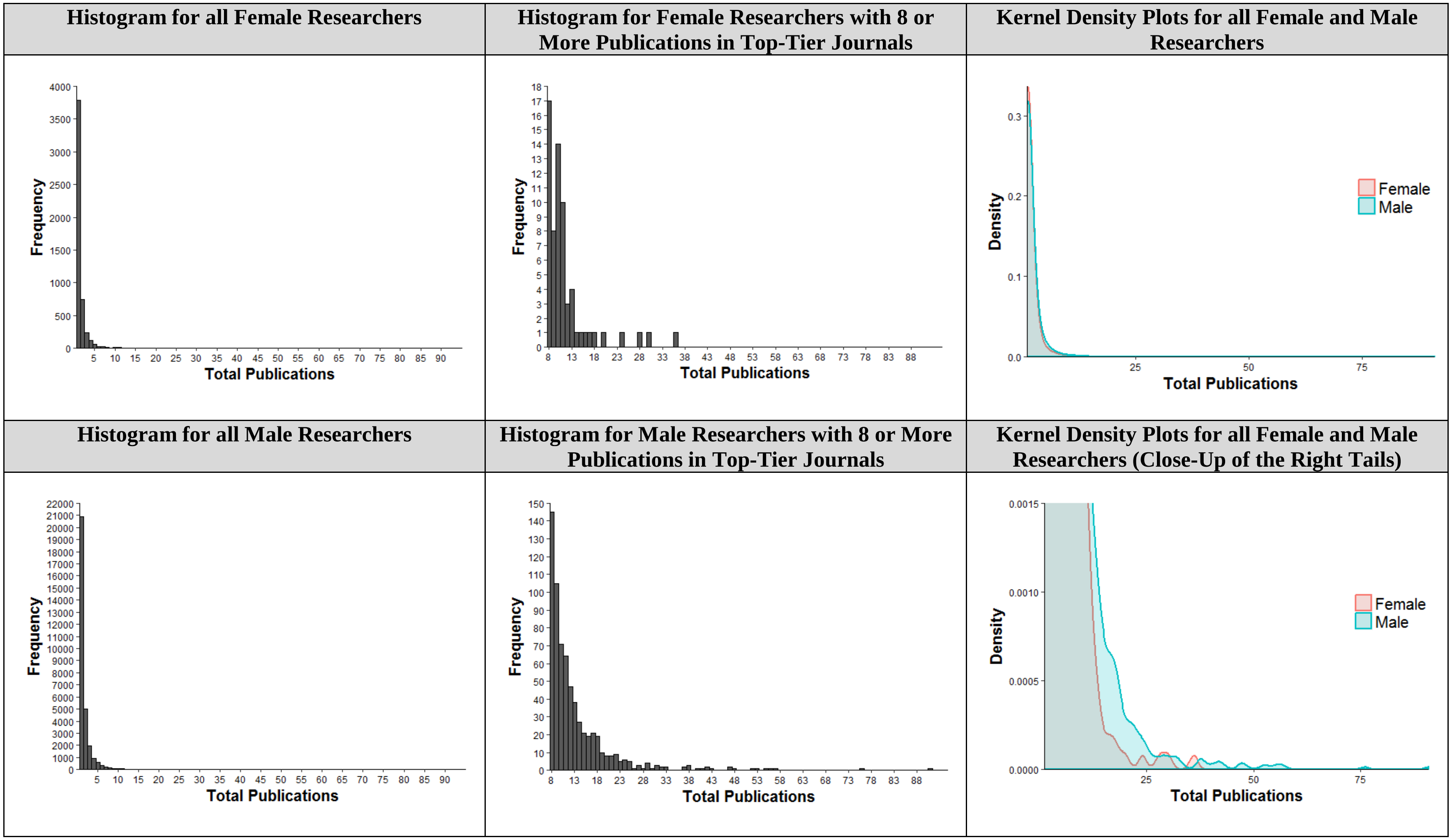


Figure 3. Histograms and Kernel Density Plots of the Research Productivity of Female and Male Researchers in Materials Sciences

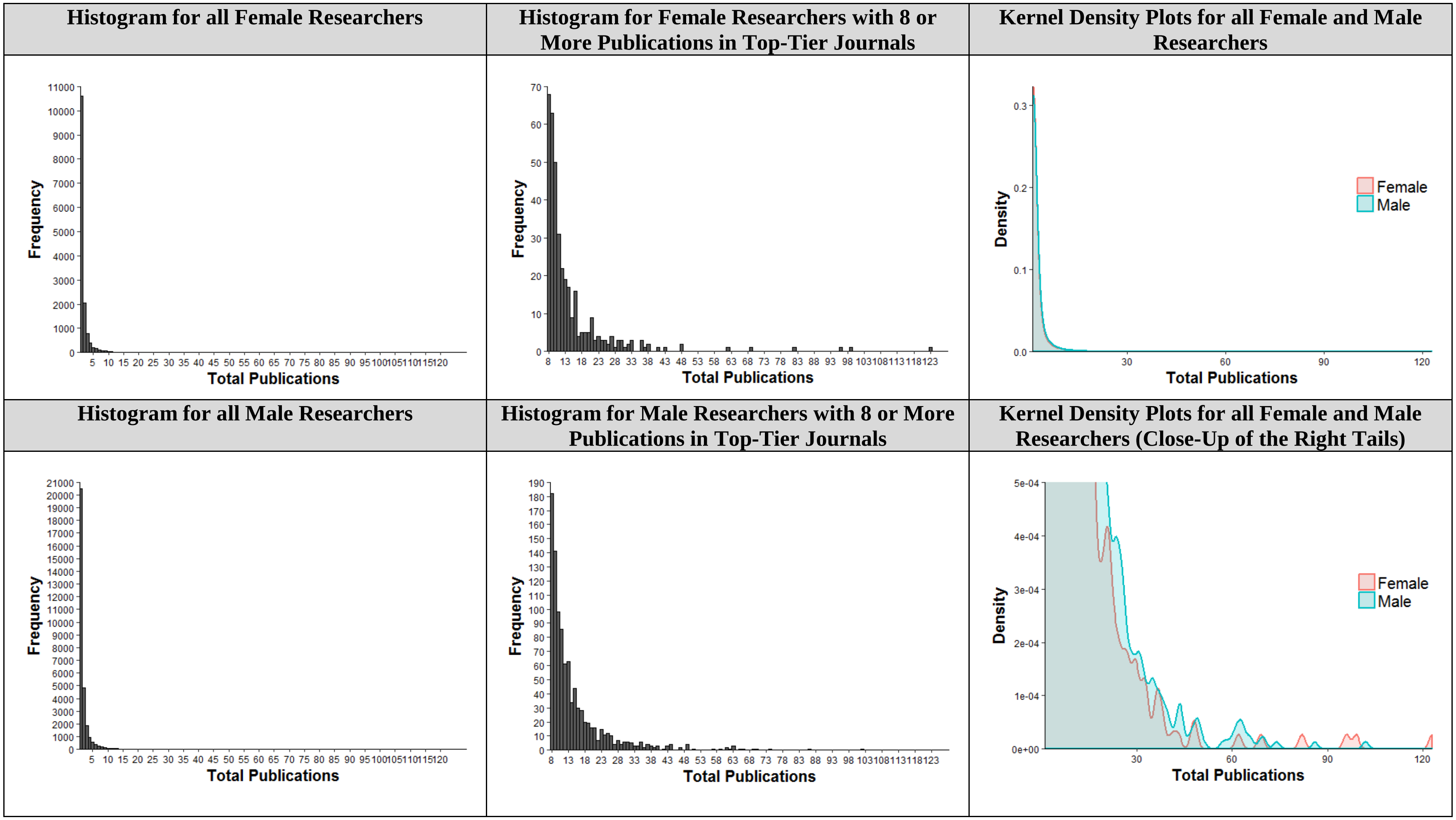


Figure 4. Histograms and Kernel Density Plots of the Research Productivity of Female and Male Researchers in Genetics

Appendix A: R Codes Used for Distribution Pitting, Fit Parameter Estimations, and Histogram and Kernel Density Plots for All Three Studies

```
#Init Libs
library(Dpit)
library(VGAM)
library(data.table)
library(moments)
library(Rcpp)
library(gsl)
library(fitdistrplus)

#Reading datasets
s1fpubs <- read.csv(file.choose(), header = TRUE)
s1mpubs <- read.csv(file.choose(), header = TRUE)
s2fpubs <- read.csv(file.choose(), header = TRUE)
s2mpubs <- read.csv(file.choose(), header = TRUE)
s3fpubs <- read.csv(file.choose(), header = TRUE)
s3mpubs <- read.csv(file.choose(), header = TRUE)

#Distribution pitting
outs1f <- Dpit(s1fpubs)
outs1m <- Dpit(s1mpubs)
outs2f <- Dpit(s2fpubs)
outs2m <- Dpit(s2mpubs)
outs3f <- Dpit(s3fpubs)
outs3m <- Dpit(s3mpubs)

#Estimating fit parameters (power law distribution's alpha)
zeta.fit(s1fpubs)
zeta.fit(s1mpubs)
zeta.fit(s2fpubs)
zeta.fit(s2mpubs)
zeta.fit(s3fpubs)
zeta.fit(s3mpubs)

#Estimating fit parameters (exponential distribution's lambda)
discexp.fit(s1fpubs)
discexp.fit(s1mpubs)
discexp.fit(s2fpubs)
discexp.fit(s2mpubs)
discexp.fit(s3fpubs)
discexp.fit(s3mpubs)

#Estimating fit parameters (lognormal distribution's SD)
```

```
fit.lnorm.disc(s2mpubs$TotalPubs)
fit.lnorm.disc(s2fpubs$TotalPubs)

#Histogram plots

#Study 1 - Women
