Journal of Applied Psychology
2017, Vol. 102, No. 7, 1022–1053

0021-9010/17/$12.00 http://dx.doi.org/10.1037/apl0000214

# Not All Nonnormal Distributions Are Created Equal: Improved Theoretical and Measurement Precision

Harry Joo
University of Dayton

Herman Aguinis
George Washington University

Kyle J. Bradley
Indiana University

We offer a four-category taxonomy of individual output distributions (i.e., distributions of cumulative results): (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, Poisson, and Weibull). The four categories are uniquely associated with mutually exclusive generative mechanisms: self-organized criticality, proportionate differentiation, incremental differentiation, and homogenization. We then introduce *distribution pitting*, a falsification-based method for comparing distributions to assess how well each one fits a given data set. In doing so, we also introduce decision rules to determine the likely dominant shape and generative mechanism among many that may operate concurrently. Next, we implement distribution pitting using 229 samples of individual output for several occupations (e.g., movie directors, writers, musicians, athletes, bank tellers, call center employees, grocery checkers, electrical fixture assemblers, and wirers). Results suggest that for 75% of our samples, exponential tail distributions and their generative mechanism (i.e., incremental differentiation) likely constitute the dominant distribution shape and explanation of nonnormally distributed individual output. This finding challenges past conclusions indicating the pervasiveness of other types of distributions and their generative mechanisms. Our results further contribute to theory by offering premises about the link between past and future individual output. For future research, our taxonomy and methodology can be used to pit distributions of other variables (e.g., organizational citizenship behaviors). Finally, we offer practical insights on how to increase overall individual output and produce more top performers.

*Keywords:* performance, output, taxonomy, generative mechanism, falsification

*Supplemental materials:* http://dx.doi.org/10.1037/apl0000214.supp



Recent research has found that individual output generally follows a nonnormal and heavily right-tailed distribution in various jobs, occupations, industries, and types of individual output measures (Aguinis, O'Boyle, Gonzalez-Mulé, & Joo, 2016; O'Boyle & Aguinis, 2012). These studies have replicated the nonnormality of individual output distributions using data consisting of researchers in more than 50 academic disciplines; actors, actresses, directors, choreographers, and lighting specialists in the movie and TV industries; fiction and nonfiction writers; musicians; elected officials in the United States and other country-level legislative bodies (e.g., Australia, Canada, Ireland, and Estonia); professional and collegiate athletes in football, baseball, basketball, cricket, swimming, track and field, skiing, tennis, and other sports; and many other types of workers including bank tellers, call center employees, grocery checkers, pelt pullers, electrical fixture assemblers, and wirers.

The nonnormality of individual output distributions is an important discovery because it affects the foundations of many organizational science theories and practices, including personnel selection, training and development, leadership, turnover, teamwork, compensation, motivation, organizational commitment, job satisfaction, and justice (Aguinis & O'Boyle, 2014). Not surprisingly, evidence that individual output follows nonnormal distribu-

This article was published Online First March 23, 2017.

Harry Joo, Department of Management and Marketing, School of Business Administration, University of Dayton; Herman Aguinis, Department of Management, School of Business, George Washington University; Kyle J. Bradley, Department of Management and Entrepreneurship, Kelley School of Business, Indiana University.

The National Science Foundation (Science of Organizations Program)—grant number 1643075 titled "Understanding the Gender Performance Gap among Star Performers in STEM Fields" to Herman Aguinis (PI)—supported part of the work described in the manuscript.

We thank Takuya Noguchi for his extensive help in compiling our R scripts into a proper R package format for multiple operating systems. We owe much gratitude to Cosma R. Shalizi, Aaron Clauset, and Yogesh Virkar for their kind responses to our questions regarding their code, on which we relied heavily to create our R package called Dpit (available on http://www.hermanaguinis.com or the Comprehensive R Archive Network [CRAN]).

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

tions has stimulated immediate and increasing scholarly attention (Beck, Beatty, & Sackett, 2014; Call, Nyberg, & Thatcher, 2015; Vancouver, Li, Weinhardt, Steel, & Purl, 2016). Results from this research stream have also received media attention from *Forbes*, *Bloomberg Businessweek*, *National Public Radio*, and many other outlets. Clearly, the nonnormality of individual output distributions is of great interest to both researchers and practitioners.

Despite the valuable knowledge produced to date, current methodological procedures lack precision and thus constitute an obstacle to better understanding the nonnormality of individual output distributions and its implications. For instance, one common methodology has been to use the chi-square statistic to determine whether the normal distribution or a particular nonnormal distribution offers a better description of individual output distributions (e.g., Beck et al., 2014; O'Boyle & Aguinis, 2012). Improving on these approaches that conceptualize normality and nonnormality as a dichotomy, subsequent research has defined nonnormality as a continuous construct and adopted pure power law distribution's parameter alpha ($\alpha$) to more precisely assess the shapes of distributions (Aguinis, Martin, Gomez-Mejia, O'Boyle, & Joo, 2017; Aguinis et al., 2016; Crawford, Aguinis, Lichtenstein, Davidsson, & McKelvey, 2015). Nonetheless, these recent improvements still lack precision because they rely on the premise that *all* nonnormal distributions fit a pure power law distribution. In fact, Aguinis et al. (2016) acknowledged this point explicitly: "We use the term [pure] power law to refer to those heavy-tailed distributions where output is clearly dominated by a small group of elites and most individuals in the distribution are far to the left of the mean" (p. 28). Thus, a necessary next step in this line of research is to specify the meaning of "nonnormality" more precisely.

Understanding the precise nonnormal shape of individual output distributions is not a mere methodological curiosity. To the contrary, enhanced precision regarding the presence of a particular distribution offers information on the mechanism that led to the distribution and, therefore, is critical for theory development and testing about when, why, and how a certain distribution exists. In support of this statement, increased precision in the definition and measurement of distributions has resulted in important theory advancements in physics, computer science, biology, engineering, and economics (Mitzenmacher, 2004; Newman, 2005). For example, research in network science has shown that the type or amount of harm to which a network structure is vulnerable is directly related to the precise nonnormal shape of the underlying distribution regarding each vertex's number of links (i.e., an individual's ties with other individuals; Barabási & Bonabeau, 2003; Mossa, Barthélémy, Stanley, & Amaral, 2002). As another example, research in economics and finance has documented that the use of incorrect distributions for modeling financial deviations (e.g., severe economic crises) exposes investments to enormous amounts of risk, and seemingly similar nonnormal distributions are poor substitutes for one another (Mandelbrot & Taleb, 2010; Taleb, 2007).

The goal of our study is to examine the extent to which individual output follows one or more distinct distributions, including multiple types of nonnormal distributions. Our theoretical approach is to develop a taxonomy containing four categories of distributions, where each category is uniquely associated with a distinct generative mechanism. We then assess empirically which distribution is better at representing individual output. In terms of methodology, we introduce and use what we refer to as *distribution pitting*, which compares distributions to assess how well each one fits a given data set. The epistemological basis of distribution pitting is falsification (Gray & Cooper, 2010; Lakatos, 1976; Leavitt, Mitchell, & Peterson, 2010; Popper, 1959)—ideally suited for theoretical domains with too many untested or undertested theories (Aguinis & Vandenberg, 2014; Hambrick, 2007). Although falsification serves as the foundation of distribution pitting, we acknowledge that the shape of an individual output distribution may be the result of multiple generative mechanisms rather than a single one. So, when implementing distribution pitting, we also introduce and use decision rules to determine the likely dominant shape and generative mechanism per observed individual output distribution.

Results suggest that, for 75% of our samples, exponential tail distributions and their generative mechanism (i.e., incremental differentiation) likely constitute the dominant distribution shape and explanation of nonnormally distributed individual output. Our findings thus contribute to the individual performance literature by suggesting that certain generative mechanisms are not as critical for explaining the existence of individual output distributions. In particular, results challenge past conclusions indicating the pervasiveness of the pure power law distribution and its generative mechanism (i.e., self-organized criticality). Reducing theory this way creates greater theoretical parsimony and allows researchers to "reduce areas of focus and avoid time spent on fruitless avenues of inquiry" (Leavitt et al., 2010, p. 645).

Our results also lead to theoretical premises regarding the link between past and future individual output. First, not all types of past individual output predict future individual output. Instead, past individual output in terms of output accumulation rate, but not initial output, predicts future individual output. Second, high variability in individual output would be followed by even higher, not lower, variability in individual output in the future. These premises, in turn, contribute to efforts to go beyond the dominant perspective that focuses on knowledge, skills, abilities, and other individual characteristics (KSAOs) as predictors of future individual output—a model that "seems to have reached a ceiling or plateau in terms of its ability to make accurate predictions about future [individual output]" (Cascio & Aguinis, 2008, p. 141). In addition, our manuscript offers contributions to future research. Our taxonomy of distributions, accompanied by a software package we make available free of charge, will facilitate distribution pitting research on other variables in domains concerned with the distribution of events. For example, our taxonomy and methodological procedures may be useful for studying distributions of organizational citizenship behaviors and counterproductive work behaviors in the individual performance literature, accidents in the safety literature, and errors (made during error management training) in the training and development literature.

The remainder of our manuscript is structured as follows. First, we offer a four-category taxonomy subsuming a total of seven distributions. Second, we introduce distribution pitting as our methodological framework as well as decision rules used to implement distribution pitting. Third, we apply distribution pitting on a data set consisting of 229 samples of individual output and including approximately 625,000 individuals across a broad range of occupations, types of individual output measures, types of collectives, and time frames. Finally, we discuss implications of our taxonomy and distribution pitting results for theory, future research, and practice.

## Taxonomy of Individual Output Distributions

We introduce a taxonomy of distributions that is new to the organizational science literature and use it specifically in the domain of individual output. The taxonomy consists of a total of seven distributions, which are grouped into four categories of distributions: (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, Poisson, and Weibull; Clauset, Shalizi, & Newman, 2009). The two distributions in the exponential tail category share several key characteristics and, therefore, are grouped together. The same reasoning applies to the three distributions in the symmetric or potentially symmetric category. Figure 1 includes generic graphic depictions for each of the distributions in our taxonomy.

Before we describe the distributions, we highlight four characteristics of our taxonomy that are crucial for theory development

