American Psychological Association
Journal of Applied Psychology

0021-9010/18/$12.00
2018, Vol. 103, No. 12, 1283–1306
http://dx.doi.org/10.1037/apl0000331

Gender Productivity Gap Among Star Performers in STEM and Other Scientific Fields

Herman Aguinis and Young Hun Ji
George Washington University

Harry Joo
University of Dayton

We examined the gender productivity gap in science, technology, engineering, mathematics, and other scientific fields (i.e., applied psychology, mathematical psychology), specifically among star performers. Study 1 included 3,853 researchers who published 3,161 articles in mathematics. Study 2 included 45,007 researchers who published 7,746 articles in genetics. Study 3 included 4,081 researchers who published 2,807 articles in applied psychology and 6,337 researchers who published 3,796 articles in mathematical psychology. Results showed that (a) the power law with exponential cutoff is the best-fitting distribution of research productivity across fields and gender groups and (b) there is a considerable gender productivity gap among stars in favor of men across fields. Specifically, the underrepresentation of women is more extreme as we consider more elite ranges of performance (i.e., top 10%, 5%, and 1% of performers). Conceptually, results suggest that individuals vary in research productivity predominantly because of the generative mechanism of incremental differentiation, which is the mechanism that produces power laws with exponential cutoffs. Also, results suggest that incremental differentiation occurs to a greater degree among men and certain forms of discrimination may disproportionately constrain women's output increments. Practically, results suggest that women may have to accumulate more scientific knowledge, resources, and social capital to achieve the same level of increase in total outputs as their male counterparts. Finally, we offer recommendations on interventions aimed at reducing constraints for incremental differentiation among women that could be useful for narrowing the gender productivity gap specifically among star performers.

Keywords: star performers, STEM fields, gender discrimination, scientific productivity

Supplemental materials: http://dx.doi.org/10.1037/apl0000331.supp

According to the 2016 Science and Engineering Indicators by the National Science Foundation (NSF), women continue to be underrepresented in science, technology, engineering, mathematics (STEM). For example, in early (K–12) education, boys and girls display similar participation rates in mathematics and science (e.g., Hyde, Lindberg, Linn, Ellis, & Williams, 2008; Xie & Shauman, 2003). However, large gender imbalances in representation occur in higher level academic fields and in the workforce. For example, although women make up half of the college-educated workforce in the United States, they only make up 29% of the STEM workforce. Also, according to a survey by the Association of American Universities, women chair only 2.7% of engineering departments, 5.9% of math or physical science departments, and 12.7% of life science departments (Niemeier & González, 2004).

The issue of gender disparities particularly in STEM fields is also hotly debated in the media and policy-making circles. For example, in August 2017 a big controversy took place at Google where engineer James Damore wrote a memo harshly criticizing the company's diversity policies and was subsequently fired for "perpetuating gender stereotypes." His memo created a firestorm across Silicon Valley, which takes pride in its progressive views regarding same-sex marriage, transgender rights, and other gender-related issues. The question posed by many companies—and echoed by the media—is this: Why are women underrepresented in the U.S. technology industry? (Fortune, 2017).

We conducted a research program involving three studies with the goal of understanding the presence and possible reasons for a gender productivity gap in STEM (i.e., mathematics, genetics) and other scientific fields (i.e., applied psychology, mathematical psychology), specifically among star performers. An examination of

This article was published Online First July 19, 2018.

Herman Aguinis and Young Hun Ji, Department of Management, School of Business, George Washington University; Harry Joo, Department of Management and Marketing, School of Business Administration, University of Dayton.

The National Science Foundation (Science of Organizations Program)—Grant 1643075 titled "Understanding the Gender Performance Gap among Star Performers in STEM Fields" awarded to Herman Aguinis (Principal Investigator)—supported part of the work described in this article. Results of Study 1 and Study 2 are based, in part, on Young Hun Ji's doctoral dissertation, which was conducted at the School of Business, George Washington University, under the supervision of Herman Aguinis. We thank committee members N. Andrew Cohen, N. Sharon Hill, and James B. Wade for constructive feedback on previous drafts.

Correspondence concerning this article should be addressed to Herman Aguinis, Department of Management, School of Business, George Washington University, Funger Hall, Suite 311, 2201 G Street, North West, Washington, DC 20052. E-mail: haguinis@gwu.edu

star performers is particularly relevant regarding the gender productivity gap because these are individuals who produce output many times greater than the rest of the individuals holding a similar job or position (Aguinis & Bradley, 2015; Aguinis & O'Boyle, 2014; Aguinis, O'Boyle, Gonzalez-Mulé, & Joo, 2016). Further, stars are highly influential to the individuals around them and often serve as role models and mentors. For example, proximity to star performers benefits the career advancement of subordinates through enrichment of the latter's social capital (Malhotra & Singh, 2016). Also, women are more inspired by outstanding female than male role models (Lockwood, 2006). Moreover, supervisors and mentors provide more psychosocial support to protégés who belong to the same gender (Koberg, Boss, & Goodman, 1998). Thus, an understanding of the gender productivity gap, specifically regarding star performers, and the mechanism that may result in this gap can be particularly useful in terms of planning interventions that can have trickle-down effects throughout organizations and even entire professional fields.

Study 1 included 3,853 researchers who published a total of 3,161 articles in the top-10 cited mathematics journals from 2006 to 2015. Study 2 included 45,007 researchers who published a total of 7,746 articles in the top-five cited genetics journals over the same time period. Study 3 included 4,081 researchers who published 2,807 articles in applied psychology and 6,337 researchers who published 3,796 articles in mathematical psychology (also for the same 2006 to 2015 time period). To assess and compare the productivity distributions by gender, we used the distribution pitting methodological approach developed by Joo, Aguinis, and Bradley (2017). Distribution pitting is particularly well-suited for testing our hypotheses because it is a falsification procedure (Gray & Cooper, 2010; Lakatos, 1976; Popper, 1959) that involves pitting theoretical distributions against one another with respect to the observed distribution. Specifically, it involves identifying distributions that do not reflect the data accurately, and, in doing so, determining the most likely dominant (i.e., sole-surviving) distribution. As we describe later, a key theoretical implication of distribution pitting is that each distribution type is associated with a unique generative mechanism. But, the identification of the surviving distribution does not completely rule out the presence of other distributions (and their generative mechanisms); rather, it identifies the predominant distribution and explanation for the existence of a particular distribution.

Our results were consistent across the three studies and showed that the power law with exponential cutoff best fit the observed individual output distributions overall and also for women and men separately. The dominance of the power law with exponential cutoff distribution implies that star performers emerge predominantly through the generative mechanism of incremental differentiation. In addition to the consistent dominance of incremental differentiation, results revealed considerable gender productivity gaps among star performers in favor of men. These gaps are reflected in the tails of the distributions, which were heavier for men, thus indicating a greater proportion of male star performers. Moreover, women were more severely underrepresented among the top performers than among all performers across all of our studies, and the degree of underrepresentation increased as levels of performance increased. In other words, female underrepresentation became more accentuated as we examined the top 10%, 5%, and 1% of performers. Our results thus offer insights into the emergence of star performers and also the existence of a gender productivity gap among star performers, thereby generating theory-based contributions regarding the mechanisms leading to the observed gaps as well as practical implications regarding interventions aimed at reducing them. Moreover, our findings build upon but also go above and beyond Joo et al. (2017), who introduced the distribution pitting methodology but did not address gender issues. In fact, Joo et al. focused on a research question different from ours and did not even report the number of women and men included in their samples—let alone any type of theory or analysis regarding gender-related issues. In short, Joo et al.'s conceptualization and analyses were not intended to examine issues about gender, which is the central goal of our studies.

Theoretical Background and Hypotheses

Our theoretical background is unique in that it bridges two different and, to date, disconnected bodies of research: (a) applied psychology and other social science research on the reasons for the existence of a gender productivity gap and (b) natural science research on the underlying mechanisms that lead to the formation of particular output distributions. In this section, we integrate these two literatures by offering hypotheses that connect specific distribution shapes with their underlying mechanisms and, in turn, relate these mechanisms to existing conceptualizations explaining the gender productivity gap. First, we describe three competing conceptualizations explaining the gap. Next, we describe various types of distribution shapes and their generative mechanisms. In doing so, we offer eight hypotheses that connect each distribution type with the existing explanations for the gender productivity gap.

Competing Conceptualizations Explaining the Gender Productivity Gap

There are three broad competing conceptualizations most frequently used to explain the underrepresentation of women in STEM and other scientific fields. These three conceptualizations can also be used to explain the gender productivity gap: (a) gender differences in abilities, (b) gender discrimination, and (c) gender differences in career and lifestyle choices. Each of the three conceptualizations offers a unique perspective. The gender differences in abilities perspective emphasizes differences in biological factors (e.g., quantitative abilities); the gender discrimination perspective emphasizes sociocultural and contextual factors that create imbalances in the opportunities and barriers faced by women compared with men; and the career and life choices perspective emphasizes gender differences in motivation and other psychological factors.

Gender differences in abilities. The first of these conceptualizations is that the gender disparity can be explained largely by sex differences in quantitative and other related abilities (Halpern, 2000; Halpern et al., 2007). According to this line of research, biological factors such as exposure to prenatal and postnatal testosterone and brain lateralization patterns 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 with 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 gender differences in abilities perspective suggests that women's underrepresentation and the gender productivity gap are mainly caused by biological factors that enable men to outperform women in quantitative tasks.

Gender discrimination. The second explanation suggests that women are pushed out of STEM and other scientific fields because of discrimination, starting from early childhood and persisting all the way to higher education and professional environments. Gender biases held by parents and teachers, for instance, may contribute to a greater allocation of resources to boys than to girls in early STEM development (Tenenbaum & Leaper, 2003). Similarly, women may face discrimination in the form of limited opportunities for advancement (Xu, 2008), be perceived as less competent than men with comparable achievements (Moss-Racusin, Dovidio, Brescoll, Graham, & Handelsman, 2012), and receive less credit for collaborative work (Sarsons, 2017). For example, reviewers and colleagues often undervalue the quality of female scientists' research outputs and are less likely to show collaboration interest toward them—a bias referred to as the Matilda effect (Knobloch-Westerwick, Glynn, & Huge, 2013; Lincoln, Pincus, Koster, & Leboy, 2012; Merton, 1968; Rossiter, 1993). Also, additional empirical evidence has provided support for gender discrimination effects in hiring, journal reviewing, and grant funding (e.g., Chesler, Barabino, Bhatia, & Richards-Kortum, 2010; Lortie et al., 2007). In short, this perspective suggests that women are underrepresented and outperformed by men mainly because of gender discrimination stemming from sociocultural processes (Carli, Alawa, Lee, Zhao, & Kim, 2016).

Gender differences in career and lifestyle choices. The third conceptualization is that women's underrepresentation in STEM and other scientific fields and the gender productivity gap are largely the result of particular ways in which women and men tend to diverge on the career and lifestyle choices they make. According to this perspective, there are gender differences with respect to a wide range of psychological factors, which in turn shape people's decision to pursue a scientific career and persist in the chosen field (Ceci & Williams, 2010). That is, men and women differ in their motivation to pursue and persist in STEM and other scientific careers (e.g., Ceci & Williams, 2010; Wang & Degol, 2013). Gender differences in psychological factors such as interests, work goals, occupational preferences, and work–life/family values may result in women being more likely to make sacrifices and tradeoffs during the course of their careers, including the decision to opt out of STEM and other scientific fields altogether. In sum, according to this conceptualization, women choose to opt out of STEM and other scientific fields at higher rates than men at all stages of their careers because of a wide range of psychological and motivational factors (Kossek, Su, & Wu, 2017).

Integrating Gender Productivity Gap Conceptualizations With Generative Mechanisms for Productivity Distributions

Empirical research in computer science, physics, zoology, and other fields (e.g., economics) has identified generative mechanisms leading to the formation of specific distributions of outcomes (Clauset, Shalizi, & Newman, 2009). In this section, we describe evidence based on dozens of studies across these fields showing that each type of observed distribution is generated by a particular mechanism and, thus, serves as a "smoking gun" for the unique underlying process leading to a specific distributional shape (Clauset et al., 2009). On the basis of this extensive literature, Joo et al. (2017) proposed a methodological approach called distribution pitting, which involves identifying the best-fitting theoretical distribution with respect to the observed productivity distribution among the following theoretical distribution types: (1) pure power law; (2) lognormal; (3) exponential tail (including exponential and power law with an exponential cutoff); and (4) symmetric or potentially symmetric (including normal, Weibull, and Poisson). Then, based on the predominant theoretical shape identified, we can infer the underlying mechanism that resulted in its formation.

Next, we describe the generative mechanism for each type of distribution and how the generative mechanisms apply to individual productivity. We also offer four competing hypotheses based on the viability of the four generative mechanisms. In addition, going beyond Joo et al. (2017), we offer a conceptual integration of research on gender with research on productivity distributions to offer four competing hypotheses about the mechanisms that would result in a gender productivity gap among star performers.

Pure power law distribution and self-organized criticality. The presence of a power law distribution is indicative of a generative mechanism referred to as *self-organized criticality* (Bak, 1996), which emphasizes the role of *output shocks* (i.e., large and unpredictable increases in output). Self-organized criticality is a process where observations (e.g., individuals) accumulate small amounts on an outcome (e.g., output) before reaching a critical state (i.e., a situation where components accumulated by an individual interconnect). After reaching a critical state, depending on the precise configuration of one's accumulated components and their interconnections, even a seemingly trivial event may trigger large output shocks. For example, research in physics has found that once enough sand grains have piled up to reach a critical slope, the drop of another sand grain will cause a sand avalanche. This process, repeated over many times, will generate a pure power law distribution of sand avalanche sizes (Bak, 1996). As another example, when a start-up reaches a critical state, even a small event such as a business plan presentation may trigger explosive growth (Crawford, Aguinis, Lichtenstein, Davidsson, & McKelvey, 2015).

In terms of parameters of this distribution, a set of values from a variable x follows a pure power law if

$$p(x) \propto x^{-\alpha} \qquad (1)$$

where α (>1) is the rate of decay, or how quickly the distribution's right tail "falls." The lower the value of α (closer to 1), the heavier is the distribution's right tail. For example, a distribution with $\alpha = 2$ has a heavier right tail compared with a distribution with $\alpha = 3$.

In the context of individual productivity, 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. 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 subsequent research productivity. In this sense, self-organized criticality involves a significant element of randomness and luck. 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; Taleb, 2007). After individuals reach such critical states, even a trivial event may cause unpredictably large output shocks. Accordingly, large paradigm-shifting breakthroughs may depend on the successful interaction of multiple components and events. Essentially, researchers may reach critical states when certain performance components interconnect (e.g., a set of interrelated projects regarding potential cures for a single disease). Subsequently, even a seemingly trivial event such as access to funding for a single project could trigger large increases in subsequent research productivity. We suggest that such interconnections may be the key that allows some scientists to experience huge leaps in scientific productivity or eureka moments. As such, large output shocks following critical states may be the key process through which star scientists differentiate their performance from those of others. Hence, differences in individuals' productivity may be driven predominantly by the pure power law distribution's generative mechanism, or self-organized criticality.