ggplot(s1fpubs, aes(TotalPubs)) + geom_histogram(binwidth = 1, color="black") + theme(panel.background = element_blank(), axis.text = element_text(color="black"), axis.line = element_line(color = "black")) + scale_y_continuous(breaks=seq(0, 300, 50), expand = c(0,0)) + scale_x_continuous(breaks = seq(0,20,1)) + theme(axis.title = element_text(face="bold", size=15)) + labs(x= "Total Publications", y = "Frequency") + theme(axis.title = element_text(face="bold", size=15)) + coord_cartesian(ylim=c(0, 300), xlim = c(1.5, 20))

#Study 1 - Men
ggplot(s1mpubs, aes(TotalPubs)) + geom_histogram(binwidth = 1, color="black") + theme(panel.background = element_blank(), axis.text = element_text(color="black"), axis.line = element_line(color = "black")) + scale_y_continuous(breaks=seq(0, 2500, 500), expand = c(0,0)) + scale_x_continuous(breaks = seq(0,20,1)) + theme(axis.title = element_text(face="bold", size=15)) + labs(x= "Total Publications", y = "Frequency") + theme(axis.title = element_text(face="bold", size=15)) + coord_cartesian(ylim=c(0, 2500), xlim = c(1.5, 20))

#Study 1 - Women (pubs = 4 or more)
ggplot(s1fpubs, aes(TotalPubs)) + geom_histogram(binwidth = 1, color="black") + theme(panel.background = element_blank(), axis.text = element_text(color="black"), axis.line = element_line(color = "black")) + scale_y_continuous(breaks=seq(0, 6, 1), expand = c(0,0)) + scale_x_continuous(breaks = seq(4,20,1)) + theme(axis.title = element_text(face="bold", size=15)) + labs(x= "Total Publications", y = "Frequency") + theme(axis.title = element_text(face="bold", size=15)) + coord_cartesian(ylim=c(0, 6), xlim = c(4.3, 20))

#Study 1 - Men (pubs = 4 or more)