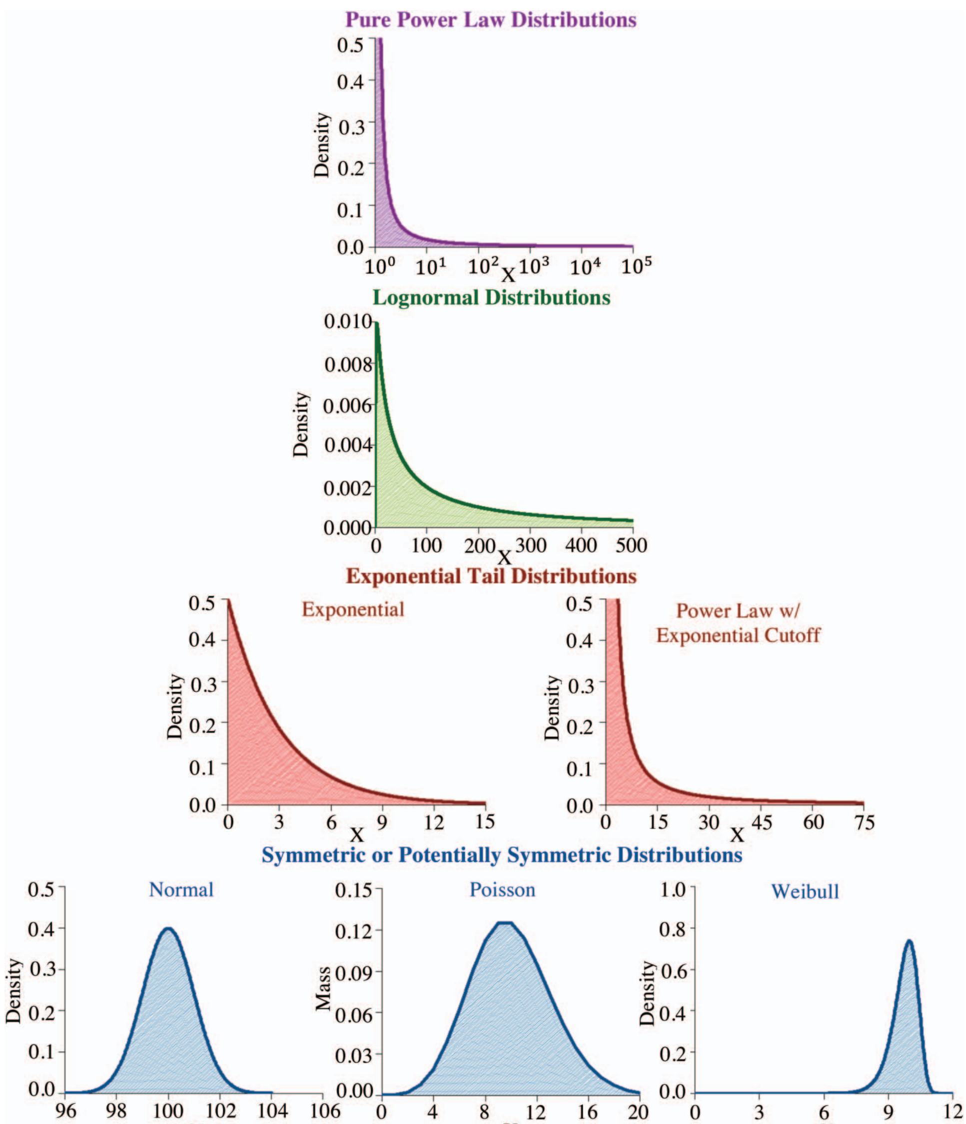


*Figure 1.* Generic visual representation of seven types of distributions (see Table 1 for more detailed descriptions of each). Pure power law ($\alpha = 1.5$); lognormal ($\mu = 5$, $\sigma = 2$); exponential tail distributions: exponential ($\lambda = 0.5$), power law with an exponential cutoff ($\alpha = 1.5$, $\lambda = 0.01$); symmetric or potentially symmetric distributions: normal ($\mu = 100$, $\sigma = 1$), Poisson ($\mu = 10$), and Weibull ($\beta = 20$, $\lambda = 10$). In each of the panels, except the one containing the Poisson distribution, the *x*-axis represents values of a continuous variable, whereas the *y*-axis ("Density") represents the likelihood of the continuous variable taking on a given value or range of values. In the panel containing the Poisson distribution, the *x*-axis represents values of a discrete variable, whereas the *y*-axis ("Mass") represents the likelihood of the discrete variable taking on a given discrete value. Each of the non-Poisson distributions shown above can model both real and discrete variables, but the Poisson distribution can only model discrete variables.

and testing. First, we focused on these seven particular distributions because one or more of the seven distributions likely explain the majority of natural phenomena (Gupta, & Kundu, 1999; Limpert, Stahel, & Abbt, 2001; Sornette, 2006). For instance, past studies have used one or more of the seven distributions to generate and test theory across a wide variety of scientific fields including physics (Bak, 1996), geology (Kile & Eberl, 2003), entomology (Lenoir, Hefetz, Simon, & Soroker, 2001), and network science (Amaral, Scala, Barthélémy, & Stanley, 2000). Our focus on these seven distributions is also consistent with how organizational science has benefitted greatly from borrowing theory and methods from other disciplines including the natural sciences (Whetten, Felin, & King, 2009). While other types of distributions have been used in past research, these distributions are typically narrower in their application. For example, the gamma distribution has been used in several fields including marketing (Platzer & Reutterer, 2016) and medicine (Yamada et al., 2016). However, the Weibull distribution can take on a wider variety of shapes, including the shapes of the gamma distribution (Balasooriya & Abeysinghe, 1994). Indeed, studies have generally favored the use of the Weibull distribution more than the gamma distribution (Gupta & Kundu, 2001).

Second, each distribution category serves as an indicator of a particular generative mechanism—a process leading to the existence of the focal distributional shape for the phenomenon under investigation. For example, the pure power law distribution and its associated generative mechanism, self-organized criticality, have been used to explain sand avalanche sizes (Bak, 1996). The lognormal distribution and its generative mechanism, proportional differentiation, have been used to explain crystal sizes (Kile & Eberl, 2003). The exponential tail distributions and their generative mechanism, incremental differentiation, have been used to explain accumulated wages (Nirei & Souma, 2007) and number of links per vertex (Amaral et al., 2000). Finally, studies have used (potentially) symmetric distributions and their generative mechanism, homogenization, to explain scent-related observations in ants (Lenoir et al., 2001) and number of specific particles (i.e., gametocytes) taken by a mosquito in its bloodmeal (Pichon, Awono-Ambene, & Robert, 2000). To enhance the relevance of our work, we also link each generative mechanism to organizational science phenomena later in the manuscript.

Third, as another important characteristic of our taxonomy related to theory development and testing, the four generative mechanisms associated with the distributions in our taxonomy are mutually exclusive. Applied to the specific domain of individual output, pure power law distribution's generative mechanism, self-organized criticality, suggests that individuals differ in terms of total output because a small proportion of individuals experience *output shocks* (i.e., unpredictable and extremely large output increases). In contrast, lognormal distribution's generative mechanism, proportional differentiation, suggests that a small proportion of individuals disproportionately benefit from *output loops* (i.e., increasingly larger output increases based on positive feedback between past and future output). Meanwhile, the generative mechanism of exponential tail distributions, incremental differentiation, suggests that some individuals enjoy larger *output increments* (i.e., linear increases in output) than others. The generative mechanism of (potentially) symmetric distributions, homogenization, suggests that individuals undergo *output homogenization* (i.e., reduction of differences in individual output).

Fourth, our taxonomy adopts a results-based definition of individual performance. In defining individual performance, prior research has focused on behaviors (Beck et al., 2014; Campbell, 1990), results (Bernardin & Beatty, 1984; Minbashian & Luppino, 2014), or both (Viswesvaran & Ones, 2000). To reflect our adoption of a results-based definition of individual performance, our discussion of theory, results, and implications focuses on *individual output*, defined as an individual's cumulative results over a certain period of time (e.g., weeks, months, years, and lifetime). Further, our focus is on explaining interindividual differences in output as opposed to intraindividual changes in output. Processes that generate differences among individuals (e.g., effects of relatively stable individual differences) are not necessarily the same as processes that generate differences within an individual over time (e.g., effects of short-term fluctuations in a variable or certain events; Dalal, Bhave, & Fiset, 2014; Dalal, Lam, Weiss, Welch, & Hulin, 2009; Molenaar, 2004; Sitzmann & Yeo, 2013).

Next, we describe the four categories of distributions and their associated generative mechanisms in greater detail. Then, from each generative mechanism, we derive implications for theory regarding the link between past and future individual output. In addition, for each of the four categories of distributions, we offer practical implications in terms of how to increase overall individual output and produce more top performers. As a preview and summary, and to offer more detailed information beyond Figure 1, Table 1 includes nontechnical and technical descriptions of each distribution (including relevant equations and parameters) as well as implications for theory and practice associated with the presence of each distribution.

## Pure Power Law Distribution

The pure power law distribution consists of a long head and often a very heavy (i.e., seemingly infinite) right tail. Compared to the other distributions in the taxonomy, the pure power law distribution has the heaviest right tail. Pure power law distribution's parameter alpha ($\alpha$) ($>1$) is the rate of decay, which denotes how quickly the distribution's right tail falls. So, the lower the value of $\alpha$ (closer to 1), the heavier is the distribution's right tail. For example, in a pure power law distribution where $\alpha = 1.5$ and sample size ($N$) $= 1{,}000$ (as shown in Figure 1), the top 10% of performers account for 99.5% of the total output, indicating a small proportion of extremely productive individuals. Moreover, the top performers may be very highly distinct (i.e., very high variability in the right tail). In fact, in the pure power law distribution shown in Figure 1, the first, second, third, and fourth largest values are 664.7, 51.0, 171.1, and 47.1% greater than the second, third, fourth, and fifth largest values, respectively. We note that "pure power law" is equivalent to "power law," and studies have used the two labels interchangeably (e.g., Clauset et al., 2009). Given the two labels, in this article, we choose to use pure power law to avoid confusing it with the power law with an exponential cutoff.

**Generative mechanism: Self-organized criticality.** The presence of a pure power law distribution indicates self-organized criticality as the generative mechanism (Andriani & McKelvey, 2009; Boisot & McKelvey, 2011; Newman, 2005). Self-organized criticality

Table 1
*Summary of the Taxonomy of Individual Output Distributions Shown Graphically in Figure 1*

| Generic description of distribution | Technical description of distribution | Implications for theory | Implications for practice |
|---|---|---|---|
| The **pure power law distribution** consists of a long head and often a very heavy (i.e., seemingly infinite) right tail. Compared to the other distributions in the taxonomy, the pure power law distribution has the heaviest right tail. Moreover, the top performers may be very highly distinct (i.e., very high variability in the right tail). See the "Pure Power Law Distribution" panel in Figure 1. | The pure power law distribution describes a set of values $x$ if $p(x) \propto x^{-\alpha}$ where alpha ($\alpha$) ($>1$) is the rate of decay, which denotes how quickly the distribution's right tail falls. So, the lower the value of $\alpha$ (closer to 1), the heavier is the distribution's right tail. For example, a distribution where $\alpha = 2$ has a heavier right tail compared to a distribution where $\alpha = 3$. | Individuals differ in terms of total output because a small proportion of individuals experience output shocks (i.e., unpredictable and extremely large output increases).<br>Differences in terms of past output may help predict future output among individuals who have not yet reached a critical state. On the other hand, among those in a critical state, differences in past output would fail to predict individuals who are more likely to experience output shocks in the future and to what extent. | To generate a greater proportion of top performers, one recommendation is to initially (but not permanently) provide similar amounts of training, opportunities, and other resources to a wide range of employees.<br>Another recommendation is to move resources out of those who have not produced an expected level of output within a prespecified period of time, and subsequently invest more heavily in those who have successfully produced the expected level of output. |
| The **lognormal distribution** consists of a heavy but ultimately finite right tail and often a bell-shaped head. Compared to other distributions in our taxonomy, the lognormal distribution has the second heaviest right tail (after the pure power law distribution). Moreover, the top performers tend to be highly distinct (i.e., high variability in the right tail). See the "Lognormal Distribution" panel in Figure 1. | The lognormal distribution describes a set of values $x$ if $p(x) \propto e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}$ where Euler's number $e \approx 2.718$. $\ln(x)$ is the natural log of $x$ and is normally distributed. Mu ($\mu$) ($>0$) is the mean. Sigma ($\sigma$) ($>0$) is the standard deviation. The higher the value of $\sigma$ (further from 0), the heavier is the right tail of the distribution. Unlike $\sigma$, $\mu$ does not affect the heaviness of the distribution's right tail. | Individuals differ in terms of total output because a small proportion of individuals disproportionately benefit from output loops (i.e., increasingly larger output increases based on positive feedback between past and future output).<br>Arbitrarily or randomly accumulated initial outputs for some individuals may be so large that other individuals with superior output accumulation rates are unable to catch up over time. | To develop and retain top performers, one recommendation is to allocate different amounts of resources across individuals based on their output accumulation rates and also past output.<br>Another recommendation is to maintain large differentiation in resource allocation not only between top and ordinary performers, but also among top performers due to their large output differences in a lognormal distribution. |

Table 1 (*continued*)

| Generic description of distribution | Technical description of distribution | Implications for theory | Implications for practice |
|---|---|---|---|
| The **exponential distribution** consists of a long head and a somewhat heavy right tail. Moreover, the top performers tend to be similar (i.e., low variability in the right tail). See the "Exponential" panel in Figure 1. | The exponential distribution describes a set of values $x$ if $$p(x) \propto e^{-\lambda x}$$ where Euler's number $e \approx 2.718$. Lambda ($\lambda$) ($>0$) is the rate of decay, which denotes how quickly the distribution's right tail falls. So, the lower the value of $\lambda$ (closer to 0), the heavier is the distribution's right tail. | Individuals differ in terms of total output because some individuals, compared to others, enjoy larger output increments (i.e., linear increases in output). That is, individuals differ in terms of output accumulation rate (i.e., output generated per opportunity to produce), which tends to have linear effects on their individual output levels rather than multiplicative and convex effects. | To generate greater overall output, one recommendation is to heavily invest in individuals with higher output accumulation rates than others. In other words, incremental differentiation suggests that it is better to allocate resources variably across individuals rather than similarly. |
| The **power law with an exponential cutoff** consists of a long head and an initially heavy but then increasingly falling right tail. Moreover, the top performers can be similar (i.e., low variability in the right tail), highly distinct (i.e., high variability in the right tail), or something in-between. Depending on its parameter values, the distribution's right tail may range from being as heavy as that of a lognormal distribution to being even lighter than that of an exponential distribution. See the "Power Law with an Exponential Cutoff" panel in Figure 1. | The power law with an exponential cutoff describes a set of values $x$ if $$p(x) \propto x^{-\alpha} e^{-\lambda x}$$ where Euler's number $e \approx 2.718$. Both alpha ($\alpha$) ($>1$) and lambda ($\lambda$) ($>0$) are rates of decay, which denote how quickly the distribution's right tail falls. So, the lower the values of $\alpha$ (closer to 1) and $\lambda$ (closer to 0), the heavier is the distribution's right tail. Between the two rates of decay, $\lambda$ is stronger in terms of making the distribution's right tail fall and thus increasingly dominates the distribution's shape along higher values of $x$. | Incremental differentiation offers greater theoretical parsimony by simply describing each individual in terms of a distinct output accumulation rate. Moreover, because incremental differentiation states that an outcome's future value is a function of accumulation rate but not initial value, incremental differentiation departs from prior research suggesting that past individual output influences future individual output through a positive feedback mechanism. | An organization may benefit from ensuring relatively small compensation differences among top performers—as indicated by the often, but not always, low variability in an exponential tail distribution's right tail. Nonetheless, exponential tail distributions of individual output still allow for large compensation differences between top and ordinary performers. |