> *Hypothesis 1a:* Individual productivity of women and men in STEM and other scientific fields follows a pure power law distribution.

In terms of gender-based differences in productivity, self-organized criticality could explain a gender productivity gap among stars in favor of men, and this gap would be consistent with both the gender discrimination and career/lifestyle choices perspectives. First, the gender discrimination perspective suggests a productivity gap among stars in favor of men, as certain forms of discrimination may lead to smaller/fewer output shocks among women. For example, prior research suggests that women in STEM fields are generally less favored in important hiring and promotion decisions. Specifically, in a study where professors in biology, physics, and chemistry evaluated applications for a lab manager position, candidates with a female name were less likely to be hired, received a lower starting salary, obtained less mentoring opportunities, and were generally perceived as less competent than other candidates with identical application materials but with a male name (Moss-Racusin et al., 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 chosen only 33.9% of the time, which meant that men were twice more likely to be chosen than women. Studies also suggest that, due largely to gender discrimination, women in STEM and other scientific fields are less likely to be promoted to leadership positions and achieve tenure status, tend to receive less research funding and support, and are often assigned heavier teaching loads (Xu, 2008). In short, the gender discrimination perspective suggests that women are less likely than their male peers to experience certain events that enable critical states and subsequent large output shocks.

Second, output shocks may be smaller among women because of gender differences in career and lifestyle choices. For example, according to this perspective, women are more likely to make tradeoffs, such as deferring career goals in pursuit of family goals or following a spouse's job location (Ceci & Williams, 2011; Singh, Zhang, Wan, & Fouad, in press)—tradeoffs that may collectively contribute to smaller and fewer output shocks. In particular, the decision to marry or have children results in disproportionate productivity losses and other career-related changes for women that lead to disadvantages in hiring and promotion (Ceci & Williams, 2011; Wang & Degol, 2013). For example, a survey of University of California graduate students found that women with children were 35% less likely to enter a tenure-track position after receiving a PhD than married men with children and are 27% less likely than men to achieve tenure (Mason & Goulden, 2009). Additionally, regarding promotion to leadership roles, the career and lifestyle choices perspective suggests that women often stay away from leadership roles because some prefer to spend more time teaching and having more opportunities for collegial collaboration (e.g., Bentley & Adamson, 2003; Robertson, Smeets, Lubinski, & Benbow, 2010).

We argue that the occurrence of sudden and large output shocks largely depends on the presence of a particularly biased (vs. supportive) supervisor or a major career and/or life decision made by the individual. In comparison, gender differences in ability and other individual differences may have a relatively smaller impact on enabling output shocks. This is expected because major career/life choices or discrimination act as greater potential barriers against female researchers to both (1) reaching critical states (where one's accumulated components interconnect) and also (2) generating larger output shocks after having reached critical states (given that output shocks occur only after reaching critical states in the self-organized criticality framework). For example, the choice to decline a leadership position or failure to obtain it because of gender-based bias may prevent a female researcher from gaining the big-picture insights necessary for integrating her existing human and social capital. Even after a female researcher has reached a critical state by having obtained and learned from such a leadership position, the subsequent decision to devote larger amounts of time for nonresearch areas of life or (subtle) gender-based biases by her fellow male leaders could lower her likelihood of experiencing very large-sized output shocks. In contrast, although any stable ability differences across the two genders may prevent female researchers from reaching critical states, such differences are less likely to affect the size of subsequent output shocks—the latter of which is more likely a function of external factors such as discrimination or abrupt factors such as a major career/life decision. Accordingly, an integration of the literatures on generative mechanisms and the gender productivity gap conceptualizations suggests that the presence of a gender productivity gap under a power law distribution would largely be the result of gender discrimination and/or women's lifestyle choices.

> *Hypothesis 1b:* The pure power law distribution of individual productivity will have a lighter right tail for women than men.

Lognormal distribution and proportionate differentiation. Lognormal distributions result from the generative process of *proportionate differentiation*, where individuals' future output is a distinct percentage of their prior output (Barabási, 2012). Proportionate differentiation suggests that individuals' prior value on an outcome (e.g., output) interacts with their accumulation rates in determining their future amounts on the same outcome. Accumulation rate refers to the average amount of a variable that an individual produces per time period (e.g., sales generated per month), whereas prior output 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. 1030)

As an example from geology, a crystal's rate of exposure to additional minerals and its initial size together may determine the crystal's subsequent sizes, leading to a lognormal distribution of crystal sizes (Kile & Eberl, 2003).

Regarding the parameters for this distribution, a set of values from a variable x follows a lognormal distribution if

$$p(x) \propto e^{-\frac{(ln(x)-\mu)^2}{2\sigma^2}} \quad (2)$$

where Euler's number $e \approx 2.718$. $ln(x)$ is the natural log of x and is normally distributed. μ (>0) is the mean. σ (>0) is the standard deviation. μ does not affect the heaviness of the distribution's right tail but σ does. The higher the value of σ (further away from 0), the heavier is the distribution's right tail.

In terms of individual productivity, the proportionate differentiation mechanism suggests that a small proportion of individuals (with the largest initial outputs and accumulation rates), compared with others, enjoy larger *output loops* (i.e., increasingly larger output increases based on positive feedback between past and future output). To illustrate, assume there are two researchers, A and B. Also assume that both are comparably talented, but A starts his tenure-track career with three publications in top-tier journals, whereas B starts with only one because of situational factors (e.g., a senior researcher joined the department and offered A the opportunity to work with him). Here, A may find it easier to produce subsequent publications because of, say, greater 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 effort, B would not be able to catch up to A in terms of number of publications unless B is able to increase her accumulation rate enough to eventually offset the impacts of A's greater initial output value.

In STEM fields and other scientific domains, individuals' past successes have a powerful impact on their subsequent access to various resources and opportunities and, thus, their potential for future success (Kwiek, 2018). In early education, for example, children who display large initial successes in STEM subjects are given more attention and resources from teachers and parents (e.g., tutoring, advanced educational programs), which then helps to beget even more success. As such, small differences in talent and early performance can lead to large differences in rewards which, in turn, translate into disproportionate levels of future success (Kwiek, 2018). In professional environments, one's past successes are similarly linked to future productivity. In a meritocratic fashion, people and organizations make greater investments (e.g., funding and collaboration opportunities) in researchers who have been successful in the past. In sociology, this phenomenon is referred to as the Matthew effect, where scientists receive greater recognition and rewards for their work on the basis of their current renown and visibility (Merton, 1968). In other words, the rich get richer, as various advantages accumulate for individuals who have already received recognition (Lincoln et al., 2012). Moreover, academics may accrue disadvantages as a result of the accumulation of failures, resulting in the poor getting poorer (Kwiek, 2018). In a recent study demonstrating performance differences among Polish scholars (e.g., the top 10% of performers produced roughly half of all journal articles), Kwiek (2018) theorized that the Matthew effect was what primarily drove such large differences in research productivity. Accordingly, individual productivity may be driven predominantly by the lognormal distribution's generative mechanism of proportionate differentiation.

> *Hypothesis 2a:* Individual productivity of women and men in STEM and other scientific fields follows a lognormal distribution.

With respect to gender differences in productivity, the gender discrimination perspective suggests that women may experience smaller output loops compared with their male colleagues. In particular, the feedback mechanism linking women's prior outputs to their future outputs may be constrained compared with men, as people often "overvalue" men's prior achievements (Reuben et al., 2014), thereby investing greater resources in men and boys than in women and girls who display comparable levels of prior success (Brown & Stone, 2016; Tenenbaum & Leaper, 2003). Accordingly, the gender discrimination literature suggests numerous ways in which gender-based actions of others may result in smaller output loops for women. One example has to do with how female (vs. male) researchers are evaluated in terms of their prior outputs and their desirability as potential collaborators, which is likely one of the most fundamental processes that enable proportionate differentiation for these individuals, given that most research conducted in these fields is collaborative in nature. Prior studies suggest that when people evaluate the past works of others, they often place disproportionate penalties on women for coauthored works (i.e., giving them less credit than their male coauthors) and perceive them less favorably as potential collaborators (Diekman, Weisgram, & Belanger, 2015; Knobloch-Westerwick et al., 2013; Sarsons, 2017). Moreover, regarding hiring and promotion, which is a process that similarly involves others evaluating an individual's past outputs and making decisions that affect future output gains, female applicants are less favored despite having equivalent qualifications as their male counterparts because of gender biases (e.g., Moss-Racusin et al., 2012; Reuben et al., 2014). Additionally, the literature on stereotype threat (Walton, Murphy, & Ryan, 2015) shows that when individuals are reminded of negative stereotypes directed at them (e.g., via biased actions from peers), their performance (e.g., women's performance on math tasks) diminishes.

As discussed, in STEM and other scientific fields, the connection between one's prior and future output is rooted primarily in the actions of other people (i.e., an external factor). In other words, one's past outputs matter in terms of future output gains, because people tend to invest more in highly productive individuals, which enables those individuals to enjoy larger output loops. In contrast, intrapersonal factors such as ability and motivation, although important for initiating positive feedback loops between initial and future output, might be less impactful in determining the size and duration of such feedback loops. Accordingly, the presence of a gender productivity gap under a lognormal distribution would largely be the result of gender discrimination.

> *Hypothesis 2b:* The lognormal distribution of individual productivity will have a lighter right tail for women than men.

Exponential tail distributions and incremental differentiation. Exponential tail distributions (i.e., exponential and the power law with exponential cutoff distributions) result from the same generative mechanism called *incremental differentiation*. This mechanism implies that individuals' output increases at an approximately linear rate based on their accumulation rate

(Amitrano, 2012). Unlike proportionate differentiation, prior output is not linked to future output through a positive feedback loop. Rather, 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).

Regarding parameters, a set of values from a variable x follows an exponential distribution if

$$p(x) \propto e^{-\lambda x} \tag{3}$$

where Euler's number $e \approx 2.718$. λ (>0) is the rate of decay, or how quickly the distribution's right tail falls. The lower the value of λ (closer to 0), the heavier is the distribution's right tail.

Regarding power law with exponential cutoff distributions, a set of values from a variable x follows this distribution if

$$p(x) \propto x^{-\alpha} e^{-\lambda x} \tag{4}$$

where Euler's number $e \approx 2.718$. Both α (>1) and λ (>0) are rates of decay, or how quickly the distribution's right tail falls. 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.

In terms of individual productivity, top performers with the highest accumulation rates enjoy larger *output increments* (i.e., linear increases in output) than others. As an example, because of their higher accumulation rates, some researchers may produce a greater number of publications compared with other researchers who began their academic careers around the same time. Individuals in STEM and other scientific fields may vary in terms of their productivity primarily because of differences in their accumulation of various input components that, together, have a stable (and linear) impact on their future output increases. Accordingly, prior studies have demonstrated that the accumulation of inputs such as social capital, training, and research hours lead to greater incremental growth in future outputs. For example, in a study involving professors from two U.S. research institutions, van Eck Peluchette and Jeanquart (2000) found that individuals who had multiple mentors experienced significantly higher levels of career success, on objective as well as subjective measures, and at all stages of their careers (i.e., early, middle, and late stage), compared with others who did not. In another study, which involved faculty members from 92 academic otolaryngology departments, findings showed that fellowship-trained otolaryngologists had significantly higher research productivity than non-fellowship-trained otolaryngologists, as measured by the h-index, which considers both the number of articles published and the number of citations received by each (Eloy, Svider, Mauro, Setzen, & Baredes, 2012). In yet another study, Kwiek (2018) found that the top 10% of Polish academics, on average, spent 5 more hours per week on research, and the working time differential was even greater among researchers in mathematics and physical sciences (i.e., 12 more hours per week). As such, researchers may vary in their total publications primarily because of differences in the rate at which they acquire important input components such as social capital (e.g., mentors and professional connections), education (e.g., AP courses, graduate degrees), advanced scientific training, and working hours. Accordingly, individual productivity may be driven mainly by the exponential tail distributions' generative mechanism, or incremental differentiation.

> *Hypothesis 3a:* Individual productivity of women and men in STEM and other scientific fields follows an exponential tail distribution.

With respect to gender differences in productivity, there may be a productivity gap in favor of men because of certain forms of gender discrimination that result in lower output increments for women compared with men with comparable accumulation rates. For example, John and Sally may have similar accumulation rates on major input components—such as knowledge, social capital, and other research-related resources—yet John may have a greater publication rate (i.e., larger output increments) than Sally because of consistent gender biases in peer reviews. Prior literatures demonstrate that women working in male-dominated fields are often perceived by others as lacking the innate talent or "genius" required to be successful (Cheryan, Ziegler, Montoya, & Jiang, 2017). When such stereotypes (i.e., tendency to undervalue female scientists' abilities) are held by gatekeepers such as mentors, potential collaborators, hiring committees, and other decision makers (e.g., journal editors), female scientists may need to overaccumulate input components and thus achieve higher accumulation rates to achieve the same outputs as their male counterparts. For example, even if John and Sally enter the same institution with identical qualifications, gender biases (and gender homophily) among top (male) mentors in the department could result in a greater publication rate for John, unless Sally increases her accumulation rate of input components to offset this effect. In addition to gender discrimination, it is also possible that men and women considerably differ in their accumulation rates because of gender differences in abilities and/or lifestyle choices. However, these differences may have a smaller impact on gender differences in output, given that prior research largely (though not exclusively) attributes situational constraints (e.g., limited resources) rather than person-based factors (e.g., ability) to explain differential levels of positive skew in exponential-tail distributions (Amaral, Scala, Barthélémy, & Stanley, 2000; Joo et al., 2017). Accordingly, the presence of a gender productivity gap under an exponential tail distribution may largely be a reflection of the impact of gender discrimination.

> *Hypothesis 3b:* The exponential tail distribution of individual productivity will have a lighter right tail for women than men.

Symmetric or potentially symmetric distributions and homogenization. Finally, three symmetric or potentially symmetric distributions (i.e., normal, Weibull, and Poisson distributions) result from the generative mechanism of *homogenization* (Araújo & Herrmann, 2010). Unlike the other generative mechanisms described so far (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). As another example, uniform expectations of production or service tend to homogenize workers' outputs (e.g., Groshen, 1991).

In terms of parameters, a set of values from a variable x follows a normal distribution if

$$p(x) \propto e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{5}$$

where Euler's number $e \approx 2.718$. μ (>0) is the mean. 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.

A set of values from a variable x follows a Weibull distribution if

$$p(x) \propto (x/\lambda)^{\beta-1} e^{-(x/\lambda)^{\beta}} \tag{6}$$

where Euler's number $e \approx 2.718$. β (>0) is the extent to which the distribution is "pulled" up and to the right. 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. λ (>0) is the extent to which the distribution is "pushed" down and stretched to the sides. The lower the value of λ (i.e., closer to 0), the higher is the height of the distribution's bell-shaped head.