ggplot(s1mpubs, aes(TotalPubs)) + geom_histogram(binwidth = 1, color="black") + theme(panel.background = element_blank(), axis.text = element_text(color="black"), axis.line = element_line(color = "black")) + scale_y_continuous(breaks=seq(0, 140, 20), expand = c(0,0)) + scale_x_continuous(breaks = seq(4,20,1)) + theme(axis.title = element_text(face="bold", size=15)) + labs(x= "Total Publications", y = "Frequency") + theme(axis.title = element_text(face="bold", size=15)) + coord_cartesian(ylim=c(0, 140), xlim = c(4.3, 20))

#Study 2 - Women
ggplot(s2fpubs, aes(TotalPubs)) + geom_histogram(binwidth = 1, color="black") + theme(panel.background = element_blank(), axis.text = element_text(color="black"), axis.line = element_line(color = "black")) + scale_y_continuous(breaks=seq(0, 4000, 500), expand = c(0,0)) + scale_x_continuous(breaks = seq(0,91,5)) + theme(axis.title = elemen
```

```
t_text(face="bold", size=15)) + labs(x= "Total Publications", y = "Frequency") + theme(
axis.title = element_text(face="bold", size=15)) + coord_cartesian(ylim=c(0, 4000), xli
m = c(4.8, 91))

#Study 2 - men
ggplot(s2mpubs, aes(TotalPubs)) + geom_histogram(binwidth = 1, color="black") + the
me(panel.background = element_blank(), axis.text = element_text(color="black"), axis.
line = element_line(color = "black")) + scale_y_continuous(breaks=seq(0, 22000, 1000)
, expand = c(0,0)) + scale_x_continuous(breaks = seq(0,91,5)) + theme(axis.title = elem
ent_text(face="bold", size=15)) + labs(x= "Total Publications", y = "Frequency") + the
me(axis.title = element_text(face="bold", size=15)) + coord_cartesian(ylim=c(0, 22000
), xlim = c(4.8, 91))

#Study 2 - Women (pubs = 8 or more)
ggplot(s2fpubs, aes(TotalPubs)) + geom_histogram(binwidth = 1, color="black") + the
me(panel.background = element_blank(), axis.text = element_text(color="black"), axis.