(*table continues*)

Table 1 (*continued*)

| Generic description of distribution | Technical description of distribution | Implications for theory | Implications for practice |
|---|---|---|---|
| The **normal distribution** consists of a bell-shaped body and symmetric tails that quickly become light along values further from the mean. Only the normal distribution never has a skew different from zero (or near zero) in our taxonomy. See the "Normal" panel in Figure 1. | The normal distribution describes a set of values $x$ if $$p(x) \propto e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$ where Euler's number $e \approx 2.718$. Mu ($\mu$) (>0) is the mean. Sigma ($\sigma$) (>0) is the standard deviation. $\mu$ does not affect the heaviness of the symmetric tails. In contrast, $\sigma$ does. The lower the value of $\sigma$ (closer to 0), the lighter are the distribution's symmetric tails. | Homogenization suggests that individuals undergo output homogenization (i.e., reduction of differences in individual output). The reason is that individuals are subject to situational constraints that act as a floor and ceiling to future output differences.<br>Variability in past individual output would be followed by lower variability in future individual output. | To increase overall output, one recommendation is to implement human resource management practices (e.g., selection, training) that lead to newcomers and incumbents with higher and more uniform levels of individual output. This would subsequently make it more cost-efficient to make compensation, job assignment, and other organizational decisions.<br>An organization may benefit from focusing on managing collective processes such as the perception that reporting mistakes and failures are valuable learning opportunities. This perception would foster group learning, reduce variability in the number of errors made by individuals, and decrease the total number of errors made by the group. |
| The **Poisson distribution** has a bell-shaped head and possibly a somewhat heavy right tail consisting of relatively low counts. Unlike other distributions in the taxonomy, the Poisson distribution can model only discrete values (i.e., counts) and, thus, always has a jagged curve. See the "Poisson" panel in Figure 1. | The Poisson distribution describes a set of values $x$ if $$p(x) \propto \mu^x / x!$$ where mu ($\mu$) (>0) is the mean, which also equals the variance of the distribution. The lower the value of $\mu$ (i.e., closer to 0), the heavier is the distribution's right tail. | | |
| The **Weibull distribution** often consists of a bell-shaped head and a slight left skew. However, the distribution is so flexible that it can take on other shapes. Depending on other $\beta$ values, the distribution's right tail may also range from being (very) heavy (i.e., $\beta < 1$) to as light as a normal distribution's right tail (e.g., $\beta = 3.5$). See the "Weibull" panel in Figure 1. | The Weibull distribution describes a set of values $x$ if $$p(x) \propto (x/\lambda)^{\beta-1} e^{-(x/\lambda)^{\beta}}$$ where Euler's number $e \approx 2.718$. Lambda ($\lambda$) (>0) is the extent to which the distribution is pushed down and stretched to the sides. So, the higher the value of $\lambda$ (further from 0), the lower is the distribution's height. Beta ($\beta$) (>0) is the extent to which the distribution's head is pushed to the right. So, the higher the value of $\beta$ (further from 0), the heavier is the distribution's left tail (e.g., $\beta = 20$). | | |

states that individuals differ in terms of total value on an outcome because, after some individuals reach critical states, seemingly trivial events trigger increases on their outcome ranging from small to extremely large. In self-organized criticality, a critical state refers to a situation where components accumulated by an individual interconnect. So, even a small event initially affecting a single component may in turn impact other components that are directly or indirectly connected to the initially affected component. Moreover, given that self-organized criticality does not require a manager to determine the specific individuals who reach critical states, the individuals are said to have self-organized to their critical states—hence the label, self-organized criticality.

As an organizational science example, a scientist may be in a critical state when he is working on a particular set of projects (i.e., accumulated components) such that the progress of one project depends on the progress made in other projects (i.e., such that the components are interconnected). So, a small event in the form of a single breakthrough on one project may in turn lead to a number of breakthroughs in other projects. Depending on how intricately the scientist's projects were interconnected as well as the number of those projects, the initial breakthrough may quickly lead to a very large number of new publications. Consistent with this example, prior research has noted that star scientists often take advantage of the "cross-talk" among various projects to derive unexpected solutions and insights (Simonton, 2003, p. 479; Poincaré, 1921). As another organizational science example, a manager may be in a critical state because she has proposed a number of initiatives (i.e., accumulated components) that complement one another (i.e., that are interconnected). So, top-management approval on one initiative may in turn lead to a number of her other initiatives being approved. Depending on how well the manager's proposed initiatives complement one another as well as the impact of those initiatives on the company, the first top-management approval might even develop into a company-wide transformation led by the manager.

As an illustration from outside of organizational science, a forest is in a critical state when multiple clusters of trees (i.e., accumulated components) are adjacent to one another (i.e., are interconnected). In this situation, a small event in the form of a single lightning strike on one tree cluster may start a fire that spreads to other tree clusters that are directly or indirectly connected to the tree cluster hit by the lightning strike. The exact amount of the forest destroyed by the fire would depend on how intricately the tree clusters were connected (Newman, 2005). As another illustration from outside of organizational science, research in physics has found that once enough sand grains (i.e., accumulated components) have piled up to reach a "critical slope" (i.e., have interconnected with each other), the drop of another sand grain on the pile will cause a sand avalanche—with a magnitude that is potentially extremely large. This process, repeated over many times, will generate a pure power law distribution of sand avalanche sizes (Bak, 1996).

**Theoretical implications of self-organized criticality.** Applied to the domain of individual output, self-organized criticality suggests that individuals differ in terms of total output because a small proportion of individuals experience *output shocks* (i.e., unpredictable and extremely large output increases). Specifically, after some individuals reach critical states, seemingly trivial events trigger increases in their output ranging from small to extremely large. In addition, increases in individual output after reaching critical states would be unpredictable, or nondeterministic (Bak, 1996; Boisot & McKelvey, 2011; Sornette & Ouillon, 2012). This is because it generally takes a long time to reach a critical state, and it is often never reached in an individual's lifetime. For instance, a scientist in search of the cure for a particular cancer may persistently engage in work behaviors (e.g., conducting experiments), but nonetheless there is a very large amount of uncertainty as to whether those behaviors will lead to the desired final result (e.g., finding the cure in her lifetime). Similarly, in many contexts involving artists, writers, inventors, entrepreneurs, and managers, individuals may doggedly pursue a difficult outcome to achieve the "grand vindication [that nevertheless] may never come" (Taleb, 2007, p. 87). The longer it takes for individuals to reach critical states, the more rapidly compounding is the degrading effect of measurement errors on prediction accuracy. Thus, it becomes necessary to measure the past with virtually perfect precision to predict distal outcomes such as output shocks, especially in complex settings involving human interactions (Taleb, 2007).

Consistent with the notion of output shocks, research in the individual performance literature has found that behaviors and results are often decoupled in a number of occupations such as sales (Aguinis, 2013; Bommer, Johnson, Rich, Podsakoff, & MacKenzie, 1995). Yet, going beyond the broad observation that behaviors and results are often decoupled, self-organized criticality more precisely suggests that the decoupling is much more severe once individuals reach critical states, after which some individuals would experience output shocks that are unpredictable. Further, self-organized criticality implies that differences in terms of past output may help predict future output among individuals who have not yet reached a critical state. On the other hand, among those in a critical state, differences in past output would fail to predict individuals who are more likely to experience output shocks in the future and to what extent. For instance, a series of experiments found that songs considered top musical quality (the "best" songs) were the most unpredictable in terms of future market share (Salganik, Dodds, & Watts, 2006).

**Practical implications of self-organized criticality.** To generate a greater proportion of top performers, one recommendation is to initially (but not permanently) provide similar amounts of training, opportunities, and other resources to a wide range of employees. This is consistent with self-organized criticality's emphasis on the unpredictability of output shocks, as well as the fact that there is no research to guide practice on how to identify and predict when certain individuals reach critical states (after which they would experience output shocks). Allocating similar amounts of resources widely across individuals is similar to, but not as precise as, using a real options-based 1/N (i.e., blind funding) approach, where a decision-maker (e.g., venture capitalist) with N investment possibilities "invest[s] in all of them in equal amounts" (Taleb, 2012, p. 230). A 1/N approach to investing or, more generally, investing similarly across many individuals can be useful if, among many uncertain investment decisions, only few investments would yield nearly unlimited payoff while other investments yield little or no payoff.

Another recommendation is to move resources out of those who have not produced an expected level of output within a prespecified period of time, and subsequently invest more heavily in those who have successfully produced the expected level of output (Berk & Kaše, 2010). This would at least better incentivize individuals to engage in various behaviors that could be (though not necessarily) conducive to reaching a critical state, after which some individuals would experience output shocks. Without incentivizing individuals to continuously strive toward a critical state, it seems less likely that an organization will benefit from output shocks.

## Lognormal Distribution

The lognormal distribution consists of a heavy but ultimately finite right tail and often a bell-shaped head. More precisely, a distribution follows a lognormal distribution if its logarithm results in a normal distribution. Though lognormal distributions may appear quite similar to pure power law distributions in that both types of distributions model the presence of extreme values, the two are different because lognormal distributions "fall" (i.e., decay) rapidly at the highest values of observations (Taleb, 2007, p. 326). Thus, compared to other distributions in our taxonomy, the lognormal distribution has the second heaviest right tail (after the pure power law distribution). Lognormal distribution's parameter mu ($\mu$) ($>0$) refers to the mean, which does not affect the heaviness of the distribution's right tail. In contrast, the higher the value of the distribution's standard deviation, or sigma ($\sigma$) ($>0$), the heavier is the right tail. For instance, in a lognormal distribution where $\mu = 5$, $\sigma = 2$, and $N = 1{,}000$ (as shown in Figure 1), the top 10% of performers account for 74.4% of the total output, indicating a small group of disproportionately productive individuals. Moreover, the top performers tend to be highly distinct (i.e., high variability in the right tail). In fact, in the lognormal distribution shown in Figure 1, the first, second, third, and fourth largest values are 20.2, 45.3, 63.5, and 6.7% greater than the second, third, fourth, and fifth largest values, respectively.

**Generative mechanism: Proportionate differentiation.** The presence of a lognormal distribution indicates proportionate differentiation as the generative mechanism (Banerjee & Yakovenko, 2010; Gibrat, 1931; Mitzenmacher, 2004). There are two key components in a proportionate differentiation process: initial value and accumulation rate. Initial value refers to the amount of a variable that each individual has accumulated during a relatively short period of time (e.g., 1 year) since the beginning of a common baseline (e.g., since the first date of employment for all individuals hired in the same year). Accumulation rate is the average amount of the variable that an individual produces per time period (e.g., sales generated per month). Given such, proportionate differentiation states that individuals differ in terms of total value on an outcome because of their differences with respect to the accumulation rate and initial value on the outcome. Further, accumulation rate and initial value on an outcome would interact (i.e., multiply with each other). So, future amounts of the outcome would increase by increasing amounts for some individuals, whereas future amounts of the same outcome would stay at low levels for many other individuals. In short, an outcome's future value is a distinct proportion (i.e., percentage) of the outcome's initial value—hence the label, proportionate differentiation.

As an organizational science example, an individual output dimension among community organizers is the number of signatures collected. Some organizers may initially obtain a larger number of signatures (i.e., higher initial value) than others and, as a result, find it easier to obtain additional signatures because people might be more willing to endorse something that other people already have. If the same organizers also tend to more quickly obtain signatures (i.e., higher accumulation rate) than others, then these organizers may accumulate additional signatures by increasingly larger amounts, thus benefiting greatly from positive feedback loops between the number of signatures obtained so far and additional signatures obtained. On the other hand, many other organizers may initially fail to obtain many signatures and also obtain signatures more slowly. As a result, these organizers may find it difficult over time to obtain additional signatures. This is analogous to how two employees with different starting salaries may further depart from each other in terms of future salary. For instance, suppose Jessie and Sam begin their tenure on the same job. If Jessie receives even a slightly higher starting annual salary (i.e., initial value) compared to Sam, then Sam will not be able to catch up to Jessie's annual salary—that is, unless Sam enjoys a higher rate of annual raises on her salary (i.e., accumulation rate) large enough to eventually offset the initial differences. Similarly, simulation research has shown that a seemingly trivial initial difference between two groups in terms of performance ratings may later lead to large differences in the two groups' promotion rates (Martell, Lane, & Emrich, 1996).