A set of values from a variable x follows a Poisson distribution if

$$p(x) \propto \frac{\mu^x}{x!} \tag{7}$$

where 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.

In the context of individual productivity, this generative mechanism suggests that individuals are subject to *output homogenization*, or the reduction of differences in individual output. The reason is that there are constraints that act as a floor and ceiling to future output differences. In STEM and other scientific fields, individuals undergo certain processes that act as a floor to future output. For example, promotion policies involving the denial of tenure to assistant professors who fail to produce a certain high number of publications act as such a floor, thus limiting the variability in researchers' productivity. In this manner, individual productivity may be driven predominantly by the (potentially) symmetric distributions' dominant generative mechanism, or homogenization.

> *Hypothesis 4a:* Individual productivity of women and men in STEM and other scientific fields follows a (potentially) symmetric distribution.

With respect to gender differences in productivity, women's productivity compared with men's may undergo greater homogenization over time because of certain gender differences in career and life choices. Specifically, women may experience disproportionately greater productivity losses as a result of certain choices, such as the decision to have children early in one's career (Ceci, Ginther, Kahn, & Williams, 2014). In certain fields, women compared with men may also place greater priority on family responsibilities and make more career sacrifices over the course of their careers (Singh et al., in press). Such choices act as a disproportionately lower ceiling to women's future productivity, resulting in greater output homogenization among women than among men. Compared with gender differences in career and life choices, which tend to be abrupt in their occurrence, the ceiling effects of gender discrimination and gender differences in abilities on productivity, we argue, are relatively more constant over time, thereby exerting a smaller impact on the strength/presence of output homogenization processes. Accordingly, the presence of a gender productivity gap under a (potentially) symmetric distribution may largely be a reflection of women's choices.

> *Hypothesis 4b:* The (potentially) symmetric distribution of individual productivity will have a lighter right tail for women than men.

In summary, integrating gender productivity gap conceptualizations with generative mechanisms for productivity distributions allowed us to offer four competing hypotheses (i.e., Hypotheses 1a, 2a, 3a, and 4a) about the specific shape of the distribution and, implicitly, the underlying mechanisms leading to that particular shape. Also, for the case of all of the distributions, we proposed a similar hypothesis (i.e., Hypotheses 1b, 2b, 3b, and 4b), stating that the right tail on the distribution will be lighter for women than men.

Study 1: Method

Sample

We examined the productivity of researchers in the field of mathematics who have published at least one article in one of the 10 most influential mathematics journals from January 2006 to December 2015. The field of mathematics is one of the most male-dominated disciplines within STEM and represents a domain where some of the most extreme gender productivity gaps might be observed. For example, only 7.3% of full professor positions in the field of mathematics are occupied by women (Ceci & Williams, 2010). The sample size was 3,853 unique researchers, of whom 360 (9.3%) were women.

The Institutional Review Board at Indiana University approved our data collection (Protocol Number: 1512087389; Title: "Understanding the Gender Performance Gap among Star Performers in STEM Fields"). Our study was judged to be exempt from institutional review board review because of the use of secondary preexisting data.

Journal Selection Criteria

We identified the 10 most influential journals from the mathematics category of Web of Science based on their mean impact factor from 2011 to 2015 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 Analysis*, *Foundations of Computational Mathematics*, *Publications Mathématiques de l'IHÉS*, *Inventiones Mathematicae*, and *Bulletin of the American Mathematical Society*. The impact factor of a journal is the average number of citations received per article 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 articles published in 2014 and 2013 received four citations each, on average, in 2015.

The total number of articles published from January 2006 to December 2015 was 3,161. This total may appear somewhat small considering that that it includes 10 journals and a 10-year period. But this is due to publication practices in mathematics that differ from those of applied psychology and related fields. As an illustrative comparison, *Journal of Applied Psychology* (JAP) alone published 1,082 articles during the 10-year period from 2001 to 2010 (Kruschke, Aguinis, & Joo, 2012).

Measures

Research productivity: Number of articles published in top-tier journals. We measured research productivity by counting the total number of articles published by each author in the previously mentioned 10 journals during the 10-year period from January 2006 to December 2015. We used the Web of Science database to identify all articles and their authors. Also, we used the metadata associated with each of the articles to record the names of all authors, and we used the Open Researcher and Contributor ID to identify unique authors as needed.

We initially considered taking authorship order into account rather than assigning an equal unit weight to each author per article published. For example, Howard, Cole, and Maxwell (1987) developed a procedure that determines authors' rank-weighted "author credits" that are proportional to their ordinal position. However, in contrast to practices in applied psychology, authorship credit in most STEM fields is not based on relative contribution such as the first author contributing the most, then the second, and so on in descending order. For example, in the field of mathematics, the listing of multiple authors is usually in alphabetical order with no relation to the degree of an author's contribution to an article. Appendix A in the online supplemental material includes a more detailed description of author ordering and its meaning in STEM journals, which led to our choice to use unit weights rather than author credits.

Gender. We recorded the gender of each author based on his or her first name. In cases where the gender associated with a first name was ambiguous (e.g., by the use of initials only or gender-neutral first names), we visited the author's web page (personal, faculty, profile on ResearchGate or Google Scholar) to ascertain their gender. In cases where the first name was ambiguous and we could not find a web page, photo, or other information that would reveal an author's gender, we used the website Namepedia.org to find the gender that is most strongly associated with a name. When a first name is entered into the Namepedia database, it generates entries for that name by country and 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 the name in that particular region is male, female, mostly male, or mostly female. For example, Jean is more likely a female name in English-speaking countries, but it is usually a male name in France. In such cases where a first name can be associated with different genders depending on the region, we first deduced the author's ethnicity and geographic region via the surname or other information such as the location of the current workplace, alma mater, and other available information. Then, we coded their gender that best matched their name according to Namepedia, given authors' geographic and ethnicity information.

As recommended by an anonymous reviewer, we also conducted all of our substantive analyses and hypothesis tests excluding ambiguous names for which we used Namepedia. Substantive results and conclusions remained unchanged. Appendix B in the online supplemental material includes tables summarizing results using the reduced samples, where sample sizes were reduced by eight (Study 1: mathematics), 81 (Study 2: genetics), 60 (Study 3: applied psychology), and 34 (Study 3: mathematical psychology).

Data Analytic Approach

Distribution pitting. To test Hypotheses 1a, 2a, 3a, and 4a, we used distribution pitting implemented with the R package Dpit, which is available on the CRAN. Dpit allows pairwise comparisons among competing theoretical distributions with respect to the observed distribution. Appendix C in the online supplemental material includes the entire R script we used for implementing distribution pitting. Also, in the interest of full transparency (Aguinis, Ramani, & Alabduljader, 2018) and as recommended by recently published American Psychological Association guidelines (Appelbaum et al., 2018), we make all of our data files available on request.

As described in detail by Joo et al. (2017), distribution pitting involves three decision rules used to ultimately identify the likely dominant distribution and associated generative mechanism for each observed distribution. The first decision rule involves generating distribution pitting statistics. That is, we used Dpit to conduct 21 pairwise fit comparisons of seven theoretical distributions: pure power law, lognormal, exponential, power law with an exponential cutoff, normal, Weibull, and Poisson. In turn, for each pairwise comparison, the R package provides the log likelihood ratio (LR) and its p value (Aguinis, Gottfredson, & Culpepper, 2013). 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 nonzero 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. As recommended, we used a p value cutoff of 0.10 (Clauset et al., 2009). These statistics were used in the first decision rule to identify theoretical distributions that can be ruled out by any of the other theoretical distributions.

As suggested by Joo et al. (2017), if the first step and decision rule did not result in only one surviving distribution, the second decision rule is to apply the principle of parsimony. If two theoretical distributions survived the first step of the distribution pitting process, we subsequently chose the distribution with fewer parameters as being a better match to the observed distribution. Distributions with more parameters have equivalent or superior fit to the observed distribution; however, they are associated with reduced parsimony and risks being associated with sampling error and chance in general that reduce generalizability beyond the specific sample (Aguinis, Cascio, & Ramani, 2017). Out of the 21 pairwise comparisons of the theoretical distributions, three comparisons involve distributions that 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, we identified the former as being the better explanation for the observed distribution.

As the third step and decision rule in the distribution pitting procedure, if the first and second steps did not result in only one surviving distribution, we again relied on the principle of parsimony—but this time, to rule out one or more of the surviving distributions such that, among multiple remaining distributions, the theoretical distribution with a greater range of possible distribution shapes lacks parsimony and, therefore, is considered the worse explanation. 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 remained after using the first and second decision rules, we identified the flexible distribution (i.e., the theoretical distribution with a greater range of possible distribution shapes) as having the worse explanation and, therefore, ruled it out. In short, if one or more of the three flexible distributions along with one or more of the four inflexible distributions still remain survivors, the appropriate decision is to rule out the flexible distribution(s) while keeping the inflexible distribution(s).

Log likelihood values. Following the recommendation by an anonymous reviewer, we also calculated log likelihood values, which serve as an index of absolute rather than relative fit—holding sample size constant (Edwards, 1972; Huzurbazar, 1948). The calculation of log likelihood values involves two main procedures. First, for each data point in the focal sample, we estimated the likelihood of observing the data point given the best-fitting theoretical distribution to the sample (i.e., point-wise likelihood per data point). Specifically, each point-wise likelihood is the natural logarithm of the data point's likelihood and is expressed as a negative value, such that smaller negative values closer to zero denote better fit to the theoretical distribution. Second, for the focal sample, we added all the point-wise log likelihood values into a single negative value. The resulting value is the log likelihood of observing the sample if the focal theoretical distribution were correct—such that smaller negative values closer to zero suggest better fit.

The following equation summarizes the aforementioned description of how we calculated log likelihood values (Edwards, 1972, p. 33; Huzurbazar, 1948, p. 185):

$$L = \sum_{i=1}^{N} \ln f(x_i | \theta) \tag{8}$$

where L = log likelihood of the sample's fit to the theoretical distribution; i = data point; N = sample size; ln = natural logarithm; x = a data point in the sample; θ = a theoretical distribution; $f(x_i \mid \theta)$ refers to the likelihood of observing the focal data point given the theoretical distribution at hand; $\ln f(x_i \mid \theta)$ is the negative log likelihood for data point i; and Σ indicates that all negative log likelihood values (i.e., point-wise log likelihood values) are added to compute the focal sample's L.

An important caveat regarding Equation 8 is that it is based on adding all point-wise likelihood values within the sample. So, even if two different samples in truth fit equally well to a theoretical distribution, the log likelihood (L) for the larger sample will have a larger negative value, which in this example incorrectly indicates worse fit. This is why, as described earlier, we computed log likelihood ratios and their p values to compare the relative fit across theoretical distributions within each sample, and not across samples.

Fit parameters and descriptive statistics. To test Hypotheses 1b, 2b, 3b, and 4b, we estimated the one or more parameters associated with the best-fitting theoretical distribution per sample. Such parameters, for example, include the mean and standard deviation of a normal distribution or the pure power law's rate of decay α. Parameter estimation is necessary to compare differences in the right-rail heaviness of the productivity distributions across genders and allows us to test Hypotheses 1b, 2b, 3b, and 4b. In other words, even if both gender groups share the same theoretical distribution and its associated generative mechanism, a comparison of fit parameters provides information on the relative heaviness of the distributions' right tails and serves as an effect size estimate. In particular, for the pure power law, lognormal, exponential, and power law with exponential cutoff distributions, a heavier right tail suggests that the particular generative mechanism is stronger for the particular gender group. In contrast, for (potentially) symmetric distributions (i.e., normal, Weibull, and Poisson), lighter tails suggest that the underlying generative mechanism (i.e., homogenization) is stronger for a particular gender group. We computed fit parameters using functions in the Dpit package, which are shown in Appendix C in the online supplemental material.

We also computed the productivity of top performers (top 10%, 5%, and 1%) relative to the total output of their gender group. Moreover, we created graphs to perform visual comparisons of the best-fitting theoretical distributions across genders. This information serves to better illustrate the significance of our results to theory and practice, specifically in terms of gender differences in the right tails. The code we used for creating these graphs are provided in Appendix C in the online supplemental material.

Finally, we conducted two additional types of analysis to supplement the descriptive results based on percentages and graphs: bootstrapping and permutation. These two additional analyses allowed us to more formally compare the relative productivity of top performers. Bootstrapping was based on 5,000 replications of each best-fitting theoretical distribution's parameter value. We used the boot or bootstrap package in R to compute two-sided 95% bootstrapped confidence intervals, with the exception that, for Study 2, we used 50,000 replications because it is necessary to use a larger number of replications than the number of data points in a sample (the values of N for women and men in Study 2 are 14,685 and 30,322, respectively). We used the bias-corrected (i.e., balanced) bootstrapping procedure that is available in the boot R package because it is less susceptible to underestimating the presence of outliers (e.g., Yam, Fehr, Keng-Highberger, Klotz, & Reynolds, 2016). Regarding the permutation analysis, we used the jmuOutlier package in R to check the statistical significance of any observed underrepresentation (i.e., lower than expected fraction) of women among the top 1%, 5%, and 10% producers of research—when compared against an expected fraction (μ) of women in the top 1%, 5%, and 10%. To compute the expected fraction value, we calculated the fraction of women in the entire sample used per study (as opposed to the fraction of women in the top 1%, 5%, or 10% of the sample). As a result, we derived a p value for each observed (i.e., actual) fraction of females among the top 1%, 5%, or 10%, where the p value was estimated based on 20,000 simulations and two-sided.

Study 1: Results

Distribution Pitting

Table 1 summarizes distribution pitting results for Study 1. In support of Hypothesis 3a, results showed that the power law with exponential cutoff distribution, which belongs to the exponential tail distribution category, had the best fit with the data for both the female and male samples. 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; 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 power law with exponential cutoff, Weibull, and lognormal distributions remained after implementing 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 (i.e., power law with exponential cutoff, Weibull, and lognormal) were nested within each other. Finally, we used the third decision rule to rule out the Weibull and lognormal distributions because these are flexible distributions as opposed to the power law with exponential cutoff distribution, which is "inflexible" and thus more parsimonious.

For men, all but the power law with exponential cutoff and Weibull distributions remained after implementing 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).

Log Likelihood Values

As noted in the Method section, we also computed absolute fit (i.e., log likelihood) values based on Equation 8. Results are included in the Appendix. Please recall the caveat mentioned earlier that these values are influenced by sample size. So, the Appendix shows that, overall, larger samples seem to have worse fit (i.e., larger negative values for log likelihood)—which is not necessarily true. So, an appropriate use of log likelihood values in terms of how well a theoretical distribution fits a sample involves keeping sample size constant. As shown in the Appendix, for each of the two genders (i.e., men and women) in the field of mathematics, the power law with an exponential cutoff has a smaller negative value of log likelihood (i.e., has better fit) than that for each of the other distributions.