line = element_line(color = "black")) + scale_y_continuous(breaks=seq(0, 18, 1), expan
d = c(0,0)) + scale_x_continuous(breaks = seq(8,91,5)) + theme(axis.title = element_te
xt(face="bold", size=15)) + labs(x= "Total Publications", y = "Frequency") + theme(axis
.title = element_text(face="bold", size=15)) + coord_cartesian(ylim=c(0, 18), xlim = c(
11.5, 91))

#Study 2 - men (pubs = 8 or more)
ggplot(s2mpubs, aes(TotalPubs)) + geom_histogram(binwidth = 1, color="black") + the
me(panel.background = element_blank(), axis.text = element_text(color="black"), axis.
line = element_line(color = "black")) + scale_y_continuous(breaks=seq(0, 150, 10), exp
and = c(0,0)) + scale_x_continuous(breaks = seq(8,91,5)) + theme(axis.title = element_
text(face="bold", size=15)) + labs(x= "Total Publications", y = "Frequency") + theme(ax
is.title = element_text(face="bold", size=15)) + coord_cartesian(ylim=c(0, 150), xlim =
 c(11.5, 91))

#Study 3 - Women
ggplot(s3fpubs, aes(TotalPubs)) + geom_histogram(binwidth = 1, color="black") + the
me(panel.background = element_blank(), axis.text = element_text(color="black"), axis.
line = element_line(color = "black")) + scale_y_continuous(breaks=seq(0, 11000, 1000)
, expand = c(0,0)) + scale_x_continuous(breaks = seq(0,123,5)) + theme(axis.title = ele
ment_text(face="bold", size=15)) + labs(x= "Total Publications", y = "Frequency") + th
eme(axis.title = element_text(face="bold", size=15)) + coord_cartesian(ylim=c(0, 1100
0), xlim = c(6.5, 123))

#Study 3 - men
ggplot(s3mpubs, aes(TotalPubs)) + geom_histogram(binwidth = 1, color="black") + the
me(panel.background = element_blank(), axis.text = element_text(color="black"), axis.