As another organizational science example, if a firm already has a large amount of a resource (e.g., R&D know-how in a certain area), the firm is in a better situation to accumulate more of the same resource compared to other firms with low amounts of the resource (Dierickx & Cool, 1989). In turn, because firms differ not only on the initial amount of the resource but also on the accumulation rate, future amounts of the resource would increase by increasing amounts for some firms, thereby creating a heavy right tail in the resulting distribution. In contrast, for many other firms, future amounts of the resource would stay at low levels, possible creating a bell-shaped head in the resulting distribution (Gabaix, 1999).

Across these examples of proportionate differentiation, each observation (e.g., individual, firm) would accumulate a certain positive amount of initial value because of the observation's accumulation rate and/or luck on the focal outcome (Barabási, 2012). Regarding time of measurement, initial value can be measured as soon as the opportunity to accumulate initial value is available (e.g., first year out of graduate school for researchers, since the foundation of a firm).

Finally, unlike in self-organized criticality (i.e., pure power law distribution's generative mechanism), it is not necessary in a proportionate differentiation process for individuals to have reached critical states for them to have large differences with respect to an outcome. Instead, in proportionate differentiation, outcome values are a function of the product between accumulation rate and initial value on the outcome. Because of this distinction between self-organized criticality and proportionate differentiation, the highest and most extreme values in a pure power law distribution are unpredictable, whereas those in a lognormal distribution can be predicted as long as the accumulation rate and initial value are known.

**Theoretical implications of proportionate differentiation.** Applied to the domain of individual output, proportionate differentiation suggests that individuals differ in terms of total output because a small proportion of individuals disproportionately benefit from *output loops* (i.e., increasingly larger output increases based on positive feedback between past and future output). Specifically, individuals differ in terms of output accumulation rate (i.e., output generated per opportunity to produce) and initial output (i.e., output accumulated so far in the beginning of an individual's career or tenure at an organization). In turn, the product between each individual's output accumulation rate and initial output means that many individuals would only reach low levels of future output, whereas some individuals would accumulate future output by increasingly larger amounts.

Consistent with the notion of output loops, research in the individual performance literature has noted that for some occupations or individual output dimensions, additional output requires fewer resources (e.g., time, effort) than does initial output (Aguinis et al., 2016). That is, the marginal cost of output decreases as more output is produced, and this helps explain why future individual output would increase proportionately based on initial individual output. Yet, going beyond the notion that additional individual output may require fewer resources than initial individual output, proportionate differentiation allows for the possibility that even if a person has exceptionally high initial output because of luck, the person's total output may eventually be surpassed by another individual with a superior output accumulation rate. That is, arbitrary or random differences in initial individual output may not lead to long-term differences in individual output (Mankiw, 2013; van de Rijt, Kang, Restivo, & Patil, 2014). On the other hand, proportionate differentiation also allows for the possibility that arbitrarily or randomly accumulated initial outputs for some individuals may be so large that other individuals with superior output accumulation rates are unable to catch up over time. Going back to the example involving community organizers, those with higher accumulation rates are more likely to initially obtain signatures with which to obtain additional signatures—though the initial number of signatures obtained, at least to some extent, may be attributed to "dumb luck" (Barabási, 2012, p. 507).

**Practical implications of proportionate differentiation.** To develop and retain top performers, one recommendation is to allocate different amounts of resources across individuals based on their output accumulation rates and also past output. The reason is that, in proportionate differentiation, an individual's future output levels are a function of the product between her output accumulation rate and initial output. Another recommendation is to maintain large differentiation in resource allocation not only between top and ordinary performers, but also among top performers because of their large output differences in a lognormal distribution.

## Exponential Tail Distributions

Exponential tail distributions have positively skewed tails that fall rapidly (i.e., decay at an exponential rate). These distributions include the exponential distribution and power law with an exponential cutoff. The exponential distribution consists of a long head and a somewhat heavy right tail. Exponential distribution's parameter lambda ($\lambda$) ($>0$) is the rate of decay, which denotes how quickly the distribution's right tail falls. So, the lower the value of $\lambda$ (closer to 0), the heavier is the distribution's right tail. For instance, in an exponential distribution where $\lambda = 0.5$ and $N = 1{,}000$ (as shown in Figure 1), the top 10% of performers account for 32.8% of the total output. Moreover, the top performers tend to be similar (i.e., low variability in the right tail). In fact, in the exponential distribution shown in Figure 1, the first, second, third, and fourth largest values are 18.7, 7.4, 4.3, and 4.4% greater than the second, third, fourth, and fifth largest values, respectively.

The power law with an exponential cutoff consists of a long head and an initially heavy but then increasingly falling right tail. Specifically, in a power law with an exponential cutoff, parameters alpha ($\alpha$) ($>1$) and lambda ($\lambda$) ($>0$) are both rates of decay, which denote how quickly the distribution's right tail falls. So, the lower the values of $\alpha$ (closer to 1) and $\lambda$ (closer to 0), the heavier is the distribution's right tail. Between the two rates of decay, the exponential rate of decay $\lambda$ is stronger in terms of making the distribution's right tail fall and thus increasingly influences the distribution's shape along higher values of $x$. For example, in power laws with exponential cutoffs where $\alpha = 1.5$ and $N = 1{,}000$, the top 10% of performers account for 64.1% or 91.8% of the total output depending on the value of $\lambda$ (i.e., 0.5 vs. 0.01, respectively). Moreover, the top performers can be similar (i.e., low variability in the right tail), highly distinct (i.e., high variability in the right tail), or something in-between. For instance, in a power law with an exponential cutoff where $\alpha = 1.5$, $\lambda = 0.5$, and $N = 1{,}000$, the first, second, third, and fourth largest values are 66.8, 2.7, 29.7, and 2.9% greater than the second, third, fourth, and fifth largest values, respectively. However, in another power law with an exponential cutoff where $\alpha = 1.5$, $\lambda = 0.01$, and $N = 1{,}000$ (as shown in Figure 1), the top performers are more distinct from one another. That is, in the latter power law with an exponential cutoff, the first, second, third, and fourth largest values are 53.9, 6.9, 35.6, and 68.5% greater than the second, third, fourth, and fifth largest values, respectively. Finally, depending on the distribution's parameter values, its right tail may range from being as heavy as that of a lognormal distribution to being even lighter than that of an exponential distribution. As an example, in a power law with an exponential cutoff where $\alpha = 1.5$, $\lambda = 20$, and $N = 1{,}000$, the top 10% of performers account for 29.6% of the total output. This percentage is lower than in an exponential distribution with $\lambda = 0.5$ and $N = 1{,}000$ (as shown in Figure 1), where the top 10% of performers account for 32.8% of the total output.

**Generative mechanism: Incremental differentiation.** The presence of an exponential tail distribution indicates incremental differentiation as the generative mechanism (Amitrano, 2012; Czirók, Schlett, Madarász, & Vicsek, 1998). Incremental differentiation states that individuals differ in terms of total value on an outcome because of their differences with respect to the accumulation rate on the outcome. In incremental differentiation, accumulation rate refers to the average amount of the variable that an individual produces per time period (e.g., sales generated per month). To be clear, incremental differentiation is distinct from proportionate differentiation (lognormal distribution's generative mechanism) in two ways. First, incremental differentiation's accumulation rate is slightly different from proportionate differentiation's accumulation rate. Though accumulation rate has a linear (i.e., additive) effect on the focal outcome in both generative mechanisms, incremental differentiation clearly acknowledges that individuals with the highest accumulation rates may be subject to diminishing returns. Second, both accumulation rate and initial value explain individuals' total values on an outcome in proportionate differentiation, whereas incremental differentiation specifies that an outcome's future value is a function of accumulation rate but not initial value. In short, individuals' values on an outcome would differentiate at an incremental rate and, hence, the generative mechanism's label.

As an organizational science example of incremental differentiation leading to an exponential distribution, differences across individuals in terms of labor productivity lead them to accumulate wages at different linear rates. The result is an exponential distribution of accumulated wages (Nirei & Souma, 2007). As an organizational science example of incremental differentiation leading to a power law with an exponential cutoff, consider medical school doctors who have three primary responsibilities: Patient care (i.e., clinical work), research, and teaching (Miller, 2000). Compared to other doctors who began their tenure at the school around the same time, some doctors may have disproportionately greater amounts of cumulative patient care output (e.g., number of successful treatments) because of their higher accumulation rates regarding patient care (e.g., treatment success rates). However, these doctors with superior accumulation rates in terms of patient care would face steep costs if they try to see more patients or spend more time per patient after reaching their full capacity (e.g., using up all the extra time and energy freed up by research and/or teaching load reductions). Such steep costs might include less time for family, hobbies, and even basic life function including sleep. Thus, medical school doctors who accumulate patient care output more quickly than others would nonetheless be subject to diminishing returns, possibly giving rise to a distribution of cumulative patient care output that follows a power law with an exponential cutoff.

As an illustration of incremental differentiation from outside of organizational science, research in network science has documented that vertices (e.g., airports) incur various costs when they make additional links with other vertices (e.g., new departures to other airports). As an airport reaches full capacity, adding new arrivals and departures requires additional construction, extra staff, and other capacity-expanding activities. Such requirements for expanding the airport's capacity would rapidly increase its cost of adding a new arrival/departure. Thus, vertices that accumulate links more quickly than other vertices would nonetheless be subject to diminishing returns, giving rise to a distribution of links per vertex that follows a power law with an exponential cutoff (Amaral et al., 2000).[1]

**Theoretical implications of incremental differentiation.** Applied to the domain of individual output, incremental differentiation suggests that individuals differ in terms of total output because some individuals, compared to others, enjoy larger *output increments* (i.e., linear increases in output). That is, individuals differ in terms of output accumulation rate (i.e., output generated per opportunity to produce), which tends to have linear effects on their individual output levels rather than multiplicative and convex effects. Moreover, individuals with the highest output accumulation rates may be subject to diminishing returns. In particular, high output accumulation rates may exhibit stronger diminishing returns on individual output in lower-complexity jobs than in higher-complexity jobs. In this sense, high output accumulation rates might be less valuable for accumulating individual output in lower-complexity jobs, where there are less likely to be positively skewed distributions of total individual output (Aguinis et al., 2016; Vancouver et al., 2016).

Consistent with the notion of output increments, research in the individual performance literature has noted that individual differences such as cognitive ability have a linear relation with individual performance (Whetzel, McDaniel, Yost, & Kim, 2010). In addition, high values on other individual differences such as conscientiousness are often subject to diminishing returns in terms of their effects on individual performance (Pierce & Aguinis, 2013; Sackett, Gruys, & Ellingson, 1998). Yet, going beyond the observation that various individual differences have linear effects on outcomes with possibly diminishing returns, incremental differentiation offers greater theoretical parsimony by simply describing each individual in terms of a distinct output accumulation rate. Furthermore, because incremental differentiation states that an outcome's future value is a function of accumulation rate but not initial value, incremental differentiation departs from prior research suggesting that past individual output influences future individual output through a positive feedback mechanism (e.g., Aguinis et al., 2016; Vancouver et al., 2016).

**Practical implications of incremental differentiation.** To generate greater overall output, one recommendation is to heavily invest in individuals with higher output accumulation rates than others. In other words, incremental differentiation suggests that it is better to allocate resources variably across individuals rather than similarly. The reason is that past output in terms of an individual's output accumulation rate determines future output. Moreover, an organization may benefit from ensuring relatively small compensation differences among top performers—as indicated by the often, but not always, low variability in an exponential tail distribution's right tail. Nonetheless, exponential tail distributions of individual output still allow for large compensation differences between top and ordinary performers.

## Symmetric or Potentially Symmetric Distributions

Distributions in this category have symmetric tails or often have nearly symmetric tails. These distributions include the normal, Poisson, and Weibull distributions. The normal distribution consists of a bell-shaped body and symmetric tails that quickly become light along values further from the mean. Normal distribution's parameter mu ($\mu$) ($>0$) refers to the mean, which does not affect the heaviness of the symmetric tails. In contrast, the lower the value of the distribution's standard deviation, or sigma ($\sigma$) ($>0$), the lighter are the distribution's symmetric tails. For instance, in a normal distribution where $\mu = 100$, $\sigma = 1$, and $N = 1{,}000$ (as shown in Figure 1), the top 10% of performers account for only 10.3% of the total output. Only the normal distribution never has a skew different from zero (or near zero) in our taxonomy.