Table 1
Distribution Pitting Results for Research Productivity of Female and Male Researchers in Study 1 (Mathematics)

Gender	N	Norm vs. PL	Norm vs. Cut PL vs. Cut	Norm vs. Weib PL vs. Weib Cut vs. Weib	Norm vs. LogN PL vs. LogN Cut vs. LogN Weib vs. LogN	Norm vs. Exp PL vs. Exp Cut vs. Exp Weib vs. Exp LogN vs. Exp	Norm vs. Pois PL vs. Pois Cut vs. Pois Weib vs. Pois LogN vs. Pois Exp vs. Pois
Women	360	−5.69 (0)	−6.10 (0) −3.46 (.01)	−6.14 (0) −1.66 (.10) .54 (.59)	−6.09 (0) −1.55 (.12) 1.08 (.28) .03 (.97)	−6.77 (0) .63 (.53) 1.76 (.08) 1.80 (.07) 1.67 (.10)	−7.71 (0) 1.73 (.08) 2.42 (.02) 2.40 (.02) 2.40 (.02) 2.87 (0)
Men	3,493	−16.55 (0)	−18.28 (0) −83.30 (0)	−18.18 (0) −7.72 (0) .31 (.75)	−18.09 (0) −7.76 (0) .88 (.38) 1.71 (.09)	−20.25 (0) 1.49 (.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. Distribution names are abbreviated as follows: Norm = normal; PL = pure power law; Cut = power law with exponential cutoff; Weib = Weibull; LogN = lognormal; Exp = exponential; Pois = Poisson. Distribution pitting titles are presented such that the first distribution is pitted against the second distribution (e.g., Norm vs. PL = 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.

Fit Parameters and Descriptive Statistics

To test Hypotheses 1b, 2b, 3b, and 4b, we computed the pure power law's parameter (α) and exponential distribution's parameter (λ) using Dpit. Both parameters were larger for women (i.e., lighter tails). For women, the α and λ parameters were $\alpha = 2.94$ (bootstrapped 95% confidence interval from 2.74 to 3.16) and $\lambda = 0.57$ (bootstrapped 95% confidence interval from 0.54 to 0.59), respectively. For men, the α and λ parameters were $\alpha = 2.39$ (bootstrapped 95% confidence interval from 2.35 to 2.44) and $\lambda = 0.47$ (bootstrapped 95% confidence interval from 0.46 to 0.48), respectively. These parameter estimates offer support for Hypothesis 3b in that, although both genders share the same likely dominant generative mechanism of incremental differentiation, the distribution for men has a heavier right tail such that there is a gender productivity gap in favor of men. Indeed, regardless of α or λ, there was no overlap between the bootstrapped confidence interval for female researchers and that for male researchers.

Figure 1 depicts histograms and kernel density plots of the research productivity of female and male researchers in Study 1. For women, the total number of publications ranged from one to seven. Additionally, the top 10% of female performers published within the range of two to seven articles, the top 5% published within the range of three to seven articles, and the top 1% published within the range of five to seven articles. For men, the number of publications ranged from one to 20. The top 10% of male performers published within the range of three to 20 articles, the top 5% published within the range of four to 20 articles, and the top 1% published within the range of eight to 20 articles. This means that each of the top 1% of male researchers, individually, outperformed each of the female researchers in terms of number of publications. In addition, results showed that among all 3,853 mathematics researchers, the top 1% of performers is composed entirely of male researchers, each with eight or more top-tier journal articles. Finally, permutation analyses showed that the actual fraction of women was significantly lower than the expected fraction of women among the top 1% (actual fraction = 0, expected fraction = 0.09, $p < .01$), top 5% (actual fraction = 0.05, expected fraction = 0.09, $p < .01$), and top 10% of all producers of research (actual fraction = 0.06, expected fraction = 0.09, $p < .05$).

Study 2: Method

Sample

We examined the productivity of all researchers in the field of genetics who have published at least one article in one of the five

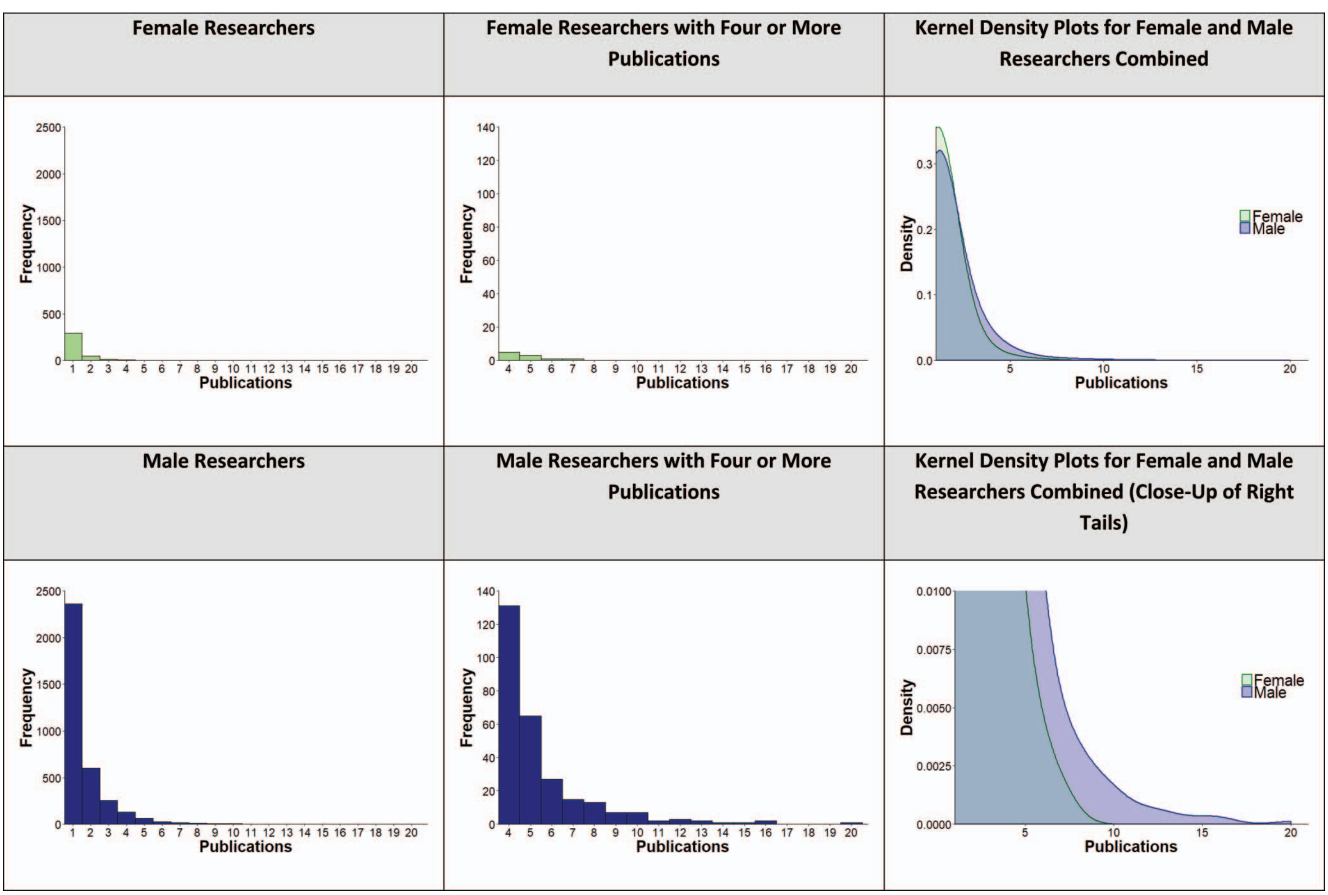

Figure 1. Histograms and kernel density plots of the productivity of 3,853 researchers in Study 1 (mathematics).

most cited genetics journals from January 2006 to December 2015. In contrast to mathematics, the field of genetics has one of the greatest concentrations of women across STEM fields (National Science Foundation, 2016). Study 2 thus complemented Study 1 in that it involved a STEM field but one that may involve different gender dynamics and processes. The sample size for Study 2 was N = 45,007 unique researchers, of whom 14,685 (32.6%) were women.

Journal Selection Criteria

We identified the five most influential journals using the same process as in Study 1 (by assessing mean impact factors from 2011 to 2015). Study 2 included five journals because genetics journals publish many more articles compared with the field of mathematics. The top five journals in the field of genetics were as follows: *Nature Reviews Genetics*, *Nature Genetics*, *Annual Review of Genetics*, *Trends in Ecology & Evolution*, and *Genome Research*. The total number of articles published from January 2006 to December 2015 was 7,746, which is comparable to the number of articles published in applied psychology journals (Kruschke et al., 2012).

Measures

We used the same procedures for gathering information on the number of articles published by each researcher in top-tier journals as in Study 1. Similarly, we followed the same procedures regarding the measurement of gender for each author.

Data Analytic Approach

The data analytic approach was identical to Study 1. In other words, it involved distribution pitting followed by the calculation of log likelihood values, fit parameters, descriptive statistics, bootstrapping, and permutation.

Study 2: Results

Distribution Pitting

Table 2 summarizes the distribution pitting results for Study 2. In support of Hypothesis 3a, results showed 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 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 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 were not nested within one another. Finally, the third decision rule was used to rule out the lognormal distribution (i.e., flexible).

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

Table 2
Distribution Pitting Results for Research Productivity of Female and Male Researchers in Study 2 (Genetics)

Gender	N	Norm vs. PL	Norm vs. Cut PL vs. Cut	Norm vs. Weib PL vs. Weib Cut vs. Weib	Norm vs. LogN PL vs. LogN Cut vs. LogN Weib vs. LogN	Norm vs. Exp PL vs. Exp Cut vs. Exp Weib vs. Exp LogN vs. Exp	Norm vs. Pois PL vs. Pois Cut vs. Pois Weib vs. Pois LogN vs. Pois Exp vs. Pois
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) −.41 (.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) −.38 (.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. Distribution names are abbreviated as follows: Norm = normal; PL = pure power law; Cut = power law with exponential cutoff; Weib = Weibull; LogN = lognormal; Exp = exponential; Pois = Poisson. Distribution pitting titles are presented such that the first distribution is pitted against the second distribution (e.g., Norm vs. PL = 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.

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).

Log Likelihood Values

As with Study 1, we computed absolute fit (i.e., log likelihood) values based on Equation 8. These results are included in the Appendix. For each of the two genders (i.e., men and women) in the field of genetics, the power law with an exponential cutoff has a smaller negative value of log likelihood (i.e., has better fit) than that for each of the other distributions.

Fit Parameters and Descriptive Statistics

As with Study 1, the power law with exponential cutoff distribution was the best fitting one for both gender groups. To assess the heaviness of the distributions' right tails, we estimated the α and λ parameters (i.e., rates of decay) for each gender group. Both parameters were larger for women (i.e., lighter tails). For women, the α and λ parameters were $\alpha = 2.43$ (bootstrapped 95% confidence interval from 2.41 to 2.46) and $\lambda = 0.44$ (bootstrapped 95% confidence interval from 0.43 to 0.45), respectively. For men, the α and λ parameters were $\alpha = 2.30$ (bootstrapped 95% confidence interval from 2.28 to 2.31) and $\lambda = 0.40$ (bootstrapped 95% confidence interval from 0.40 to 0.41), respectively. Regardless of α or λ, there was no overlap between the bootstrapped confidence interval for female researchers and that for male researchers. Thus, in support of Hypothesis 3b, these parameter estimates show that, although both genders share the same likely dominant generative mechanism of incremental differentiation, the right tail is lighter for women than for men.

Figure 2 shows histograms and kernel density plots of the research productivity of female versus male researchers in Study 2. Women exceeded men in terms of their total range of publications (i.e., one to 123 for women vs. one to 102 for men). The top 10% of female researchers published articles within the range of three to 123 articles, the top 5% within the range of five to 123 articles, and the top 1% within the range of 12 to 123 articles. The top 10% of male researchers published within the range of four to 102 articles, the top 5% within the range of six to 102 articles, and the top 1% within the range of 15 to 102 articles. However, among the top 1% of all 45,007 researchers in Study 2, 26.2% were women. Given that women comprised 32.6% of the total sample of genetics

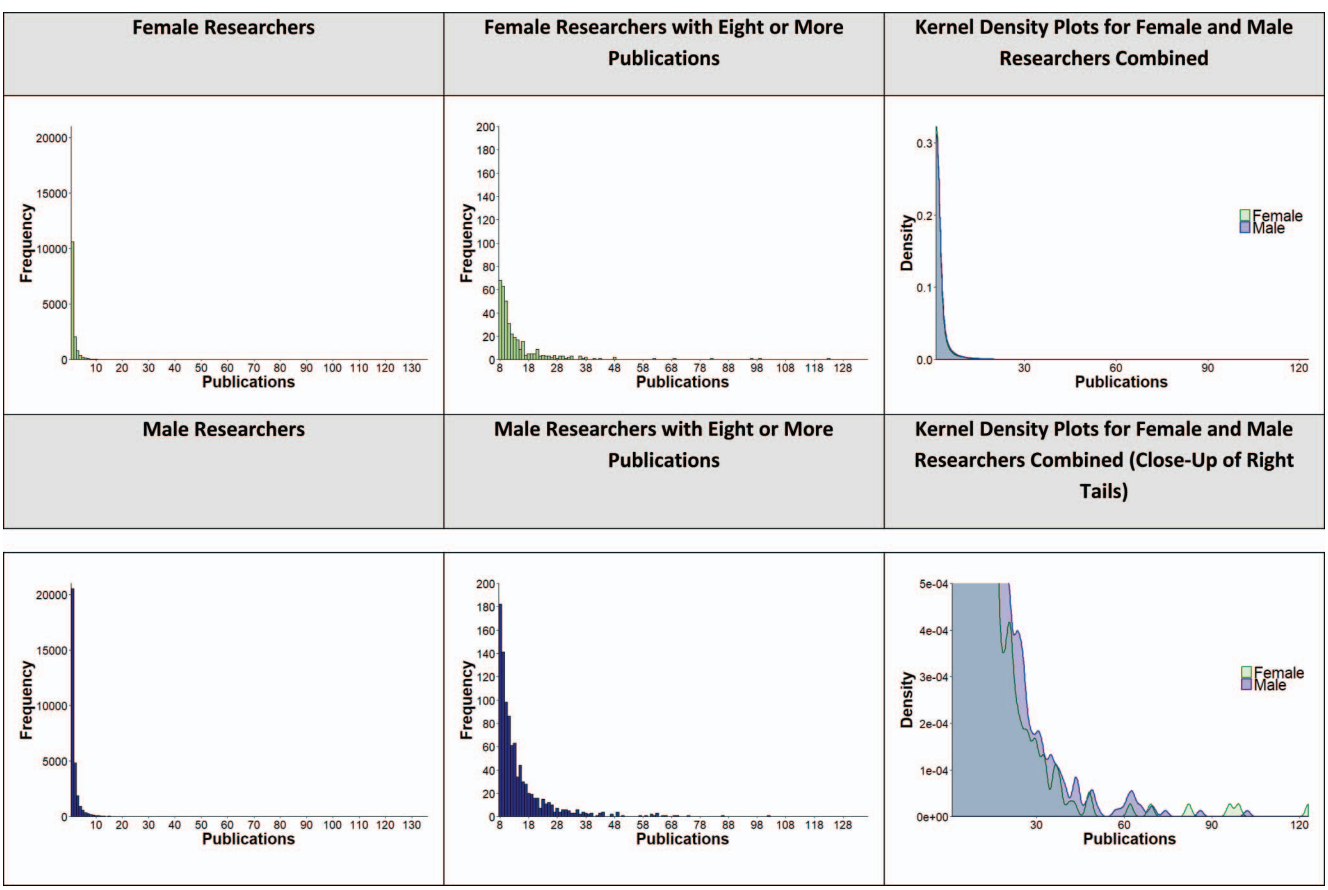

Figure 2. Histograms and kernel density plots of the productivity of 45,007 researchers in Study 2 (genetics).

researchers, the fact that women make up 26.2% of the top 1% genetics researchers in our sample shows that women are underrepresented at the upper tail of the combined productivity distribution. Finally, permutation analyses showed that the actual fraction of women was significantly lower than the expected fraction of women among the top 1% (actual fraction = 0.26, expected fraction = 0.33, $p < .01$), top 5% (actual fraction = 0.28, expected fraction = 0.33, $p < .01$), and top 10% of all producers of research (actual fraction = 0.31, expected fraction = 0.33, $p < .01$).