line = element_line(color = "black")) + scale_y_continuous(breaks=seq(0, 21000, 1000)
, expand = c(0,0)) + scale_x_continuous(breaks = seq(0,123,5)) + theme(axis.title = ele
ment_text(face="bold", size=15)) + labs(x= "Total Publications", y = "Frequency") + th
```

```
eme(axis.title = element_text(face="bold", size=15)) + coord_cartesian(ylim=c(0, 2100
0), xlim = c(6.5, 123))

#Study 3 - Women (pubs = 8 or more)
ggplot(s3fpubs, aes(TotalPubs)) + geom_histogram(binwidth = 1, color="black") + the
me(panel.background = element_blank(), axis.text = element_text(color="black"), axis.
line = element_line(color = "black")) + scale_y_continuous(breaks=seq(0, 70, 10), expa
nd = c(0,0)) + scale_x_continuous(breaks = seq(8,123,5)) + theme(axis.title = element_
text(face="bold", size=15)) + labs(x= "Total Publications", y = "Frequency") + theme(ax
is.title = element_text(face="bold", size=15)) + coord_cartesian(ylim=c(0, 70), xlim = c
(13, 123))

#Study 3 - men (pubs = 8 or more)
ggplot(s3mpubs, aes(TotalPubs)) + geom_histogram(binwidth = 1, color="black") + the
me(panel.background = element_blank(), axis.text = element_text(color="black"), axis.
line = element_line(color = "black")) + scale_y_continuous(breaks=seq(0, 190, 10), exp
and = c(0,0)) + scale_x_continuous(breaks = seq(8,123,5)) + theme(axis.title = element
_text(face="bold", size=15)) + labs(x= "Total Publications", y = "Frequency") + theme(a
xis.title = element_text(face="bold", size=15)) + coord_cartesian(ylim=c(0, 190), xlim
= c(13, 123))


#Kernel density plots

#Reading combined datasets
s1all <- read.csv(file.choose(), header = TRUE)
s2all <- read.csv(file.choose(), header = TRUE)
s3all <- read.csv(file.choose(), header = TRUE)

#Study 1 Density plot
ggplot(s1all, aes(TotalPubs, fill=Gender, color=Gender))+geom_density(bw=1, lwd=0.8
, alpha=0.2) + theme(panel.background = element_blank(), axis.text = element_text(col
or="black"), axis.line = element_line(color = "black")) + scale_x_continuous(expand =
c(0, 0)) + scale_y_continuous(expand = c(0, 0)) + labs(x= "Total Publications", y = "De
nsity") + theme(axis.title = element_text(face="bold", size=15)) + theme(legend.text=el
ement_text(size=15), legend.title = element_text(size=15)) + theme(legend.title=eleme
nt_blank(), legend.position=c(.9,.60))

#Study 1 Density plot (close-up of the tail-end)
ggplot(s1all, aes(TotalPubs, fill=Gender, color=Gender))+geom_density(bw=1, lwd=0.8
, alpha=0.2) + theme(panel.background = element_blank(), axis.text = element_text(col
or="black"), axis.line = element_line(color = "black")) + scale_x_continuous(expand =
c(0, 0)) + scale_y_continuous(expand = c(0, 0)) + labs(x= "Total Publications", y = "De
nsity") + theme(axis.title = element_text(face="bold", size=15)) + theme(legend.text=el
ement_text(size=15), legend.title = element_text(size=15)) + theme(legend.title=eleme
nt_blank(), legend.position=c(.9,.60)) + coord_cartesian(ylim=c(0, 0.01))
```

```r
#Study 2 Density plot
ggplot(s2all, aes(TotalPubs, fill=Gender, color=Gender))+geom_density(bw=1, lwd=0.8, alpha=0.2) + theme(panel.background = element_blank(), axis.text = element_text(color="black"), axis.line = element_line(color = "black")) + scale_x_continuous(expand = c(0, 0)) + scale_y_continuous(expand = c(0, 0)) + labs(x= "Total Publications", y = "Density") + theme(axis.title = element_text(face="bold", size=15)) + theme(legend.text=element_text(size=15), legend.title = element_text(size=15)) + theme(legend.title=element_blank(), legend.position=c(.9,.60))

#Study 2 Density plot (close-up of the tail-end)