The Poisson distribution has a bell-shaped head and possibly a somewhat heavy right tail consisting of relatively low counts. Poisson distribution's parameter mu ($\mu$) ($>0$) is the mean, which also equals the variance. The lower the value of $\mu$ (i.e., closer to 0), the heavier is the distribution's right tail. As an example, in a Poisson distribution where $\mu = 100$ and $N = 1{,}000$, the top 10%

[1] Incremental differentiation does not require decreases in each individual's output accumulation rate over time. Instead, the increasing cost per additional output at the highest levels of output accumulated results in diminishing returns and, therefore, a power law with an exponential cutoff.

of performers account for 11.9% of the total output. However, in another Poisson distribution where $\mu = 10$ and $N = 1{,}000$ (as shown in Figure 1), the top 10% of performers account for 15.8% of the total output. Unlike other distributions in the taxonomy, the Poisson distribution can model only discrete values (i.e., counts) and, thus, always has a jagged curve.

The Weibull distribution often consists of a bell-shaped head and a slight left skew. Specifically, one parameter of the Weibull distribution, lambda ($\lambda$) ($>0$), is the extent to which the distribution is pushed down and stretched to the sides. So, the higher the value of $\lambda$ (further from 0), the lower is the distribution's height. Another parameter of the Weibull distribution, beta ($\beta$) ($>0$), is the extent to which the distribution's head is pushed to the right. So, the higher the value of $\beta$ (further from 0), the heavier is the distribution's left tail. For example, in a discrete Weibull distribution where $\beta = 20$, $\lambda = 10$, and $N = 1{,}000$ (as shown in Figure 1), the median and mean were 10 and 9.73, respectively, indicating a slight left skew. However, the distribution is so flexible that it can take on other shapes. Depending on other $\beta$ values, the distribution's right tail may also range from being (very) heavy (i.e., $\beta < 1$) to as light as a normal distribution's right tail (e.g., $\beta = 3.5$).

**Generative mechanism: Homogenization.** The presence of symmetric or potentially symmetric distributions indicates homogenization as the generative mechanism (Araujo & Herrmann, 2010; Seminogov, Semchishen, Panchenko, Seiler, & Mrochen, 2002; Spear & Chown, 2008). Homogenization is a process that reduces differences among individuals in terms of their values on an outcome. Unlike other generative mechanisms we discussed previously (i.e., self-organized criticality, proportionate differentiation, and incremental differentiation), which refer to how differences among individuals increase (i.e., increased heterogeneity), homogenization refers to how differences decrease over time (i.e., increased homogeneity)—hence the label, homogenization.

As an organizational science example, differences across individuals regarding characteristics such as attitudes often decrease over time through the mechanisms of attraction, selection, and attrition (Ployhart, Weekley, & Baughman, 2006; Schneider, 1987). In addition, uniform expectations of production or service reduce the differences in output among assembly line workers (e.g., Groshen, 1991) and service workers (e.g., Das, Das, & Mackenzie, 1996; Tepeci, 1999). As an illustration of homogenization from outside of organizational science, research in entomology found that mouth-to-mouth feeding, social grooming, and other physical contact (i.e., group dynamics) among ants homogenized their scent and led to (nearly) normal distributions of scent-related observations (Lenoir et al., 2001).

Even if situational constraints consist of a floor and ceiling that are *un*equally strong, the result may nonetheless be reduced differences across individuals in terms of their values on an outcome. That is, in the presence of an unequally strong floor and ceiling, random fluctuations around a mean value alone may lead to a (nearly) normal distribution, a Poisson distribution possibly characterized by a somewhat heavy right tail (Vancouver et al., 2016), or a Weibull distribution possibly characterized by a (slight) left or right skew (Newby & Winterton, 1983; Rinne, 2008). Examples of such instances include income quota by taxi drivers per day (Camerer, Babcock, Loewenstein, & Thaler, 1997) and number of arrests made by police officers per month (Meng & Burris, 2013)—where output differences among individuals are reduced despite production or service expectations that constitute an unequally strong floor and ceiling.

**Theoretical implications of homogenization.** Applied to the domain of individual output, homogenization suggests that individuals undergo *output homogenization* (i.e., reduction of differences in individual output). The reason is that individuals are subject to situational constraints that act as a floor and ceiling to future output differences. Consistent with the notion of output homogenization, research in the individual performance literature has found that individual output dimensions that are central to a job or position tend to approximate a normal distribution. For example, a normal distribution can describe the number of publications among pretenured industrial-organizational psychologists working in departments with research-oriented doctoral programs (Beck et al., 2014). Indeed, promotion policies in research-oriented departments generally lead to terminating scholars who fail to produce a certain high number of publications (i.e., floor), thereby limiting variability in the number of publications among scholars in research-oriented departments. In short, in homogenization, variability in past individual output would be followed by lower variability in future individual output.

**Practical implications of homogenization.** To increase overall output, one recommendation is to implement human resource management practices (e.g., selection, training) that lead to newcomers and incumbents with higher and more uniform levels of individual output. This would subsequently make it more cost-efficient to make compensation, job assignment, and other organizational decisions (Aguinis & O'Boyle, 2014). Moreover, an organization may benefit from focusing on managing collective processes such as the perception that reporting mistakes and failures are valuable learning opportunities. This perception would foster group learning, reduce variability in the number of errors made by individuals, and decrease the total number of errors made by the group (Leroy et al., 2012).

### Summary

Our taxonomy consists of four categories of distributions: (1) pure power law; (2) lognormal; (3) exponential tail (consisting of exponential and power law with an exponential cutoff); and (4) symmetric or potentially symmetric (consisting of normal, Poisson, and Weibull). From a theoretical perspective, these four categories of distributions are associated with the following generative mechanisms, respectively: self-organized criticality, proportionate differentiation, incremental differentiation, and homogenization. Next, we describe our empirical study aimed at understanding which distribution more closely approximates each observed distribution of individual output.

## Method

### Data

Our study had four critical data requirements. First, because most distributions in our taxonomy (can) have heavy right tails, we needed to include long time scales so that some of the observations have enough time to develop into outliers (Andriani & McKelvey, 2009). For example, in the domain of finance, researchers have noted that "assessing extreme risks . . . at [short] time scales of 1 or 5 min leads to . . . dramatic under-estimations of the amount of risk" (Malevergne, Pisarenko, & Sornette, 2005, p. 380). Accordingly, our data set includes a variety of time frames (e.g., weeks, months, years, and lifetime).

Second, we needed to ensure that our data set was not limited to only one or a few contexts because our aim was to obtain a general understanding about the shape of individual output distributions. Accordingly, our study used data collected from a broad range of occupations (i.e., researchers, entertainers, politicians, athletes, and additional occupations consisting of manufacturing, service, and clerical jobs), types of individual output measures (e.g., products produced, number of appearances, number of times won, and revenue generated), and types of collectives (e.g., an entire profession, an organization, a unit in an organization).

Third, given our focus on interindividual output, we used data that reflect output between individuals and not within individuals over time. While an understanding of intraindividual output is important, the generative mechanisms we previously discussed focus on explaining interindividual differences and not necessarily why a specific individual's output would change over time.

Fourth, our study required distributions of individual output (i.e., distribution level of analysis) rather than observations of individual output (i.e., individual level of analysis). In other words, we needed to collect a number of samples, where each sample refers to a distribution (i.e., column, input vector) of individual output. Because we required a large amount of data compared to the typical organizational science study (e.g., Shen et al., 2011), we used a data set consisting of 229 samples of individual output collected via procedures as described in detail by Aguinis et al. (2016) (we thank Aguinis, O'Boyle, Gonzalez-Mulé, and Joo for allowing us to use the data). The total number of observations across all samples is 633,876 (or approximately 625,000, adjusting for multiple counts of a small number of observations across samples). Further, each observation refers to an individual's cumulative output within a given period of time. We emphasize that although we used the same raw data as Aguinis et al. (2016), our theory, methodological approach, and analytic procedures are entirely new.

## Analysis

**Overview of distribution pitting.** Finding that a distribution fits a sample is necessary but not sufficient evidence that a specific generative mechanism is present (Stumpf & Porter, 2012). This is because other distributions, which serve as indicators of alternative generative mechanisms, may also fit the sample. So, it is important to compare distributions with one another in terms of their fit to the sample. To meet this requirement, we conducted *distribution pitting*, a novel falsification approach-based method used for comparing distributions to identify those that do not represent the data well. For implementation, we used a new R package we developed called Dpit, which is available on http://www.hermanaguinis.com or the Comprehensive R Archive Network (CRAN). To compare the pure power law distribution with the other six distributions shown in Figure 1 (i.e., 7 instances of distribution pitting), the R package uses code available at http://tuvalu.santafe.edu/~aaronc/powerlaws/. The package also includes code to compare the other six distributions with one another (i.e., 14 additional instances of distribution pitting). After loading the Dpit package and our data set of 229 samples, we entered one command line in R: out <-Dpit(samples). This command led to comparing all seven distributions with each other per sample (i.e., 21 instances of distribution pitting per sample), completing a total of 4,809 instances of distribution pitting (= 21 instances of distribution pitting ∗ 229 samples).

Because we needed to complete a large number of analyses, we incorporated two features into the Dpit package designed for facilitating its use in the future. First, Dpit incorporates "for loop" functions to create a separate row containing results for each sample until the entire data set has been analyzed. The for loop functions also automatically clean each sample by removing missing cases and nonpositive values that lead to incalculable expressions (e.g., the log of zero is undefined). Second, Dpit skips over any unsuccessful calculations and continues analyzing the remainder of the data. When calculations fail (e.g., sample size was too small), the package prints "NA" entries in the relevant cells of the results matrix before continuing with subsequent calculations. These automation features are not available in other software packages, which are also often limited to comparing the pure power law distribution with a subset of nonnormal distributions included in our taxonomy (e.g., poweRlaw package in R; Gillespie, 2014).

**Decision rules.** Because the shape of an individual output distribution may be the result of multiple mechanisms, we applied three decision rules that are ipsative (i.e., comparative) in nature to identify the likely dominant distribution per individual output sample. First, we used distribution pitting statistics generated from our Dpit package. Per sample, and for each instance of distribution pitting involving two distributions, the Dpit package provides two types of statistics: a loglikelihood ratio (LR) and its associated $p$ value. Given that our R package, or Dpit, treats one of the two focal distributions as the "first" distribution and the other as the "second" distribution, LR quantifies the degree to which the first distribution fits better than the second distribution. So, a positive LR value means that the second distribution fits worse, whereas a negative LR value means that the first distribution fits worse. The $p$ value of each LR value indicates the extent to which random fluctuations alone likely explain the presence of a nonzero LR value, such that $LR = 0$ constitutes the null hypothesis (Clauset et al., 2009). The lower the $p$ value, the less likely that the LR value is simply because of chance. We adopted the $p$ value cutoff of 0.10 (Clauset et al., 2009). Per individual output sample, if only one type of distribution was never identified as being the worse fitting distribution, then we concluded that the particular type of distribution was the likely dominant distribution. However, if multiple types of distributions were never identified as being the worse fitting distribution, we then applied the remaining two decision rules, as explained in the following.

In the second decision rule, we applied the principle of parsimony to samples for which the likely dominant distribution was not yet determined (i.e., samples for which multiple types of distributions were never identified as being the worse fitting distribution despite implementing the first decision rule). So, for two nested distributions, the distribution with more parameters is the worse explanation for the observed distribution at hand. Though the distribution with more parameters is guaranteed to have equivalent or superior fit to the data, this comes at the price of reduced parsimony and, therefore, increases the risk that the fitted model will be sample-specific (i.e., not generalizable). Our taxonomy contains three pairs of nested distributions: (a) power law with an exponential cutoff (two parameters) and pure power law distribution (one parameter); (b) power law with an exponential cutoff (two parameters) and exponential distribution (one parameter); and (c) Weibull distribution (two parameters) and exponential distribution (one parameter). So, for example, if the first decision rule identified neither the power law with an exponential cutoff nor the pure power law distribution as being the worse fitting distribution, we then used the second decision rule to identify the former distribution as being the worse explanation for the observed distribution. Next, we applied the third decision rule to samples for which the likely dominant distribution was still not determined (i.e., samples for which multiple types of distributions were never identified as being the worse fitting distribution even after implementing the first two decision rules).