Study 3: Method

Sample

We examined the productivity distributions of male and female researchers in two psychology subfields: applied psychology and mathematical psychology. We chose applied psychology because of its relevance for a JAP readership, and we chose mathematical psychology because it is more closely related to STEM. Regarding the sample sizes for Study 3, there were 4,081 applied psychology researchers, of whom 1,595 (39.1%) were women. In addition, there were 6,337 mathematical psychology researchers, of whom 2,177 (34.4%) were women. The relative representation of women in these two non-STEM fields is similar to that in genetics as described in Study 2 (i.e., 32.6%) and clearly larger than that in mathematics as described in Study 1 (i.e., 9.3%).

Journal Selection Criteria

We identified the five most influential journals in each of the two psychology subfields using the same process as in Studies 1 and 2 (i.e., based on the mean impact factor from 2011 to 2015). In applied psychology, the top five journals are the following: *Journal of Management*, *Journal of Applied Psychology*, *Organizational Research Methods*, *Personnel Psychology*, and *Journal of Organizational Behavior*. In mathematical psychology, the top five journals are the following: *Psychonomic Bulletin & Review*, *Behavior Research Methods*, *British Journal of Mathematical & Statistical Psychology*, *Psychometrika*, and *Journal of Mathematical Psychology*. The total number of articles published in the top five journals from January 2006 to December 2015 was 2,807 for applied psychology and 3,796 for mathematical psychology.

Measures and Data Analytic Approach

We used the same procedures for gathering information on the number of articles published by each researcher in top-tier journals as in Studies 1 and 2. Similarly, we followed the same procedures regarding the measurement of gender for each author. We also used the same data analytic approach as in Studies 1 and 2.

Study 3: Results

Distribution Pitting

Tables 3 and 4 summarize the distribution pitting results for Study 3. Consistent with the results of Studies 1 and 2, in applied psychology and mathematical psychology and for both gender groups, the power law with exponential cutoff distribution had the best fit with the data. First, for women in applied psychology, 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; the pure power law had significantly worse fit than the power law with exponential cutoff distribution; the Poisson distribution had significantly worse fit than all of the other distributions except

Table 3
Distribution Pitting Results for Research Productivity of Female and Male Researchers in Study 3 (Applied Psychology)

Gender	N	Norm vs. PL	Norm vs. Cut PL vs. Cut	Norm vs. Weib PL vs. Weib Cut vs. Weib	Norm vs. LogN PL vs. LogN Cut vs. LogN Weib vs. LogN	Norm vs. Exp PL vs. Exp Cut vs. Exp Weib vs. Exp LogN vs. Exp	Norm vs. Pois PL vs. Pois Cut vs. Pois Weib vs. Pois LogN vs. Pois Exp vs. Pois
Women	1,595	−15.33 (0)	−16.25 (0)	−16.35 (0)	−16.17 (0)	−18.86 (0)	−25.82 (0)
			−20.01 (0)	−17.23 (0)	−4.01 (0)	4.22 (0)	6.60 (0)
				1.58 (.11)	2.20 (.03)	6.12 (0)	7.25 (0)
					−.55 (.58)	6.37 (0)	7.31 (0)
						6.01 (0)	7.22 (0)
							7.54 (0)
Men	2,486	20.67 (0)	−21.89 (0)	−22.07 (0)	−21.69 (0)	−25.62 (0)	−14.16 (0)
			−40.21 (0)	−4.27 (0)	−5.78 (0)	7.40 (0)	9.87 (0)
				3.45 (0)	4.24 (0)	9.74 (0)	10.38 (0)
					−1.49 (.14)	10.03 (0)	10.47 (0)
						9.45 (0)	10.21 (0)
							10.30 (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. Distribution names are abbreviated as follows: Norm = normal; PL = pure power law; Cut = power law with exponential cutoff; Weib = Weibull; LogN = lognormal; Exp = exponential; Pois = Poisson. Distribution pitting titles are presented such that the first distribution is pitted against the second distribution (e.g., Norm vs. PL = 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 4
Distribution Pitting Results for Research Productivity of Female and Male Researchers in Study 3 (Mathematical Psychology)

Gender	*N*	Norm vs. PL	Norm vs. Cut PL vs. Cut	Norm vs. Weib PL vs. Weib Cut vs. Weib	Norm vs. LogN PL vs. LogN Cut vs. LogN Weib vs. LogN	Norm vs. Exp PL vs. Exp Cut vs. Exp Weib vs. Exp LogN vs. Exp	Norm vs. Pois PL vs. Pois Cut vs. Pois Weib vs. Pois LogN vs. Pois Exp vs. Pois
Women	2,177	–6.90 (0)	–7.03 (0)	–7.04 (0)	–7.00 (0)	–7.29 (0)	–7.77 (0)
			−7.69 (0)	−.78 (.44)	−2.43 (.02)	3.34 (0)	4.07 (0)
				1.39 (.16)	.31 (.75)	4.15 (0)	4.43 (0)
					−1.19 (.23)	4.29 (0)	4.43 (0)
						4.06 (0)	4.38 (0)
							4.54 (0)
Men	4,160	−20.65 (0)	−21.38 (0)	−21.53 (0)	−21.26 (0)	−23.59 (0)	−29.73 (0)
			−35.58 (0)	−2.34 (.02)	−5.45 (0)	8.21 (0)	10.35 (0)
				3.27 (0)	3.49 (0)	10.22 (0)	10.96 (0)
					−2.74 (.01)	10.85 (0)	11.06 (0)
						10.03 (0)	10.89 (0)
							10.99 (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. Distribution names are abbreviated as follows: Norm = normal; PL = pure power law; Cut = power law with exponential cutoff; Weib = Weibull; LogN = lognormal; Exp = exponential; Pois = Poisson. Distribution pitting titles are presented such that the first distribution is pitted against the second distribution (e.g., Norm vs. PL = 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.

normal; and the exponential distribution had significantly worse fit than the power law with exponential cutoff and Weibull distributions. Thus, the power law with exponential cutoff, lognormal, and Weibull 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 one another. Finally, the third decision rule was used to rule out the lognormal and Weibull distributions (i.e., flexible).

For men in applied psychology, the normal, pure power law, Weibull, Poisson, 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 Weibull distribution had significantly worse fit than the lognormal distribution; the Poisson distribution had significantly worse fit than all of the other distributions except normal; and the exponential distribution had significantly worse fit than the power law with exponential cutoff, Weibull, and lognormal 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 any of the remaining distributions because they are not nested within one another. Finally, the third decision rule was used to rule out the lognormal distribution (i.e., flexible).

Next, for both men and women in mathematical psychology, the normal, pure power law, Weibull, Poisson, 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 Weibull distribution had significantly worse fit than the pure power law, power law with exponential cutoff, and lognormal distributions; the Poisson distribution had significantly worse fit than all of the other distributions except normal; and the exponential distribution had significantly worse fit than the pure power law, power law with exponential cutoff, Weibull, and lognormal 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 any of the remaining distributions because they are not nested within one another. Finally, the third decision rule was used to rule out the lognormal distribution (i.e., flexible).

Log Likelihood Values

As with Studies 1 and 2, we computed absolute fit (i.e., log likelihood) values based on Equation 8. These results are included in the Appendix. For each of the two non-STEM fields (i.e., applied psychology and mathematical psychology) and for each of the two genders (i.e., men and women), the power law with an exponential cutoff has a smaller negative value of log likelihood (i.e., has better fit) than that for each of the other distributions.

Fit Parameters and Descriptive Statistics

As with Studies 1 and 2, the power law with exponential cutoff distribution was the best fitting one for both gender groups. To assess the heaviness of the distributions' right tails, we estimated the α and λ parameters (i.e., rates of decay) for each gender group. In both the applied and mathematical psychology subfields, the α and λ parameters were greater for women (i.e., lighter tails). For women in applied psychology, the α and λ parameters were α = 2.40 (bootstrapped 95% confidence interval from 2.33 to 2.47) and λ = 0.46 (bootstrapped 95% confidence interval from 0.44 to 0.47), respectively. For men in applied psychology, the α and λ parameters were α = 2.14 (bootstrapped 95% confidence interval from 2.10 to 2.18) and λ = 0.37 (bootstrapped 95% confidence interval from 0.36 to 0.39), respectively. Next, for women in mathematical psychology, the α and λ parameters were α = 2.95 (bootstrapped 95% confidence interval from 2.86 to 3.04) and λ = 0.56 (bootstrapped 95% confidence interval from 0.55 to 0.57), respectively. For men in mathematical psychology, the α and λ parameters were α = 2.41 (bootstrapped 95% confidence interval from 2.37 to 2.45) and λ = 0.45 (bootstrapped 95% confidence interval from 0.44 to 0.47), respectively. For each of the two non-STEM fields, regardless of α or λ, there was no overlap between the bootstrapped confidence interval for female researchers and that for male researchers. These parameter estimates are consistent with the results of Studies 1 and 2 in that, while both genders share the same likely dominant generative mechanism of incremental differentiation, the right tail is lighter for women than for men.

Figures 3 and 4 depict histograms and kernel density plots of the research productivity of female versus male researchers in Study 3. In both fields, men exceeded women in terms of their total range of publications: one to 19 for women versus one to 35 for men in applied psychology, and one to 21 for women versus one to 32 for men in mathematical psychology. In applied psychology, the top 10% of female researchers published articles within the range of three to 19, the top 5% within the range of five to 19, and the top 1% within the range of 10 to 19. The top 10% of male researchers published within the range of five to 35 articles, the top 5% within the range of seven to 35 articles, and the top 1% within the range of 16 to 35 articles. In addition, among the top 1% of all 4,081 researchers in applied psychology, only 14.6% were women, although women made up 39% of the combined sample. Permutation analyses showed that the actual fraction of women was significantly lower than the expected fraction of women among the top 1% (actual fraction = 0.15, expected fraction = 0.39, $p < .01$), top 5% (actual fraction = 0.31, expected fraction = 0.39, $p < .05$), and top 10% of all producers of research (actual fraction = 0.33, expected fraction = 0.39, $p < .01$).

Next, in mathematical psychology, the top 10% of female researchers published articles within the range of two to 21, the top 5% within the range of three to 21, and the top 1% within the range of six to 21. The top 10% of male researchers published within the range of three to 32 articles, the top 5% within the range of five to 32 articles, and the top 1% within the range of 10 to 32 articles. In addition, among the top 1% of all 6,337 researchers in mathematical psychology, only 6.3% were women, although women made up 34% of the combined sample. Permutation analyses showed that the actual fraction of women was significantly lower than the expected fraction of women among the top 1% (actual fraction =

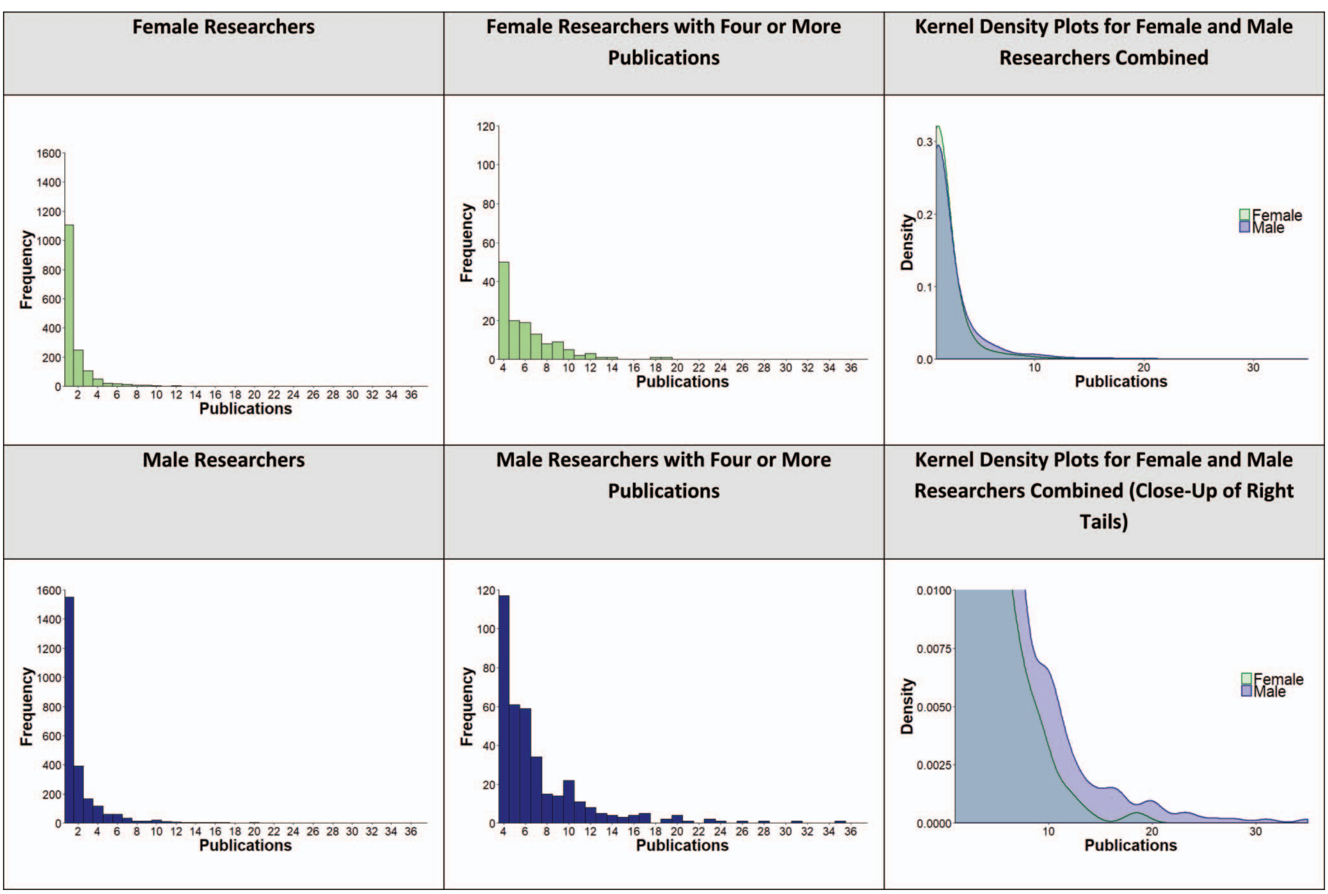

Figure 3. Histograms and kernel density plots of the productivity of 4,081 researchers in Study 3 (applied psychology).