ggplot(s2all, aes(TotalPubs, fill=Gender, color=Gender))+geom_density(bw=1, lwd=0.8, alpha=0.2) + theme(panel.background = element_blank(), axis.text = element_text(color="black"), axis.line = element_line(color = "black")) + scale_x_continuous(expand = c(0, 0)) + scale_y_continuous(expand = c(0, 0)) + labs(x= "Total Publications", y = "Density") + theme(axis.title = element_text(face="bold", size=15)) + theme(legend.text=element_text(size=15), legend.title = element_text(size=15)) + theme(legend.title=element_blank(), legend.position=c(.9,.60)) + coord_cartesian(ylim=c(0, 0.0015))

#Study 3 Density plot
ggplot(s3all, aes(TotalPubs, fill=Gender, color=Gender))+geom_density(bw=1, lwd=0.8, alpha=0.2) + theme(panel.background = element_blank(), axis.text = element_text(color="black"), axis.line = element_line(color = "black")) + scale_x_continuous(expand = c(0, 0)) + scale_y_continuous(expand = c(0, 0)) + labs(x= "Total Publications", y = "Density") + theme(axis.title = element_text(face="bold", size=15)) + theme(legend.text=element_text(size=15), legend.title = element_text(size=15)) + theme(legend.title=element_blank(), legend.position=c(.9,.60))

#Study 3 Density plot (close-up of the tail-end)
ggplot(s3all, aes(TotalPubs, fill=Gender, color=Gender))+geom_density(bw=1, lwd=0.8, alpha=0.2) + theme(panel.background = element_blank(), axis.text = element_text(color="black"), axis.line = element_line(color = "black")) + scale_x_continuous(expand = c(0, 0)) + scale_y_continuous(expand = c(0, 0)) + labs(x= "Total Publications", y = "Density") + theme(axis.title = element_text(face="bold", size=15)) + theme(legend.text=element_text(size=15), legend.title = element_text(size=15)) + theme(legend.title=element_blank(), legend.position=c(.9,.60)) + coord_cartesian(ylim=c(0, 0.0005))
```

Appendix B: Inputs from Experts Regarding Relative Author Contributions and Author Ordering in the Fields of Mathematics, Materials Sciences, and Genetics

First, in the field of mathematics, according to R. Xiaofeng, E. Robinson, and A.Naor, the listing of multiple authors is always alphabetical with no relation to the degree of an author's contribution to a paper (personal communication, February 20, 2017). Similarly, M. Gualdani explained that while it is "common practice" to list authors in alphabetical authors, in rare cases they will be rank-ordered (personal communication, February 19, 2017). A. Naor, on the other hand, firmly stated that the listing of authors in alphabetical order is universal across "all areas and all journals" in mathematics (personal communication, February 20, 2017). The explanations provided by these individuals is consistent with a statement made by the American Mathematical Society (AMS) in 2004. According to the AMS, authors are listed alphabetically in more than 75% of mathematics papers (and over 90% of pure mathematics papers) with at least one U.S. author.

In the field of material sciences, the listing of authors is such that the first author is 'responsible for most of the work,' whereas the last author is reserved for the project leader/supervisor/principal investigator who oversaw the project and provided the funding, etc. According to S. Hsu, the ordering of the rest (i.e., excluding the first and last authors) can be "equal, or random, or in the order of contribution" (personal communication, February 21, 2017). Similarly, M. Sailor explained that "The 1st author is the person who did most of the key work (usually a graduate student or a post-doc), and it falls off in order from there" but "When the relative degree of contribution is in doubt, the author list can revert to a random ordering such as alphabetical" (personal communication, February 20, 2017).