In the third decision rule, we again applied the principle of parsimony—but this time, we focused on parsimony in terms of choosing the distribution with fewer possible distribution shapes. Specifically, a flexible distribution constitutes a broader category of distribution shapes encompassing the shapes of an inflexible distribution. With respect to our taxonomy, flexible distributions include the lognormal, Poisson, and Weibull distributions, whereas inflexible distributions include the pure power law, exponential, power law with an exponential cutoff, and normal distributions. For example, a Poisson distribution (i.e., a flexible distribution) can either have a certain amount of right skew or approximate a normal distribution with symmetric tails depending on its $\mu$ value. In contrast, a normal distribution (i.e., an inflexible distribution) will always have symmetric tails regardless of its $\mu$ and $\sigma$ values. As another example, a discrete Weibull distribution (i.e., a flexible distribution) where $\beta = 20$ and $\lambda = 10$ has a slight left skew, whereas an exponential distribution (i.e., an inflexible distribution) will always have a right skew and never a left skew regardless of its $\lambda$ value. In short, inflexible distributions are nested within flexible distributions in terms of shape. Thus, when a flexible distribution and an inflexible distribution remain after using the second decision rule, the principle of parsimony dictates that we choose the distribution with fewer possible distribution shapes (i.e., the inflexible distribution) rather than the other distribution with more possible distribution shapes (i.e., the flexible distribution). To be clear, the output produced from the Dpit package allows users to derive results after implementing the first decision rule, the first two decision rules, and all three decision rules.

**Accuracy of distribution pitting and decision rules.** We conducted a simulation study to assess the accuracy of our methodological procedures (i.e., distribution pitting and the three decision rules). Results indicated that accurate decisions were overwhelmingly more frequent compared to false positive and false negative decisions. Across both the discrete and continuous data we simulated, our procedures correctly identified the dominant distribution 91.2% of the time, while Type 1 and 2 error rates were 8.5 and 8.8%, respectively. Our methodological procedures are even more accurate when results are derived based on how well they identify the correct distribution category. Across both the discrete and continuous data we simulated, our procedures correctly identified the dominant distribution category 98.5% of the time, while the Type 1or 2 error rate was only 1.5%. Appendix A includes a detailed description of the simulation study.

Moreover, we conducted a follow-up study to check the extent to which the absence of the third decision rule reduces the accuracy of results. To do so, we used the same discrete and continuous data simulated in Appendix A. However, unlike the study described in Appendix A, the follow-up study only applied the first two decision rules to the simulated data. Results indicated that using decision rules #1 and #2 while not using the third decision rule reduces the accuracy of conclusions regarding the likely dominant distribution. Across both discrete and continuous data, using the first two decision rules while not the third decision rule only led to correctly identifying the dominant distribution 58% of the time (= [67 + 84]/260). This accuracy rate is much lower than the accuracy rate of 91.2%, which we obtained in Appendix A by using all three decision rules. Appendix B includes a detailed description of this follow-up study. Because of these reasons, we based our conclusions and implications on results derived after using all three decision rules, instead of results derived after using the first decision rule or the first two decision rules only, as described next.

## Results

Table 2 provides a summary of results showing the number of times each distribution was identified as the likely dominant distribution. Further, for each of the 229 samples, Table 3 shows the likely dominant distribution or whether the likely dominant distribution was undetermined—after implementing the first decision rule, the first two decision rules, and all three decision rules. Also, as an illustration, Table 4 shows detailed distribution pitting statistics for the first two and last two samples in our data set. A complete table including detailed results for each of the 229 samples is available as supplemental material at http://dx.doi.org/10.1037/apl0000214.supp.

As shown in Table 2, the power law with an exponential cutoff was the likely dominant distribution for 49.34% of our samples (113 out of 229), and the exponential distribution was the likely dominant distribution for 25.33% of our samples (58 out of 229). So, for 75% of the samples (i.e., 171 out of 229), the likely

Table 2
*Summary of Distribution Pitting Results: Number of Times Each Distribution was Identified as the Likely Dominant Distribution*

| Distribution | Generative mechanism | Count | Percentage |
|---|---|---|---|
| Pure power law | Self-organized criticality | 5 | 2.18% |
| Lognormal | Proportionate differentiation | 13 | 5.68% |
| Exponential | Incremental differentiation | 58 | 25.33% |
| Power law with an exponential cutoff | Incremental differentiation | 113 | 49.34% |
| Normal | Homogenization | 11 | 4.80% |
| Poisson | Homogenization | 0 | .00% |
| Weibull | Homogenization | 8 | 3.49% |
| Undetermined | NA | 21 | 9.17% |
| | Total | 229 | 100% |

Table 3
*Distribution Pitting Results After Implementing the First, First Two, or All Three Decision Rules*

| ID | Sample | After decision rule #1 | After decision rules #1 and #2 | After all three decision rules | ID | Sample | After decision rule #1 | After decision rules #1 and #2 | After all three decision rules |
|---|---|---|---|---|---|---|---|---|---|
| | **Researchers** | | | | | | | | |
| 1 | Agriculture | Undetermined | Undetermined | PL w/ cutoff | 53 | Women studies | Undetermined | Undetermined | Undetermined |
| 2 | Agronomy | Undetermined | Undetermined | Undetermined | 54 | Zoology | Undetermined | Undetermined | PL w/ cutoff |
| 3 | Anthropology | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff | | **Entertainers** | | | |
| 4 | Astronomy | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff | 55 | AVN noms. actor | Undetermined | Undetermined | Exponential |
| 5 | Biological psych | Undetermined | Undetermined | PL w/ cutoff | 56 | AVN noms. actress | Undetermined | Undetermined | Exponential |
| 6 | Clinical psych | Undetermined | Undetermined | Undetermined | 57 | AVN noms. actor | Undetermined | Undetermined | PL w/ cutoff |
| 7 | Computer science | Undetermined | Undetermined | PL w/ cutoff | 58 | AVN noms. actress | Undetermined | Undetermined | Exponential |
| 8 | Criminology | Undetermined | Undetermined | PL w/ cutoff | 59 | AVN noms. director | Undetermined | Undetermined | Exponential |
| 9 | Demography | Undetermined | Undetermined | PL | 60 | Cable ACE noms. actress | Undetermined | Undetermined | Undetermined |
| 10 | Dentistry | Lognormal | Lognormal | Lognormal | 61 | Country Music Awards noms. | Undetermined | Undetermined | Exponential |
| 11 | Dermatology | Lognormal | Lognormal | Lognormal | 62 | Edgar Allan Poe Awards noms. | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff |
| 12 | Developmental psych | Undetermined | Undetermined | PL w/ cutoff | 63 | Emmy noms. actor | Undetermined | Undetermined | PL w/ cutoff |
| 13 | Ecology | Undetermined | Undetermined | PL w/ cutoff | 64 | Emmy noms. actress | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff |
| 14 | Economics | Undetermined | Undetermined | PL w/ cutoff | 65 | Emmy noms. art direction | Undetermined | Undetermined | PL w/ cutoff |
| 15 | Education | Undetermined | Undetermined | PL w/ cutoff | 66 | Emmy noms. casting | Undetermined | Undetermined | Exponential |
| 16 | Educational psych | Undetermined | Undetermined | Undetermined | 67 | Emmy noms. choreography | Undetermined | Undetermined | PL w/ cutoff |
| 17 | Environmental science | Undetermined | Undetermined | PL w/ cutoff | 68 | Emmy noms. cinematography | Undetermined | Undetermined | PL w/ cutoff |
| 18 | Ergonomics | Undetermined | Undetermined | PL w/ cutoff | 69 | Emmy noms. direction | Undetermined | Undetermined | PL w/ cutoff |
| 19 | Ethics | Undetermined | Undetermined | PL w/ cutoff | 70 | Emmy noms. editing | Undetermined | Undetermined | PL w/ cutoff |
| 20 | Ethnic studies | Undetermined | Undetermined | PL w/ cutoff | 71 | Emmy noms. lighting | Undetermined | Undetermined | PL w/ cutoff |
| 21 | Finance | Undetermined | Undetermined | PL w/ cutoff | 72 | Emmy noms. writing | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff |
| 22 | Forestry | Undetermined | Undetermined | Undetermined | 73 | Golden Globe noms. actor | Undetermined | Undetermined | PL w/ cutoff |
| 23 | Genetics | Lognormal | Lognormal | Lognormal | 74 | Golden Globe noms. actress | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff |
| 24 | History | Undetermined | Undetermined | Exponential | 75 | Golden Globe noms. direction | Exponential | Exponential | Exponential |
| 25 | Hospitality | Undetermined | Undetermined | PL w/ cutoff | 76 | Golden Globe noms. TV actor | Exponential | Exponential | Exponential |
| 26 | Industrial relations | Undetermined | Undetermined | PL w/ cutoff | 77 | Golden Globe noms. TV actress | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff |
| 27 | Intl. relations | PL | PL | PL | 78 | Grammy noms. | Undetermined | Undetermined | PL w/ cutoff |
| 28 | Law | Undetermined | Undetermined | PL w/ cutoff | 79 | Man Booker Prize Fiction noms. | Undetermined | Undetermined | PL w/ cutoff |
| 29 | Linguistics | Undetermined | Undetermined | PL w/ cutoff | 80 | MTV VMA noms. | Undetermined | Undetermined | PL w/ cutoff |
| 30 | Material sciences | Lognormal | Lognormal | Lognormal | 81 | NYT Best Seller fiction | Undetermined | Undetermined | PL |
| 31 | Mathematics | Lognormal | Lognormal | Lognormal | 82 | NYT Best Seller nonfiction | Undetermined | Undetermined | Undetermined |
| 32 | Medical ethics | Undetermined | Undetermined | PL w/ cutoff | 83 | Oscar noms. actor | Undetermined | Undetermined | PL w/ cutoff |
| 33 | Parasitology | Undetermined | Undetermined | PL w/ cutoff | 84 | Oscar noms. art direction | Undetermined | Undetermined | PL w/ cutoff |
| 34 | Pharmacology | Undetermined | Undetermined | PL w/ cutoff | 85 | Oscar noms. direction | Exponential | Exponential | Exponential |
| 35 | Physics | Undetermined | Undetermined | PL | 86 | Oscar noms. actress | Undetermined | Undetermined | PL w/ cutoff |
| 36 | Public administration | Undetermined | Undetermined | PL w/ cutoff | 87 | Oscar noms. cinematography | Exponential | Exponential | Exponential |
| 37 | Radiology | Undetermined | Undetermined | PL w/ cutoff | 88 | PEN award voting | PL | PL | PL |
| 38 | Rehabilitation | Lognormal | Lognormal | Lognormal | 89 | Pulitzer Prize noms. drama | Undetermined | Undetermined | Undetermined |
| 39 | Rheumatology | Undetermined | Undetermined | Undetermined | 90 | Rolling Stone Top 500 albums | Undetermined | Undetermined | Exponential |
| 40 | Robotics | Undetermined | Undetermined | PL w/ cutoff | 91 | Rolling Stone Top 500 songs | Undetermined | Undetermined | Exponential |
| 41 | Social psych | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff | 92 | Tony noms. actress | Undetermined | Undetermined | PL w/ cutoff |
| 42 | Social work | Undetermined | Undetermined | PL w/ cutoff | 93 | Tony noms. choreography | Undetermined | Undetermined | PL w/ cutoff |
| 43 | Sociology | Undetermined | Undetermined | Undetermined | 94 | Tony noms. actor | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff |
| 44 | Sports medicine | Undetermined | Undetermined | PL w/ cutoff | 95 | Tony noms. director | Undetermined | Undetermined | PL w/ cutoff |
| 45 | Statistics | Lognormal | Lognormal | Lognormal | 96 | Actors | Undetermined | Exponential | Exponential |
| 46 | Substance abuse | Undetermined | Undetermined | PL w/ cutoff | 97 | Actors | Weibull | Weibull | Weibull |
| 47 | Thermodynamics | Lognormal | Lognormal | Lognormal | 98 | Directors | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff |
| 48 | Urban studies | Undetermined | Undetermined | PL w/ cutoff | 99 | Directors | Undetermined | Undetermined | Exponential |
| 49 | Urology | Lognormal | Lognormal | Lognormal | 100 | Producers | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff |
| 50 | Vet. science | Undetermined | Undetermined | PL w/ cutoff | 101 | Producers | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff |
| 51 | Virology | Undetermined | Undetermined | PL w/ cutoff | 102 | Cinematographers | Exponential | Exponential | Exponential |
| 52 | Water science | Undetermined | Undetermined | PL w/ cutoff | 103 | Cinematographers | Exponential | Exponential | Exponential |