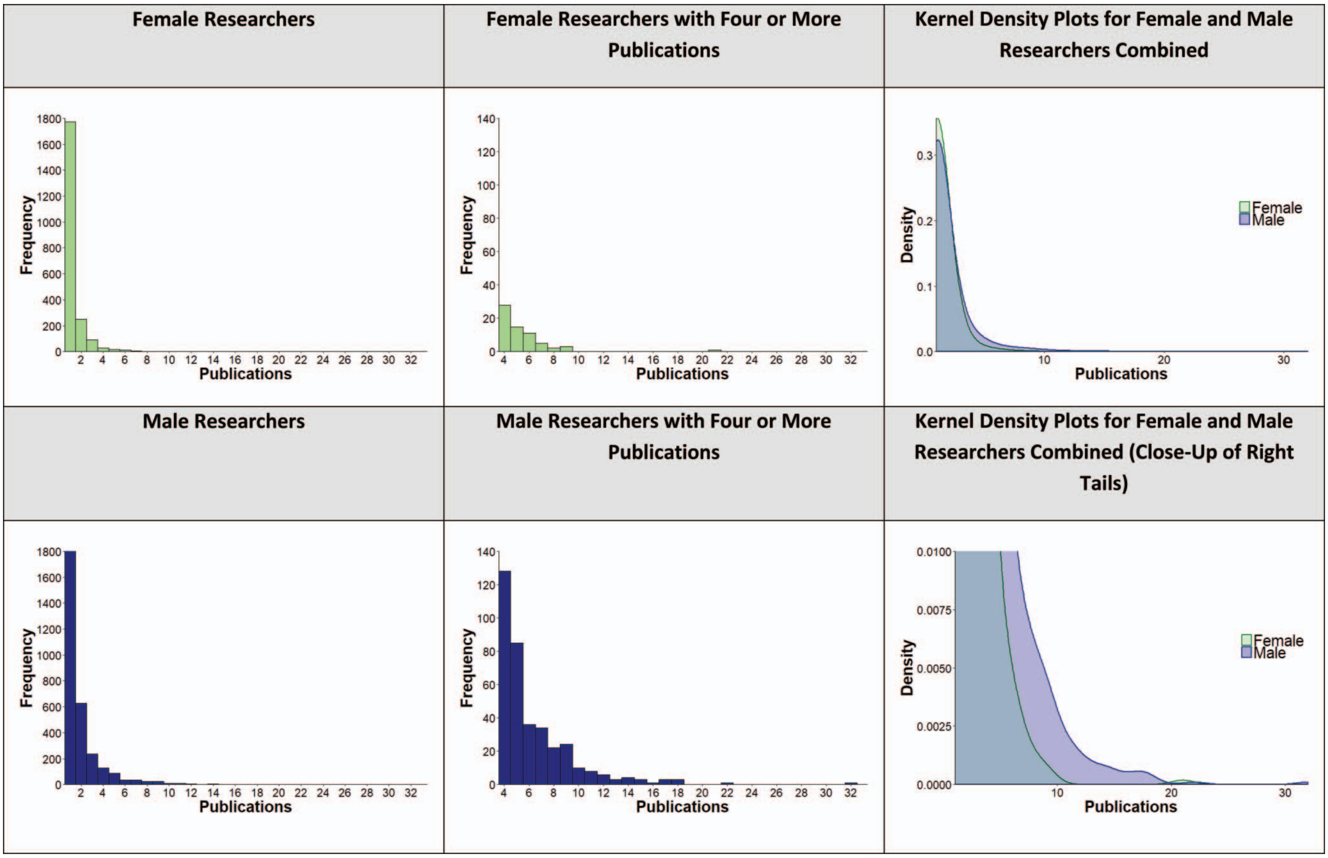

Figure 4. Histograms and kernel density plots of the productivity of 6,337 researchers in Study 3 (mathematical psychology).

0.06, expected fraction = 0.34, $p < .01$), top 5% (actual fraction = 0.21, expected fraction = 0.34, $p < .01$), and top 10% of all producers of research (actual fraction = 0.24, expected fraction = 0.34, $p < .01$).

General Discussion

Regarding Hypotheses 1a, 2a, 3a, and 4a, in Study 1 (mathematics), Study 2 (genetics), and Study 3 (applied psychology and mathematical psychology), we found that the power law with an exponential cutoff distribution best fits the productivity distributions of both male and female researchers. Given that incremental differentiation is the generative mechanism associated with this particular distribution (i.e., power law with exponential cutoff), findings suggest that, for both genders, individual variation in cumulative productivity is predominantly driven by differences in accumulation rates (i.e., average amount of a result produced per time period).

Regarding Hypotheses 1b, 2b, 3b, and 4b, we found significant gender-based differences in the heaviness of the distributions' right tails (i.e., lighter for women). Moreover, we found that the underrepresentation of women is more and more extreme as we consider more elite ranges of performance (i.e., top 10%, 5%, and 1% of performers). We include a summary of results regarding the underrepresentation of women in Table 5. For example, considering the top 1% of researchers, there was not even one woman in mathematics, only 26.2% were women in genetics, only 14.6% in applied psychology, and only 6.3% in mathematical psychology. This pattern of female underrepresentation among the top 1% of researchers stands in contrast to the overall female representation in our mathematics (9.3%), genetics (32.6%), applied psychology (39.1%), and mathematical psychology (34.4%) samples.

Implications for Theory

First, results help us understand the likely generative mechanism leading to the emergence of star performers in STEM and other scientific fields. To identify the likely generative mechanism, we adopted an epistemological approach based on falsification, which allowed us to rule out mechanisms that did not fit the data as well. Results were consistent across all studies and revealed that the generative mechanism of incremental differentiation is likely dominant for both gender groups. Because the incremental differentiation mechanism means that individuals' output increases at an approximately linear rate based on their accumulation rates, we conclude that star researchers vary in their cumulative productivity largely because of individual differences in accumulation rates, which allow some individuals to enjoy greater output increments than others. Note that because distribution pitting is a falsification-based approach, we conclude that incremental differentiation is

Table 5
Summary of Results Regarding the Underrepresentation of Women in Study 1 (Mathematics), Study 2 (Genetics), and Study 3 (Applied Psychology and Mathematical Psychology)

	Study 1: Mathematics		Study 2: Genetics		Study 3: Applied Psychology		Study 3: Mathematical Psychology	
	Women (N = 360)	Men (N = 3,493)	Women (N = 14,685)	Men (N = 30,322)	Women (N = 1,595)	Men (N = 2,486)	Women (N = 2,177)	Men (N = 4,160)
Parameter alpha and 95% CI	α = 2.94	α = 2.39	α = 2.43	α = 2.30	α = 2.40	α = 2.14	α = 2.95	α = 2.41
	[2.74, 3.16]	[2.35, 2.44]	[2.41, 2.46]	[2.28, 2.31]	[2.33, 2.47]	[2.10, 2.18]	[2.86, 3.04]	[2.37, 2.45]
Parameter lambda and 95% CI	λ = .57	λ = .47	λ = .44	λ = .40	λ = .46	λ = .37	λ = .56	λ = .45
	[.54, .59]	[.46, .48]	[.43, .45]	[.40, .41]	[.44, .47]	[.36, .39]	[.55, .57]	[.44, .47]
Actual vs. expected number and percent among the top 10% of producers	24 vs. 36	361 vs. 349	1,377 vs. 1,467	3,123 vs. 3,033	133 vs. 160	275 vs. 249	214 vs. 304	668 vs. 579
	6.2% vs. 9.3%	93.8% vs. 90.7%	30.6% vs. 32.6%	69.4% vs. 67.4%	32.6% vs. 39.1%	67.4% vs. 60.9%	24.3% vs. 34.4%	75.7% vs. 65.6%
Actual vs. expected number and percent among the top 5% of producers	10 vs. 18	183 vs. 175	637 vs. 734	1,614 vs. 1,517	63 vs. 80	141 vs. 124	90 vs. 152	351 vs. 289
	5.2% vs. 9.3%	94.8% vs. 90.7%	28.3% vs. 32.6%	71.7% vs. 67.4%	30.9% vs. 39.1%	69.1% vs. 60.9%	20.5% vs. 34.4%	79.5% vs. 65.6%
Actual vs. expected number and percent among the top 1% of producers	0 vs. 4	39 vs. 35	118 vs. 147	332 vs. 303	6 vs. 16	35 vs. 25	6 vs. 30	83 vs. 58
	0% vs. 9.3%	100% vs. 90.7%	26.2% vs. 32.6%	73.8% vs. 67.4%	14.6% vs. 39.1%	85.4% vs. 60.9%	6.3% vs. 34.4%	93.7% vs. 65.6%

Note. N = sample size; CI = confidence interval. Estimates of the pure power law's α and exponential distribution's λ parameters for each sample are shown in Rows 1–2. Each entry is followed by the parameter value's bootstrapped 95% confidence intervals (in brackets). The smaller the α and λ parameters, the heavier is the distribution's right tail. Rows 3–5 display the actual versus expected numbers (as well as percentages) of women and men among the top 10%, 5%, and 1% of producers of research for each sample.

dominant over the other generative mechanisms with respect to the observed output distributions. However, this does not mean that other mechanisms are completely irrelevant. It is possible that some researchers experience, to a certain extent, the output shocks of self-organized criticality, output loops of proportionate differentiation, and output homogenization processes that produce (potentially) symmetric distributions. Nonetheless, incremental differentiation consistently emerged as dominant over the other mechanisms. In sum, with regard to the overriding putative cause of the emergence of star researchers, our results provide stronger evidence for explanations based on incremental differentiation and associated output increments over explanations based on output shocks, output loops, or output homogenization.

Second, our results regarding the likely dominance of incremental differentiation provide a building block for further theory development about why individuals vary in their total outputs and how star performers emerge. The incremental differentiation mechanism suggests that researchers differ substantially in their accumulation rates on a set of major variables or components that, together, have a more-or-less linear effect on total research productivity. In other words, researchers with higher accumulation rates on input components such as social capital, scientific knowledge, work hours, and funding may experience larger increments in total research productivity—analogous to the way workers with higher labor productivity enjoy larger increments in accumulated wages. For instance, holding all other factors constant, individuals who accumulate network ties at a greater rate than others may experience larger increments in future productivity, because their larger networks would enable a greater influx of useful research-related advice, resources, and other forms of social capital. In short, each unit of increase in one's accumulation rate would be followed by a concordant increase in total outputs, thus allowing incremental differentiation to occur.

Third, because results regarding the likely dominance of the power law with exponential cutoff were so consistent in mathematics, genetics, applied psychology, and mathematical psychology, our studies contribute to the current and important debate in applied psychology and related fields about the need to produce cumulative knowledge (Kepes & McDaniel, 2013). Moreover, our results regarding the likely dominance of the power law with an exponential cutoff and its associated mechanism of incremental differentiation are also consistent with those of Joo et al. (2017) based on star performers in a variety of industries, occupations, and jobs including sports, entertainment, politics, and manufacturing. Such consistency across different samples and also consistency with past research are particularly important in a domain that is highly contentious such as gender issues and with important implications for organizational practices, policy making, and society in general. As such, our article follows the advice by Cortina, Aguinis, and DeShon (2017) who noted that "our top journals must encourage and publish high-quality constructive replications" (p. 283). In short, we see the consistency in results as an important contribution to the need to produce cumulative knowledge and, more specifically, such consistency provides further evidence of the pervasiveness of incremental differentiation in the production of star performers across various contexts.

Fourth, our findings provide insights into the likely predominant cause of the observed gender productivity gaps among stars in STEM and other scientific fields. To do so, we first integrated conceptualizations regarding the reasons for a gender productivity gap from applied psychology and other social sciences with the literature from the natural sciences regarding distribution shapes and their generative mechanisms. The existence of a gender productivity gap under the incremental differentiation mechanism is most strongly aligned with the gender discrimination explanation for the gender productivity gap. Given such, our results suggest that incremental differentiation occurs to a greater degree among men, because certain forms of discrimination may disproportionately constrain the output increments resulting from women's accumulation rates. For example, two researchers, John and Sally, may be nearly identical in terms of their accumulation rates, and they may produce high quality articles at similar rates. However, John may experience greater increments in his total publications compared with Sally because of gender biases in peer reviews (in favor of men). Similarly, John and Sally may accumulate network ties with other researchers at equivalent rates, yet John may accrue greater benefits (i.e., more publications) from his network because of various gender biases among collaborators and mentors (e.g., tendency to overvalue the work of star men over that of star women). Hence, for each unit of increase in accumulation rate, John is likely to experience greater increments in total productivity compared with Sally. Extending this logic to large groups of individuals, the same distribution of accumulation rates would lead to greater incremental differentiation and thus a more heavily right-tailed productivity distribution for men compared with women. The effects of gender discrimination in analogous to a situation where, between two groups of individuals with similar labor productivity distributions, one group is able to accumulate wages at a higher rate than the other because of systematic biases in salaries (Martell, Lane, & Emrich, 1996).

Fifth, our results demonstrate that the gender productivity gap among stars may be prevalent in scientific domains that do not have the reputation of being traditionally masculine, such as applied psychology. To our surprise, in Study 3, we found a substantial productivity gap favoring men in applied psychology and mathematical psychology. In fact, as shown in Table 5, there is more underrepresentation of female stars in applied psychology than in genetics. Consider the results for the top 1% of performers in each discipline we examined. In genetics, because 32.6% of the entire sample was female, we also expected 32.6% of the top 1% of genetics researchers to be women. However, we found that women make up only 26.2% of the top 1% of genetics researchers. In applied psychology, the underrepresentation of women at the top of the productivity distribution was even greater. We expected 39.1% of the top 1% of applied psychology researchers to be women, but we found that women make up only 14.6% of that group. In retrospect, perhaps the underrepresentation of female stars in both STEM and non-STEM fields is not surprising in light of past research findings. For example, consider the fact that Kozlowski, Chen, and Salas (2017) reviewed articles published in the first 100 years of JAP and concluded that "research published in the journal is increasingly team based and the size of the author teams is increasing over time. This indicates that the quaint notion of the sole investigator/author has long since passed in time" (p. 243). Given the increased number of coauthored articles, the Matilda effect to which we referred earlier (i.e., reviewers and colleagues often undervalue the quality of female scientists' research outputs and are less likely to show collaboration interest toward them) may be relevant in applied psychology as well. These results are also consistent with those of a recent study involving 511 management professors, showing that women were less likely to receive named professorships, even when controlling for publication records and citations (Treviño, Gomez-Mejia, Balkin, & Mixon, 2018). In sum, our results suggest that it is likely that similar gender discrimination processes and "masculine-gendered environments" (Treviño et al., 2018) affect the production of star performers not only in STEM fields but also in non-STEM fields—including those that include a greater overall representation of women in general such as applied psychology.