Likewise, in the field of genetics, the first author(s) correspond to the individuals who "performed the majority of the experiments and analyses" and the last author(s) correspond to the "principal investigators who supervised the project," (K. Vogan, personal communication, February 20, 2017). A journal editor at Nature Genetics explained that "there's an informal rule that the person who contributed the most (in terms of actually conducting the research) and is followed by other authors who contributed less, in a descending order" and that "for the last authors (aka senior authors) the order is usually reversed" (F. Tiago, personal communication, February 20, 2017). According to J. Akey, if one was to "graph author order in genetics papers (x-axis) versus a measure of contribution or importance it would look like a V" (personal communication, February 19, 2017). There are apparently exceptions to the norm, however, as A. Martin added that "some manuscripts published by big collaborative efforts, such as "Genome Sequencing Consortia", sometimes use the alphabetical order" (personal communication, February 19, 2017).

As a result, I learned that the author credits measure developed by Howard and colleagues (1987) would not be a suitable measure of research productivity in these fields. First, in the field of Mathematics, it is pretty much taken as an axiom that co-authors are listed alphabetically, with no regard to relative contributions. Likewise, in the fields of material sciences and genetics, author order does not provide much information about the relative contributions of co-authors besides the fact that the first and last authors can generally be assumed as being most important. One of my correspondents described the relationship between author order and relative contributions as following a V shape in these fields. However, it cannot be assumed that this accurate reflects the relationship between author order and practically meaningful outcomes such as promotion, pay, and tenure, etc. I instead surmise that, unless an individual was one of the first and/or last authors, author order alone has minimal effect on how others in the field judge that individual's research output, especially on projects with a large number of authors. In other words, differences in the perceived research value added by, say, the 50th author versus the 25th author on a paper with 100 authors, may be minimal in the eyes of others in the field. Moreover, even if I could develop a reliable measure that assigns author credits in a U-shape with respect to ordinal author position, it cannot be assumed in the first place that, in these fields, the middlemost author always contribute the least, as there are apparently cases that deviate from the norm (e.g., cases when some or all co-authors are listed in alphabetical order).

Nonetheless, I did consider the possibility of coding the relative contributions of co-authors through assessment of any supplementary materials released alongside these papers that may provide such information. In Nature journals, for example, authors are required to submit such supplementary information. I eventually realized that this approach would also not be feasible. First, in the field of mathematics, I found that information about the relative contributions of co-authors is rarely, if ever, released. This appears to be a result of the nature of mathematics research and corresponding beliefs about how joint research is produced in this field. In a 2004 statement about the 'culture of research and scholarship in mathematics," the American Mathematical Society (AMS) explained that, in mathematics, "joint research is a sharing of ideas and skills that cannot be attributed to the individuals separately," and that "the roles of researchers are seldom differentiated (in the way they are in laboratory sciences, for example)." According to the AMS, it is for these reasons that the relative contributions of co-authors are considered as being equal, which also explains why author listing is alphabetical in this field.

In the fields of material sciences and genetics, such supplementary information is available, but, I realized that ranking the relative contributions of co-authors using this information would not be feasible. These supplementary statements do not include a 'ranking' of author contributions, but instead, provide information about the different roles occupied by the authors. To illustrate, Speliotes and colleagues (2010) published a paper in Nature Genetics, which, at the time of writing, has received 1932 citations. The supplementary information released alongside that paper includes a four-page summary of author contributions. Not surprisingly, the roles occupied by the first and last authors are listed at the top of the summary. According to the summary, of the 377 total authors on the project, 26 "oversaw the project," and 14 "drafted and edited the manuscript." From there, the document proceeds to list 15 other roles with field-specific descriptions such as "polygene analyses," "gene-expression (eQTL) analyses," and "gene-by-gene interaction analyses." As a layman especially, the relative importance of each of those roles is nearly impossible to determine. Furthermore, many of the authors contributed to more than one role. For example, I found that, of the 26 authors who oversaw the project, 8 of them also contributed to the pivotal role of writing the manuscript. Ultimately, I decided against the 'author credits' approach entirely and proceeded to use the total count of publications as my only measure of research performance despite its potential limitations.