Table 3 (*continued*)

| ID | Sample | After decision rule #1 | After decision rules #1 and #2 | After all three decision rules | ID | Sample | After decision rule #1 | After decision rules #1 and #2 | After all three decision rules |
|---|---|---|---|---|---|---|---|---|---|
| 104 | Screenwriters | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff | 157 | EPL goals | Undetermined | Undetermined | Undetermined |
| 105 | Screenwriters | Exponential | Exponential | Exponential | 158 | NBA coaches career wins | Undetermined | Undetermined | PL w/ cutoff |
| 106 | Composers | Undetermined | Undetermined | PL w/ cutoff | 159 | NBA career points | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff |
| 107 | Composers | Exponential | Exponential | Exponential | 160 | PGA career wins | Undetermined | Undetermined | PL w/ cutoff |
| | **Politicians** | | | | 161 | Olympics medals swim (M) | Undetermined | Undetermined | PL w/ cutoff |
| 108 | Alabama Legislature | Undetermined | Undetermined | PL w/ cutoff | 162 | Olympics medals swim (F) | Undetermined | Undetermined | PL w/ cutoff |
| 109 | Australia House (1969) | Exponential | Exponential | Exponential | 163 | Olympics medals track (M) | Undetermined | Undetermined | PL w/ cutoff |
| 110 | Australia House (2009) | Exponential | Exponential | Exponential | 164 | Olympics medals track (F) | Undetermined | Undetermined | Exponential |
| 111 | Canadian Legislature | Weibull | Weibull | Weibull | 165 | Olympics medals alpine (M) | Undetermined | Undetermined | Exponential |
| 112 | Connecticut Legislature | Undetermined | Undetermined | PL w/ cutoff | 166 | Olympics medals alpine (F) | Undetermined | Undetermined | Exponential |
| 113 | Denmark Parliament | Undetermined | Undetermined | PL w/ cutoff | 167 | PBA titles | Undetermined | Undetermined | PL w/ cutoff |
| 114 | Finland Parliament | Exponential | Exponential | Exponential | 168 | NFL career coaches wins | Undetermined | Undetermined | PL w/ cutoff |
| 115 | Georgia House | Undetermined | Undetermined | Exponential | 169 | NFL career kick return yards | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff |
| 116 | Illinois Legislature | Weibull | Weibull | Weibull | 170 | NFL career TD receptions | Undetermined | Undetermined | Exponential |
| 117 | Iowa Legislature | Exponential | Exponential | Exponential | 171 | NFL career field goals | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff |
| 118 | Ireland Parliament | Exponential | Exponential | Exponential | 172 | NFL career sacks | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff |
| 119 | Ireland Senate | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff | 173 | NFL career rushing yards | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff |
| 120 | Kansas House | Undetermined | Undetermined | PL w/ cutoff | 174 | NFL career passing yards | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff |
| 121 | Kansas Senate | Lognormal | Lognormal | Lognormal | 175 | NHL defender assists | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff |
| 122 | Kentucky Legislature | Exponential | Exponential | Exponential | 176 | NHL center points | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff |
| 123 | Louisiana House | Undetermined | Undetermined | Undetermined | 177 | NHL right wing points | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff |
| 124 | Maine Legislature | Undetermined | Undetermined | Exponential | 178 | NHL left wing points | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff |
| 125 | Maryland Legislature | Undetermined | Undetermined | Exponential | 179 | NHL goalie saves | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff |
| 126 | Massachusetts House | Exponential | Exponential | Exponential | 180 | Tennis grand slams men | Exponential | Exponential | Exponential |
| 127 | Minnesota House | Undetermined | Undetermined | Exponential | 181 | Tennis grand slams women | Undetermined | Undetermined | PL w/ cutoff |
| 128 | New York Assembly | Exponential | Exponential | Exponential | 182 | NCAA basketball 2008 points | Undetermined | Undetermined | Exponential |
| 129 | New Zealand Legislature | Exponential | Exponential | Exponential | 183 | MLB career errors 1B | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff |
| 130 | North Carolina Assembly | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff | 184 | MLB career errors 2B | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff |
| 131 | Nova Scotia Legislature | Undetermined | Undetermined | Exponential | 185 | MLB career errors 3B | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff |
| 132 | Oklahoma Legislature | Undetermined | Undetermined | Undetermined | 186 | MLB career errors C | Exponential | Exponential | Exponential |
| 133 | Ontario Legislature | Lognormal | Lognormal | Lognormal | 187 | MLB career errors OF | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff |
| 134 | Oregon Legislature | Exponential | Exponential | Exponential | 188 | MLB career errors SS | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff |
| 135 | Oregon Senate | Exponential | Exponential | Exponential | 189 | EPL yellow cards | Undetermined | Undetermined | PL w/ cutoff |
| 136 | Pennsylvania House | Undetermined | Undetermined | Exponential | 190 | NBA fouls 2005 to 2008 | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff |
| 137 | Quebec Legislature | Undetermined | Undetermined | PL w/ cutoff | 191 | NFL RB fumbles | Undetermined | Undetermined | Exponential |
| 138 | South Carolina House | Exponential | Exponential | Exponential | 192 | NFL QB interceptions | Exponential | Exponential | Exponential |
| 139 | Tasmania Assembly | Exponential | Exponential | Exponential | 193 | NHL DEF penalty minutes | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff |
| 140 | Tennessee House | Undetermined | Undetermined | Exponential | 194 | NHL CR penalty minutes | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff |
| 141 | UK Parliament | Weibull | Weibull | Weibull | 195 | NHL RW penalty minutes | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff |
| 142 | US House | Undetermined | Undetermined | PL w/ cutoff | 196 | NHL LW penalty minutes | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff |
| 143 | US Senate | Undetermined | Undetermined | PL w/ cutoff | 197 | NCAA 2008 QB int. | Weibull | Weibull | Weibull |
| 144 | Virginia Assembly | Undetermined | Undetermined | PL w/ cutoff | | **Additional occupations** | | | |
| 145 | Wisconsin Legislature | Exponential | Exponential | Exponential | 198 | Bank tellers | Undetermined | Undetermined | Normal |
| | **Athletes** | | | | 199 | Bank tellers | Undetermined | Undetermined | Normal |
| 146 | MLB career strikeouts | Undetermined | Undetermined | Exponential | 200 | Bank tellers | Weibull | Weibull | Weibull |
| 147 | MLB career HR | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff | 201 | Bank tellers | Weibull | Weibull | Weibull |
| 148 | MLB career manager wins | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff | 202 | Bank tellers | Undetermined | Undetermined | Normal |
| 149 | NCAA baseball DIV1 HR | Exponential | Exponential | Exponential | 203 | Bank tellers | Undetermined | Undetermined | Undetermined |
| 150 | NCAA baseball DIV2 HR | Exponential | Exponential | Exponential | 204 | Bank tellers | Undetermined | Undetermined | Normal |
| 151 | NCAA baseball DIV3 HR | Undetermined | Undetermined | Exponential | 205 | Bank tellers | Undetermined | Undetermined | Undetermined |
| 152 | NCAA 2008 RB yards | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff | 206 | Retail sales associates | Undetermined | Undetermined | Normal |
| 153 | NCAA 2008 WR yards | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff | 207 | Call center employees | Undetermined | Undetermined | PL w/ cutoff |
| 154 | NCAA 2008 TE yards | Undetermined | Undetermined | PL w/ cutoff | 208 | Call center employees | Undetermined | Undetermined | Normal |
| 155 | Cricket runs | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff | 209 | Call center employees | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff |

(*table continues*)

3 (*continued*)

| ID | Sample | After decision rule #1 | After decision rules #1 and #2 | After all three decision rules | ID | Sample | After decision rule #1 | After decision rules #1 and #2 | After all three decision rules |
|---|---|---|---|---|---|---|---|---|---|
| 156 | Cricket wickets | Undetermined | Undetermined | Exponential | 210 | Fundraising callers | Undetermined | Undetermined | Normal |
| 211 | Fundraising callers | Undetermined | Undetermined | PL w/ cutoff | 221 | Typists | Weibull | Weibull | Weibull |
| 212 | Fundraising callers | Undetermined | Undetermined | Exponential | 222 | Application blank sorters | Lognormal | Lognormal | Lognormal |
| 213 | Call center employees | Undetermined | Exponential | Exponential | 223 | Card punch operators | Undetermined | Undetermined | Normal |
| 214 | Call center employees | PL w/ cutoff | PL w/ cutoff | PL w/ cutoff | 224 | Lamp shade sewers | Undetermined | Undetermined | Exponential |
| 215 | Paper sorters | Undetermined | Undetermined | Normal | 225 | Lamp shade sewers | Undetermined | Undetermined | Undetermined |
| 216 | Grocery checkers | Undetermined | Undetermined | Undetermined | 226 | Card punch operators | Lognormal | Lognormal | Lognormal |
| 217 | Pelt pullers | Undetermined | Undetermined | Undetermined | 227 | Card punch operators | Undetermined | Undetermined | Normal |
| 218 | Curtain and drapery salespeople | Undetermined | Undetermined | Undetermined | 228 | Electrical fixture assemblers | Undetermined | Undetermined | Normal |
| 219 | Casting shop employees | Exponential | Exponential | Exponential | 229 | Wirers | Undetermined | Undetermined | Undetermined |
| 220 | Toll-ticket sorters | Undetermined | Undetermined | Undetermined | | | | | |

*Note.* ID = sample number which matches those in Aguinis et al. (2016); AVN = Adult Video News; ACE = Award for Cable Excellence; MTV = Music Television; VMA = Video Music Awards; NYT = New York Times; PEN = Poets, Playwrights, Editors, Essayists, and Novelists; UK = United Kingdom; US = United States; MLB = Major League Baseball; HR = home run; NBA = National Basketball Association; NCAA = National Collegiate Athletic Association; QB = quarterback; RB = running back; TE = tight end; WR = wide receiver; TD = touchdown; DIV = division; NHL = National Hockey League; NFL = National Football League; PBA = Professional Bowlers Association; PGA = Professional Golfers' Association; EPL = English Premier League; 1B = first baseman; 2B = second baseman; 3B = third baseman; C = catcher; OF = outfielder; SS = shortstop; DEF = defender; CR = center; RW = right wing; LW = left wing; PL = power law; and noms. = nominations. Data set sources: 1–95 and 108–197: O'Boyle and Aguinis (2012); 96–107: information courtesy of Box Office Mojo (2013); 198–205: Sliter, Sliter, and Jex (2012); 206: Erdogan and Bauer (2009); 207–208: Grant, Nurmohamed, Ashford, and Dekas (2011); 209: Grant and Wrzesniewski (2010); 210–212: Grant and Sumanth (2009); 213–214: Grant (2012); 215: Hearnshaw (1937); 216–219: Lawshe (1948); 220–221: Maier and Verser (1982); 222–227: Stead and Shartle (1940); and 228–229: Tiffin and McCormick (1965). Output data refer to: number of publications in top five field-specific journals (for the researchers sample group); number of award nominations, expert rankings, domestic total gross revenue, or number of movies (for the entertainers sample group); number of years served by members of the legislative branch (for the politicians sample group); frequency (e.g., count) of each statistic stated in the sample title (for the athletes sample group); or sales, customer service ratings, idle minutes, hourly revenue, total revenue, number of calls per hour, words per minute typing, number of tasks completed, or ratio of production time per unit (for the additional occupations sample group).

dominant distribution was one of the two exponential tail distributions that are indicators of incremental differentiation as the generative mechanism. Because not using the third decision rule while using just the first two decision rules reduces the accuracy of conclusions (as described in the Method section), we based our conclusions and implications on results derived after using all three decision rules regarding the likely dominant distribution. Nonetheless, as shown in Table 3, results derived after using the first decision rule, the first two decision rules, and all three decision rules all point to the general conclusion that exponential tail distributions likely constitute the dominant model describing individual output distributions. Even after using the first decision rule only, there were many more samples categorized as following one of the two exponential tail distributions (31.01% of samples) compared to samples categorized as following other types of distributions (10.04% of samples; and 58.95% of samples were undetermined). Similarly, after using the first two decision rules and not the third decision rule, there were many more samples categorized as following one of the two exponential tail distributions (31.88% of samples) compared to samples categorized as following other types of distributions (10.04% of samples; 58.08% of samples were undetermined).