Finally, stars are highly visible, are sought after, and can play a powerful role in shaping people's attitudes and organizational policies (Aguinis & Bradley, 2015; Call, Nyberg, & Thatcher, 2015). So, our results about the existence of large gender imbalances specifically among stars can explain a chain of consequences that could ripple throughout the entire domain. In particular, the existence of large gender productivity gaps among stars may partially explain why the overall gender representation gap remains more persistent in certain fields over others. For example, the disproportionate dearth of female stars in fields such as mathematics may, in turn, substantially contribute to negative stereotypes about women's math abilities in general, stymie the influx of women into the field, and ultimately perpetuate gender gaps in terms of overall representation and productivity in the field. In comparison, the relatively greater presence of female stars in fields such as genetics may catalyze organizational efforts to mitigate gender biases, stimulate a greater influx of women into the field, and contribute to further improvements in women's overall representation and productivity in the field. Thus, we suggest that future theory development focus on not only better understanding the underlying causes of women's underrepresentation among stars, but also the broader consequences of such underrepresentation.

Implications for Practice

General implications. First, our findings suggest that incremental differentiation is more constrained among women stars than among men stars. Female star researchers may accumulate productivity components at similar rates as their male colleagues but experience smaller increments in productivity because of a myriad of gender biases. According to our results, women stars may need to overaccumulate or "do more" (e.g., acquire more knowledge, build more relationships, put in more research hours) to achieve the same level of increase in outputs as their male counterparts. The present studies' replication of the likely dominance of incremental differentiation is especially important given current discussions about the science-practice divide, the need for applied psychology to provide reproducible research findings to narrow the divide, and the need for applied psychology to have greater impact on organizational practices and interventions (Cascio & Aguinis, 2008; McHenry, 2007).

Second, our results point to the need for organizations to specifically address female underrepresentation among star performers. In academic domains, the difference between publishing one versus two articles in a top-tier journal can be the decisive factor with respect to whether one receives a particular grant, promotion, and other important rewards (e.g., summer funding, teaching reduction). As such, even a seemingly small gender productivity gap may entail large consequences for individual researchers, especially in fields such as mathematics where researchers publish at relatively lower rates. Moreover, a gender productivity gap among stars implies a chain of consequences with trickle-down effects onto other performers. As mentioned previously, many factors have been shown to contribute to the general underrepresentation of women in STEM and other scientific fields, including biological differences, gender discrimination, and different career and lifestyle choices. But, our analyses into the dominant mechanism through which researchers differentiate their productivity suggest that, when pertaining specifically to the disproportionate gender productivity gap among stars, gender discrimination likely plays the strongest role. Given how influential star performers can be, an implication of our results is that organizations may wish to pursue efforts to specifically address star performers—in addition to their focus on the broader issue of promoting greater female participation in general.

Third, our results suggest that simply giving women and girls more opportunities and resources (e.g., via college scholarships, mentoring programs, career workshops), so that more of them enter the field, is at best an indirect solution to the issue of gender productivity gaps specifically among star performers. Indeed, making such resources and opportunities more available to women could certainly attract more female researchers into these fields and thus improve gender diversity in the overall population of researchers. However, once these women enter a particular scientific field, it is disproportionately more challenging for exceptional women to rise to the top of the productivity distribution—which is a separate challenge. Indeed, we found that the proportion of women is lower and lower as we consider higher and higher levels of performance. For example, as summarized in Table 5, 9.3% of mathematics researchers are female in the entire sample, but the percent is 6.2% if we consider the top 10% of mathematics researchers, 5.2% if we consider the top 5%, and zero if we consider the top 1%. In mathematical psychology, the percent of women is 34.4% in the entire sample. Although this 34.4% figure may not be perceived as an extreme underrepresentation, the percent of women in the mathematical psychology sample is 24.3% if we consider the top 10% of researchers, 20.5% if we consider the top 5%, and a much smaller 6.3% if we consider the top 1%. As far as women who currently work in scientific fields are concerned, initiatives designed to make resources and opportunities more available to them are effective only insofar as they present additional options for accumulating productivity components. In other words, they do not address the core of the challenge revealed by our results, which is that women in these fields need to accumulate more in the first place to produce the same amount of output as men. Initiatives aimed at attracting more women into STEM and other scientific fields may thus improve the overall gender representation gap, but they do not necessarily help deconstrain incremental differentiation among high-end female performers. In fact, overreliance on technocratic and top-down approaches to improving women's representation—such as gender quotas or female-exclusive positions—may actually be ineffective in the long run, sending people the message that their female colleagues "are there due to preferential treatment" and providing "an alibi for not modifying attitudes in depth" (Helmer, Schottdorf, Neef, & Battaglia, 2017, p. 11).

Specific implications for organizations. What can organizations and professional fields do to minimize gender discrimination and, therefore, allow incremental differentiation to occur more unconstrained for women? First, academic departments can implement cluster hiring, which refers to the practice of hiring multiple scholars into a department based on shared research interests (Kossek et al., 2017). Compared with methods such as workplace gender quotas, cluster hiring could be more effective in terms of promoting greater gender diversity and cross-gender interactions among researchers while also minimizing tokenism. Several institutions including the University of Illinois, University of Chicago, and North Carolina State University have implemented cluster hiring and found success in terms of retention and socialization of minority members (Sgoutas-Emch, Baird, Myers, Camacho, & Lord, 2016). Nonetheless, most institutions are hesitant to adopt cluster hiring, as it entails a great degree of organizational change and can generate backlash if implemented without sufficient faculty buy-in (Muñoz et al., 2017). Thus, overall, we suggest that cluster hiring could be a challenging yet effective form of intervention if implemented correctly.

Second, organizations may wish to emphasize greater fairness and transparency regarding policies related to hiring, promotion, and funding that can have profound effects on individual researchers' future outputs (Aguinis & Bradley, 2015). In particular, organizations can focus on identifying stars based on objective measures, and then implement policies that guarantee greater opportunities for growth, equally for women and men. For example, academic institutions may implement a policy where individuals who produce above a certain level of output (e.g., number of publications) are given a lower cap on their subsequent allocation of teaching loads and miscellaneous service responsibilities, thus allowing them more time for research. This is especially relevant for high-performing female researchers, who are more likely than their male colleagues to be allocated greater teaching responsibilities and are less likely to be promoted to leadership roles, despite comparable achievements (Niemeier & González, 2004; Xu, 2008). As another example, a policy where high-end producers—again, based on objective measures—are guaranteed a greater allocation of research funding would help minimize any gender-based discrimination in research funding among (potential) stars.

Third, organizations can consider utilizing stars, especially female stars, as a source of mentoring and coaching to help other women in those fields increase their performance. As past research shows, proximity to stars is linked to the career advancement of subordinates (Malhotra & Singh, 2016), and individuals who receive coaching from star performers are more likely to increase their own performance (Aguinis & Bradley, 2015). In particular, mentoring from star performers who are also women may be the most effective in terms of facilitating greater female star emergence, given the homophilous aspects of mentor-protégé relationships (Lockwood, 2006). Current initiatives focusing on mentoring mostly target young women and students, and such programs are less common within professional research institutions. As such, we suggest that institutions implement practices that "facilitate mentoring relationships with their stars that extends through all levels of the organization" (Aguinis & Bradley, 2015, p. 165), especially between female stars and other women.

Fourth, we recommend that organizations consider allocating greater resources toward increasing (female) star retention. Again, although incremental differentiation emerged as the likely dominant mechanism in our analyses, this does not definitively rule out the presence of other mechanisms. Similarly, our results do not suggest that gender discrimination is the sole cause of the observed gender-star gap. In particular, it is possible that a part of the gap reflects the result of greater turnover among female stars because of gender differences in career and life choices. According to Aguinis and O'Boyle (2014), a star's network tends to be less mobile than the star him- or herself. Further, to reduce star turnover, Aguinis and O'Boyle recommended that organizations implement policies that integrate the stars' important network connections into the teams and organizations that the star serves. Research also suggests that female stars may be less mobile than male stars because of gender differences in career choices (Ceci & Williams, 2011; Xu, 2008). Thus, we suggest that the effectiveness of such initiatives may be more pronounced for female stars. For example, turnover among stars—female stars in particular—may be reduced if organizations provide greater assistance to top/desirable candidates' spouses regarding employment within the same university, relocating to a new home, and so on (Aguinis & O'Boyle, 2014).

Finally, as part of their efforts to increase female star representation in their organizations, managers may wish to offer idiosyncratic work arrangements (i.e., I-deals)—defined as working arrangements that are customized to each star performer for the purpose of attracting, developing, motivating, and retaining the star (Aguinis & Bradley, 2015; Aguinis & O'Boyle, 2014; Hornung, Rousseau, Glaser, Angerer, & Weigl, 2010). Examples of I-deals include increasing the number of days that a star can telecommute (e.g., per week), allowing greater flexibility in terms of when to work (e.g., taking the day off on Monday), and providing additional benefits that are not included in the standard benefits package (e.g., company-subsidized daycare, ergonomic equipment), among others. Though the practice of providing considerably higher levels of pay and performance-based pay for stars is likely a powerful way to attract, motivate, and retain an organization's star performers, I-deals can further help the organization improve its attraction, motivation, and retention of elite-level performers by serving as supplementary nonmonetary rewards (Aguinis, Joo, & Gottfredson, 2013). In addition, by increasing the flexibility of a star's schedule with respect to time and location for work, I-deals can help free up extra time for additional training and development, which would lead to even higher levels of star performance in the future. A key caveat here is that managers should clearly communicate to all workers (i.e., to both stars and nonstars) regarding the reasons for providing I-deals to stars. Without legitimate reasons underlying I-deals or without clearly communicating those reasons to all workers, I-deals may increase voluntary turnover, including turnover among nonstars (Aguinis & O'Boyle, 2014). Overall, these recommendations pertain to improvements in the organization's performance management system (Aguinis, 2019).

Limitations and Suggestions for Future Research

First, as noted earlier, the presence of a dominant generative mechanism does not imply that it is the only mechanism at play. Although incremental differentiation was identified as the likely dominant mechanism, this result does not preclude the possibility that other mechanisms also play a role—albeit a lesser one—in creating differences in cumulative productivity. For example, differences in ability may partially contribute to differences in male and female star performers' accumulation rates (e.g., rates at which they acquire advanced scientific knowledge and mastery of various technical skills). It is also possible that differences in ability contribute to the observed gender productivity gaps.

Second, we examined two STEM and two non-STEM fields. Future research could examine additional fields in an attempt to replicate the likely dominance of incremental differentiation as the generative mechanism. For example, there are several subfields within the domain of materials sciences that could be investigated. In fact, we conducted preliminary analyses based on a subset of materials sciences journals and found that, consistent with results from Studies 1 and 2, power law with exponential cutoff was the best-fitting distribution for women (and, therefore, incremental differentiation was likely the underlying dominant generative mechanism). On the other hand, for men, the lognormal and the power law with exponential cutoff distributions had equivalently acceptable fit with the data. Given our results, a clear direction for future research is to replicate the best-fit of the power law with exponential cutoff distribution across other STEM as well as non-STEM fields. Doing so would lead to a better understanding of the emergence of stars and the prevalence of gender productivity gaps among stars.

Concluding Remarks

Adopting a falsification epistemological approach and using the distribution pitting methodology to implement it, we examined the research productivity of 59,278 researchers who have published at least one article in the most cited journals in mathematics, genetics, applied psychology, and mathematical psychology from 2006 to 2015. Results revealed that productivity distributions follow the power law with exponential cutoff for both women and men. This finding points to incremental differentiation as the likely dominant generative mechanism for the production of star performers because power laws with exponential cutoffs are generated via incremental differentiation. As another unique contribution, results showed that the right-tails of the productivity distributions are significantly lighter for women compared with men—across all the scientific fields we examined. This finding indicates that the gender productivity gap is even more extreme for star performers and, more specifically, underrepresentation of women is more and more extreme as we move higher and higher along the productivity continuum. Taken together, our results make a contribution to our understanding of the emergence of star performers and the predominant reason for the existence of a gender productivity gap among star performers: gender-based differences in accumulation rates, which are better explained by gender discrimination than gender-based differences in abilities or career and life choices.

Our results also suggest that interventions aimed at reducing constraints for incremental differentiation among women can be useful for narrowing the gender productivity gap specifically among star performers. Narrowing this gap is especially important among stars because they are highly visible and play a powerful role in shaping people's attitudes and organizational policies. Overall, based on the finding that women are even more underrepresented among stars, our results highlight the urgent need to address the gender productivity gap in STEM and other scientific fields.

References

Aguinis, H. (2019). *Performance management* (4th ed.). Chicago, IL: Chicago Business Press. https://chicagobusinesspress.com/book/perfmgmt4

Aguinis, H., & Bradley, K. J. (2015). The secret sauce for organizational success: Managing and producing star performers. *Organizational Dynamics, 44,* 161–168. http://dx.doi.org/10.1016/j.orgdyn.2015.05.001

Aguinis, H., Cascio, W. F., & Ramani, R. S. (2017). Science's reproducibility and replicability crisis: International business is not immune. *Journal of International Business Studies, 48,* 653–663. http://dx.doi.org/10.1057/s41267-017-0081-0

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. http://dx.doi.org/10.1177/0149206313478188

Aguinis, H., Joo, H., & Gottfredson, R. K. (2013). What monetary rewards can and cannot do: How to show employees the money. *Business Horizons, 56,* 241–249. http://dx.doi.org/10.1016/j.bushor.2012.11.007

Aguinis, H., & O'Boyle, E., Jr. (2014). Star performers in twenty-first-century organizations. *Personnel Psychology, 67,* 313–350. http://dx.doi.org/10.1111/peps.12054

Aguinis, H., O'Boyle, E., Jr., Gonzalez-Mulé, E., & Joo, H. (2016). Cumulative advantage: Conductors and insulators of heavy-tailed productivity distributions and productivity stars. *Personnel Psychology, 69,* 3–66. http://dx.doi.org/10.1111/peps.12095

Aguinis, H., Ramani, R. S., & Alabduljader, N. (2018). What you see is what you get? Enhancing methodological transparency in management research. *The Academy of Management Annals, 12,* 83–110. http://dx.doi.org/10.5465/annals.2016.0011

Aguinis, H., Suarez-González, I., Lannelongue, G., & Joo, H. (2012). Scholarly impact revisited. *The Academy of Management Perspectives, 26,* 105–132. http://dx.doi.org/10.5465/amp.2011.0088

Amaral, L. A. N., Scala, A., Barthélémy, M., & Stanley, H. E. (2000). Classes of small-world networks. *Proceedings of the National Academy of Sciences of the United States of America, 97,* 11149–11152. http://dx.doi.org/10.1073/pnas.200327197

Amitrano, D. (2012). Variability in the power-law distributions of rupture events. *The European Physical Journal*. Special Topics, *205,* 199–215. http://dx.doi.org/10.1140/epjst/e2012-01571-9

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist, 73,* 3–25. http://dx.doi.org/10.1037/amp0000191

Araújo, N. A., & Herrmann, H. J. (2010). Explosive percolation via control of the largest cluster. *Physical Review Letters, 105,* 035701. http://dx.doi.org/10.1103/PhysRevLett.105.035701

Bak, P. (1996). *How nature works: The science of self-organized criticality*. New York, NY: Copernicus. http://dx.doi.org/10.1007/978-1-4757-5426-1

Barabási, A. L. (2012). Network science: Luck or reason. *Nature, 489,* 507–508. http://dx.doi.org/10.1038/nature11486

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

Bentley, J. T., & Adamson, R. (2003). *Gender differences in the careers of academic scientists and engineers: A literature review*. Arlington, VA: National Science Foundation.

Brown, C. S., & Stone, E. A. (2016). chap. four-gender stereotypes and discrimination: How sexism impacts development. *Advances in Child Development and Behavior, 50,* 105–133. http://dx.doi.org/10.1016/bs.acdb.2015.11.001

Call, M. L., Nyberg, A. J., & Thatcher, S. M. (2015). Stargazing: An integrative conceptual review, theoretical reconciliation, and extension for star employee research. *Journal of Applied Psychology, 100,* 623–640. http://dx.doi.org/10.1037/a0039100

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. http://dx.doi.org/10.1177/0361684315622645

Cascio, W. F., & Aguinis, H. (2008). Research in industrial and organizational psychology from 1963 to 2007: Changes, choices, and trends. *Journal of Applied Psychology, 93,* 1062–1081. http://dx.doi.org/10.1037/0021-9010.93.5.1062

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. http://dx.doi.org/10.1177/1529100614541236

Ceci, S. J., & Williams, W. M. (2010). Sex differences in math-intensive fields. *Current Directions in Psychological Science, 19,* 275–279. http://dx.doi.org/10.1177/0963721410383241

Ceci, S. J., & Williams, W. M. (2011). Understanding current causes of women's underrepresentation in science. *Proceedings of the National Academy of Sciences of the United States of America, 108,* 3157–3162. http://dx.doi.org/10.1073/pnas.1014871108

Cheryan, S., Ziegler, S. A., Montoya, A. K., & Jiang, L. (2017). Why are some STEM fields more gender balanced than others? *Psychological Bulletin, 143,* 1–35. http://dx.doi.org/10.1037/bul0000052

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. http://dx.doi.org/10.1007/s10439-010-9958-9

Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Review, 51,* 661–703. http://dx.doi.org/10.1137/070710111

Cortina, J. M., Aguinis, H., & DeShon, R. P. (2017). Twilight of dawn or of evening? A century of research methods in the Journal of Applied Psychology. *Journal of Applied Psychology, 102,* 274–290. http://dx.doi.org/10.1037/apl0000163

Crawford, G. C., Aguinis, H., Lichtenstein, B., Davidsson, P., & McKelvey, B. (2015). Power law distributions in entrepreneurship: Implications for theory and research. *Journal of Business Venturing, 30,* 696–713. http://dx.doi.org/10.1016/j.jbusvent.2015.01.001

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. http://dx.doi.org/10.1111/sipr.12010

Edwards, A. W. F. (1972). *Likelihood*. Cambridge, UK: Cambridge University Press.

Eloy, J. A., Svider, P. F., Mauro, K. M., Setzen, M., & Baredes, S. (2012). Impact of fellowship training on research productivity in academic otolaryngology. *The Laryngoscope, 122,* 2690–2694. http://dx.doi.org/10.1002/lary.23749

Fortune. (2017, August 8). *Google's gender problem is actually a tech problem*. Retrieved from http://fortune.com/2017/08/08/google-gender-struggle-tech/

Gray, P. H., & Cooper, W. H. (2010). Pursuing failure. *Organizational Research Methods, 13,* 620–643. http://dx.doi.org/10.1177/1094428109356114

Groshen, E. L. (1991). Five reasons why wages vary among employers. *Industrial Relations, 30,* 350–381. http://dx.doi.org/10.1111/j.1468-232X.1991.tb00793.x

Halpern, D. F. (2000). *Sex differences in cognitive abilities* (3rd ed.). Mahwah, NJ: Lawrence 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. http://dx.doi.org/10.1111/j.1529-1006.2007.00032.x

Helmer, M., Schottdorf, M., Neef, A., & Battaglia, D. (2017). Gender bias in scholarly peer review. *eLife, 6,* e21718. http://dx.doi.org/10.7554/eLife.21718

Hornung, S., Rousseau, D. M., Glaser, J., Angerer, P., & Weigl, M. (2010). Beyond top-down and bottom-up work redesign: Customizing job content through idiosyncratic deals. *Journal of Organizational Behavior, 31,* 187–215. http://dx.doi.org/10.1002/job.625

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,* 975–986. http://dx.doi.org/10.1037/0003-066X.42.11.975

Huzurbazar, V. S. (1948). The likelihood equation, consistency and the maxima of the likelihood function. *Annals of Eugenics, 14,* 185–200. http://dx.doi.org/10.1111/j.1469-1809.1947.tb02394.x

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Diversity. Gender similarities characterize math performance. *Science, 321,* 494–495. http://dx.doi.org/10.1126/science.1160364

Joo, H., Aguinis, H., & Bradley, K. J. (2017). Not all nonnormal distributions are created equal: Improved theoretical and measurement precision. *Journal of Applied Psychology, 102,* 1022–1053. http://dx.doi.org/10.1037/apl0000214

Kepes, S., & McDaniel, M. A. (2013). How trustworthy is the scientific literature in industrial and organizational psychology? *Industrial and Organizational Psychology: Perspectives on Science and Practice, 6,* 252–268. http://dx.doi.org/10.1111/iops.12045

Kile, D. E., & Eberl, D. D. (2003). On the origin of size-dependent and size-independent crystal growth: Influence of advection and diffusion. *The American Mineralogist, 88,* 1514–1521. http://dx.doi.org/10.2138/am-2003-1014

Knobloch-Westerwick, S., Glynn, C. J., & Huge, M. (2013). The Matilda effect in science communication: An experiment on gender bias in publication quality perceptions and collaboration interest. *Science Communication, 35,* 603–625. http://dx.doi.org/10.1177/1075547012472684

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. http://dx.doi.org/10.1006/jvbe.1997.1607

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. http://dx.doi.org/10.1177/0149206316671582

Kozlowski, S. W. J., Chen, G., & Salas, E. (2017). One hundred years of the Journal of Applied Psychology: Background, evolution, and scientific trends. *Journal of Applied Psychology, 102,* 237–253. http://dx.doi.org/10.1037/apl0000192

Kruschke, J. K., Aguinis, H., & Joo, H. (2012). The time has come: Bayesian methods for data analysis in the organizational sciences. *Organizational Research Methods, 15,* 722–752. http://dx.doi.org/10.1177/1094428112457829

Kwiek, M. (2018). High research productivity in vertically undifferentiated higher education systems: Who are the top performers? *Scientometrics, 115,* 415–462. http://dx.doi.org/10.1007/s11192-018–2644-7

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. 205–229). New York, NY: Cambridge University Press. http://dx.doi.org/10.1007/978-94-010-1863-0_14

Lincoln, A. E., Pincus, S., Koster, J. B., & Leboy, P. S. (2012). The Matilda effect in science: Awards and prizes in the U. S., 1990s and 2000s. *Social Studies of Science, 42,* 307–320. http://dx.doi.org/10.1177/0306312711435830

Lockwood, P. (2006). "Someone like me can be successful": Do college students need same gender role models? *Psychology of Women Quarterly, 30,* 36–46. http://dx.doi.org/10.1111/j.1471-6402.2006.00260.x

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. http://dx.doi.org/10.1111/j.0030-1299.2007.15686.x

Malhotra, P., & Singh, M. (2016). Indirect impact of high performers on the career advancement of their subordinates. *Human Resource Management Review, 26,* 209–226. http://dx.doi.org/10.1016/j.hrmr.2016.01.002

Martell, R. F., Lane, D. M., & Emrich, C. (1996). Male-female differences: A computer simulation. *American Psychologist, 51,* 157–158. http://dx.doi.org/10.1037/0003-066X.51.2.157

Mason, M. A., & Goulden, M. (2009). *UC doctoral student career and life survey*. Berkeley, CA: UC Berkeley.

McHenry, J. (2007, April). *We are the very model*. Presidential address delivered at the Annual Meeting of the Society for Industrial and Organizational Psychology, New York, New York.

Merton, R. K. (1968). The Matthew effect in science. *Science, 159,* 56–63. http://dx.doi.org/10.1126/science.159.3810.56

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences of the United States of America, 109,* 16474–16479. http://dx.doi.org/10.1073/pnas.1211286109

Muñoz, S. M., Basile, V., Gonzalez, J., Birmingham, D., Aragon, A., Jennings, L., & Gloeckner, G. (2017). (Counter) narratives and complexities: Critical perspectives from a university cluster hire focused on diversity, equity, and inclusion. *Journal of Critical Thought and Praxis, 6,* 1–21.

National Science Foundation. (2016). *Science and engineering indicators*. Retrieved from https://www.nsf.gov/statistics/2016/nsb20161/uploads/1/nsb20161.pdf

Niemeier, D. A., & González, C. (2004). Breaking into the guildmasters' club: What we know about women science and engineering department chairs at AAU universities. *National Women's Studies Association Journal, 16,* 157–171. http://dx.doi.org/10.2979/NWS.2004.16.1.157

Nirei, M., & Souma, W. (2007). A two factor model of income distribution dynamics. *Review of Income and Wealth, 53,* 440–459. http://dx.doi.org/10.1111/j.1475-4991.2007.00242.x

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

Reuben, E., Sapienza, P., & Zingales, L. (2014). How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences of the United States of America, 111,* 4403–4408. http://dx.doi.org/10.1073/pnas.1314788111

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. http://dx.doi.org/10.1177/0963721410391442

Rossiter, M. W. (1993). The Matthew effect in science. *Social Studies of Science, 23,* 325–341. http://dx.doi.org/10.1177/030631293023002004

Sarsons, H. (2017). *Recognition for group work* [Working paper]. Retrieved from https://scholar.harvard.edu/files/sarsons/files/groupwork_full_v6.pdf

Sgoutas-Emch, S., Baird, L., Myers, P., Camacho, M., & Lord, S. (2016). We're not all white men: Using a cohort/cluster approach to diversify STEM faculty hiring. *Thought & Action, 32,* 91–107.

Singh, R., Zhang, Y., Wan, M., & Fouad, N. A. (in press). Why do women engineers leave the engineering profession? The roles of work-family conflict, occupational commitment, and perceived organizational support. *Human Resource Management*. Advance online publication. http://dx.doi.org/10.1002/hrm.21900

Spear, D., & Chown, S. L. (2008). Taxonomic homogenization in ungulates: Patterns and mechanisms at local and global scales. *Journal of Biogeography, 35,* 1962–1975. http://dx.doi.org/10.1111/j.1365-2699.2008.01926.x

Taleb, N. N. (2007). *The black swan: The impact of the highly improbable*. New York, NY: Random House.

Tenenbaum, H. R., & Leaper, C. (2003). Parent-child conversations about science: The socialization of gender inequities? *Developmental Psychology, 39,* 34–47. http://dx.doi.org/10.1037/0012-1649.39.1.34

Treviño, L. J., Gomez-Mejia, L. R., Balkin, D. B., & Mixon, F. G., Jr. (2018). Meritocracies or masculinities? The differential allocation of named professorships by gender in the academy. *Journal of Management, 44,* 972–1000. http://dx.doi.org/10.1177/0149206315599216

van Eck Peluchette, J., & Jeanquart, S. (2000). Professionals' use of different mentor sources at various career stages: Implications for career success. *The Journal of Social Psychology, 140,* 549–564. http://dx.doi.org/10.1080/00224540009600495

Wai, J., Cacchio, M., Putallaz, M., & Makel, M. C. (2010). Sex differences in the right tail of cognitive abilities: A 30-year examination. *Intelligence, 38,* 412–423. http://dx.doi.org/10.1016/j.intell.2010.04.006

Walton, G. M., Murphy, M. C., & Ryan, A. M. (2015). Stereotype threat in organizations: Implications for equity and performance. *Annual Review of Organizational Psychology and Organizational Behavior, 2,* 523–550. http://dx.doi.org/10.1146/annurev-orgpsych-032414-111322

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. http://dx.doi.org/10.1016/j.dr.2013.08.001

Xie, Y., & 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. http://dx.doi.org/10.1007/s11162-008-9097-4

Yam, K. C., Fehr, R., Keng-Highberger, F. T., Klotz, A. C., & Reynolds, S. J. (2016). Out of control: A self-control perspective on the link between surface acting and abusive supervision. *Journal of Applied Psychology, 101,* 292–301.

Appendix

Log Likelihood (i.e., Absolute Fit) Values of Each Sample to Each Theoretical Distribution

	Study 1: Mathematics		Study 2: Genetics		Study 3: Applied Psychology		Study 3: Mathematical Psychology	
Theoretical distribution	Women (N = 360)	Men (N = 3,493)	Women (N = 14,685)	Men (N = 30,322)	Women (N = 1,595)	Men (N = 2,486)	Women (N = 2,177)	Men (N = 4,160)
Pure power law	−256	−3,892	−15,836	−36,883	−1,772	−3,514	−1,534	−4,582
Lognormal	−287	−4,189	−18,772	−42,520	−1,983	−3,895	−1,793	−5,220
Power law with exponential cutoff	−253	−3,808	−15,812	−36,756	−1,752	−3,474	−1,526	−4,547
Exponential	−570	−6,172	−26,982	−58,056	−2,855	−4,963	−3,464	−7,482
Normal	−423	−6,207	−37,041	−78,207	−3,076	−6,063	−3,001	−8,401
Poisson	−447	−5,372	−28,774	−64,848	−2,603	−5,385	−2,785	−6,997
Weibull	−407	−5,378	−26,116	−56,706	−2,582	−4,814	−2,699	−6,878

Note. N = sample size. The smaller the log likelihood's negative value, the better is the sample's fit to the theoretical distribution (see Equation 8). As a cautionary note, it is not appropriate to compare log likelihood values across different samples with different sizes because a log likelihood value is a function of not only fit but also sample size. So, a log likelihood value indicates how well a theoretical distribution fits a sample as long as it is compared with other log likelihood values for other theoretical distributions given the same sample size.

Received September 18, 2017
Revision received May 10, 2018
Accepted May 10, 2018 ■