To illustrate our findings, consider our first sample (agriculture) for which the likely dominant distribution was the power law with an exponential cutoff, as indicated in Table 3. To arrive at this conclusion, we ran the Dpit package on the agriculture sample and, based on the first decision rule, interpreted the resulting statistics shown in Table 4. From the third to the last column in Table 4, and for the first six rows of statistics corresponding to the agriculture sample (i.e., 21 cells), each loglikelihood ratio (i.e., value *not* in parentheses) and its *p* value (i.e., value in parentheses) indicate how well the two distributions at hand fit the sample. For example, in the cell corresponding to the title NormvPL (i.e., normal vs. pure power law), the loglikelihood ratio and its *p* value were −42.6 and 0, respectively. The loglikelihood ratio had a negative value, which means that the distribution mentioned first in the NormvPL title (i.e., normal distribution) fits the agriculture sample worse than the other distribution (i.e., pure power law). In addition, the *p* value was below the 0.10-cutoff, indicating that normal distribution's worse fit was statistically significant (i.e., not simply because of chance). After interpreting the remaining 20 cells of loglikelihood ratios and their *p* values for the agriculture sample, we found that the power law with an exponential cutoff and the lognormal distribution were never identified as being the worse fitting distribution. We then applied the second decision rule, which states that for two nested distributions, the distribution with more parameters is the worse explanation for the observed distribution. For the agriculture sample, the two remaining distributions (i.e., the power law with an exponential cutoff and the lognormal distribution) are not nested, so neither distribution could be identified as the worse fitting distribution in this step. We then applied the third decision rule, which states that a flexible distribution is the worse explanation for the observed distribution than is an inflexible distribution. Because the lognormal distribution is flexible while the power law with an exponential cutoff is inflexible, we concluded that the power law with an exponential cutoff was likely the dominant distribution for the agriculture sample. We repeated this process for all 229 samples in the study.

Table 4
*Distribution Pitting Statistics: Illustration Based on the First Two and Last Two Samples*

| Occupation | $N$ | NormvPL | NormvCut PLvCut | NormvWeib PLvWeib CutvWeib | NormvLogN PLvLogN CutvLogN WiebvLogN | NormvExp PLvExp CutvExp WeibvExp LogNvExp | NormvPois PLvPois CutvPois WeibvPois LogNvPois ExpvPois |
|---|---|---|---|---|---|---|---|
| 1. Researchers (agriculture) | 25,006 | −42.6 (0) | −43.6 (0) | −43.96 (0) | −43.53 (0) | −47.72 (0) | −57.31 (0) |
| | | | −168.13 (0) | −2.17 (.03) | −10.98 (0) | 20.06 (0) | 21.79 (0) |
| | | | | 6.08 (0) | −.57 (.57) | 22.86 (0) | 22.45 (0) |
| | | | | | −6.37 (0) | 23.65 (0) | 22.65 (0) |
| | | | | | | 22.98 (0) | 22.45 (0) |
| | | | | | | | 22.01 (0) |
| 2. Researchers (agronomy) | 8,923 | −18.36 (0) | −18.9 (0) | −18.88 (0) | −18.89 (0) | −20.21 (0) | −22.97 (0) |
| | | | −53.5 (0) | −4.03 (0) | −5.92 (0) | 6.48 (0) | 9 (0) |
| | | | | .47 (.64) | −2.24 (.03) | 8.67 (0) | 9.95 (0) |
| | | | | | −1.22 (.22) | 9.01 (0) | 10.22 (0) |
| | | | | | | 8.88 (0) | 10.01 (0) |
| | | | | | | | 10.53 (0) |
| 228. Electrical fixture assemblers | 40 | 4.46 (0) | 2.88 (0) | .29 (.77) | −.72 (.47) | 3.19 (0) | 1.68 (.09) |
| | | | −10.97 (0) | −5.38 (0) | −4.65 (0) | −11.35 (0) | −.8 (.42) |
| | | | | −3.67 (0) | −3.11 (0) | 7.62 (0) | .04 (.96) |
| | | | | | −.64 (.52) | 3.91 (0) | 1.45 (.15) |
| | | | | | | 3.4 (0) | 1.77 (.08) |
| | | | | | | | −.12 (.9) |
| 229. Wirers | 35 | 1.37 (.17) | .11 (.91) | −1.7 (.09) | −1.87 (.06) | .36 (.72) | 2.21 (.03) |
| | | | −6.78 (0) | −3.03 (0) | −2.61 (.01) | −5.4 (0) | 1.36 (.17) |
| | | | | −1.66 (.1) | −1.32 (.19) | 3.38 (0) | 1.79 (.07) |
| | | | | | −.08 (.93) | 1.89 (.06) | 2.2 (.03) |
| | | | | | | 1.54 (.12) | 2.27 (.02) |
| | | | | | | | 1.71 (.09) |

*Note.* $N$ = sample size; LR = loglikelihood ratio. Distribution pitting results are presented in the final six columns of the table. For each instance of distribution pitting, the LR value is presented followed by its $p$-value in parentheses. In the first row of the table, distribution names are abbreviated: Norm = Normal; PL = Pure power law; Cut = Power law with an exponential cutoff; Weib = Weibull; LogN = Lognormal; Exp = Exponential; Pois = Poisson. Distribution pitting titles are presented such that the first distribution is compared with the second distribution (e.g., NormvPL = Normal distribution vs. pure power law). Positive LR = superior fit for the first distribution as listed in the distribution pitting title. Negative LR = superior fit for the second distribution as listed in the distribution pitting title. Poisson's LR and $p$-values are not available for continuous data. In the full table version derived by our Dpit R package, there are additional descriptive statistics per sample (e.g., mean, standard deviation) not shown in the abbreviated version here. A complete table including detailed results for each of the 229 samples is available as supplemental material at http://dx.doi.org/10.1037/apl0000214.supp.

## Discussion

Our taxonomy advances theory regarding individual output distributions by offering greater precision beyond the normal versus nonnormal dichotomy adopted in prior studies. Indeed, past research in various fields has noted that without taxonomies, "there could be no advanced conceptualization, reasoning, language, data analysis or, for that matter, social science research" (Bailey, 1994, p. 1). For example, in biology, taxonomies were essential to Darwin's theory about the evolution of organisms (Ereshefsky, 1997; Leather & Quicke, 2009). Research in astronomy used taxonomies to understand various phenomena such as potentially habitable planets (Bailey, 2007). Research in organizational science also documented the importance of more precisely categorizing the multiple dimensions of a construct, such as feedback seeking (Gong, Wang, Huang, & Cheung, 2014), human resource policy regarding older workers (van Dalen, Henkens, & Wang, 2015), organizational context relevant to newcomer socialization (Wang, Kammeyer-Mueller, Liu, & Li, 2015), and personality (Erdheim, Wang, & Zickar, 2006). In our study, similar to past studies demonstrating the importance of taxonomies, we developed a taxonomy that contributes to a better understanding of individual output distributions and their generative mechanisms.

We then offered a novel distribution pitting approach for assessing which types of distributions are better at representing individual output distributions. To facilitate future research, we developed an R package, or Dpit, that implements distribution pitting. Dpit serves as a catalyst for theory advancement via falsification—that is, "pursuit of failure" (Gray & Cooper, 2010) or "theory pruning" (Leavitt et al., 2010). This way, Dpit is a useful methodological tool for reducing dense theoretical landscapes. However, because the shape of an individual output distribution may be the result of multiple mechanisms, we also developed and used three decision rules to determine the likely dominant (rather than the only) shape and generative mechanism per observed distribution.

We subsequently applied Dpit and the three decision rules to 229 samples of individual output, which include about 625,000 individuals across a broad range of occupations, types of individual output measures, types of collectives, and time frames. For 75% of the samples (i.e., 171 out of our 229 individual output samples), results suggest that exponential tail distributions (i.e., exponential and power law with an exponential cutoff) and their associated generative mechanism, incremental differentiation, likely constitute the dominant distribution shape and explanation. Specifically, right-skewed distributions are observed because some individuals enjoy larger output increments (i.e., linear increases in output) than others, and individuals with the highest output accumulation rates may be subject to diminishing returns. Thus, our results challenge past conclusions indicating the pervasiveness of other distributions and their generative mechanisms discussed in our taxonomy—in particular, the pure power law distribution and its generative mechanism (i.e., self-organized criticality). The overwhelming presence of exponential tail distributions and incremental differentiation as their underlying generative mechanism have important implications for theory, future research, and practice, as we explain next.

## Implications for Theory

**Implications for the individual performance literature.** Our results contribute to debates in organizational science regarding the distributional nature of individual performance as well as generative mechanisms of performance distributions (e.g., Aguinis et al., 2016; Beck et al., 2014). In particular, Vancouver et al. (2016) showed that a number of generative mechanisms could operate simultaneously, thereby influencing the shape and amount of positive skew in the resulting performance distributions. In other words, depending on the generative mechanisms, there are many possibilities as to the precise type of distribution describing individual output distributions. Our results help narrow down these possibilities by suggesting that exponential tail distributions likely constitute the dominant model describing individual output distributions.

More importantly, in terms of theory advancement, our study provides evidence regarding the likely dominant generative mechanism underlying observed performance distributions. Specifically, the likely dominant generative mechanism for the shape of individual output distributions is incremental differentiation. In incremental differentiation, some individuals enjoy larger output increments than others, and individuals with the highest output accumulation rates may be subject to diminishing returns. Further, in light of the nature of the samples we analyzed, incremental differentiation suggests that there are significant antecedent differences (i.e., differences in output accumulation rates) even among individuals producing top levels of output—which explain their differences in total output. Specifically, the first four sample groups we used consisted of an already elite subpopulation (i.e., the "cream of the crop"), as shown in Table 3: researchers with publications in top five field-specific journals; award-winning or otherwise highly regarded entertainers; national and state-level politicians; and professional and collegiate athletes. For instance, we examined researchers in an entire field who publish in top journals no matter what their institutional affiliation, as opposed to prior studies that focused specifically on researchers at Research-1 (R1) institutions based on the Carnegie Classification of Institutions of Higher Education (i.e., Beck et al., 2014; Vancouver et al., 2016). The reason is that our focus on an entire field more fully captures the presence of individuals producing elite-level output. A star researcher may choose to go to or stay at a non-R1 school because she is given an idiosyncratic deal (Hornung, Rousseau, Glaser, Angerer, & Weigl, 2010). Indeed, studies indicate that researchers who publish frequently in top journals do not always work at R1 institutions, and not all researchers in R1 institutions publish frequently in top journals (e.g., Aguinis, Suarez-González, Lannelongue, & Joo, 2012). In short, incremental differentiation also helps explain conditional distributions (i.e., subpopulations) of individual output—that is, individual output distributions that are conditional on elite-level output.

In addition, given incremental differentiation as the likely dominant generative mechanism of individual output distributions, our results suggest that it may be largely unnecessary to invoke other generative mechanisms discussed in our taxonomy (i.e., self-organized criticality, proportionate differentiation, and homogenization) to explain individual output distributions. In other words, our results suggest that output differences among individuals generally do not homogenize or increase at a nonlinear/explosive rate. This way, our findings depart from prior studies that have heavily relied on or emphasized homogenization (e.g., Groshen, 1991) or nonlinear increases of individual output differences (e.g., Andriani & McKelvey, 2009; Vancouver et al., 2016). By decreasing the importance of borrowing from generative mechanisms other than incremental differentiation, our results also contribute to theory reduction. Reducing theory is important because it creates greater theoretical parsimony and allows researchers to "reduce areas of focus and avoid time spent on fruitless avenues of inquiry" (Leavitt et al., 2010, p. 645).

In particular, we found that the pure power law distribution and its generative mechanism, self-organized criticality, are not as useful for explaining individual output distributions. This finding largely challenges previous research invoking the pure power law distribution or self-organized criticality to explain individual output distributions. For example, according to extant theory, pure power law distribution's generative mechanism may lead to extreme events that are "not predictable" and may affect multiple units of analysis including individuals (Andriani & McKelvey, 2009, p. 1066). Prior research has also proposed that pure power law distributions of individual output require theory focused on "plausible anticipation" rather than "prediction" of extreme outcomes (Crawford, 2012, p. 79). Though self-organized criticality may be needed to explain various strategy and entrepreneurship issues (Andriani & McKelvey, 2009; Crawford et al., 2015), our results suggest that self-organized criticality is not a very important framework for explaining individual output distributions.

We also found that the lognormal distribution and its generative mechanism, proportionate differentiation, are not as important for explaining individual output distributions. This finding largely departs from prior studies on individuals that have invoked concepts reflecting proportionate differentiation, which suggests the presence of positive feedback between past and future individual output. For instance, research has found that heavily right-tailed distributions of individual output are more likely to emerge in contexts where it is easier for individuals to draw on past success to create future success (Aguinis et al., 2016, pp. 10–11). Further, Vancouver et al. (2016, Simulation 10) found that heavily right-tailed distributions of individual output emerge when the more highly productive individuals receive greater amounts of resources to be even more productive over time, while the less productive individuals receive the same or lower amounts of resources and thus become limited in their future output. However, our results suggest that proportionate differentiation is not the most likely dominant generative mechanism of individual output distributions.

We do not mean, however, to imply that it is incorrect to consider the involvement of any feedback or other contextual processes to explain output differences among individuals. Incremental differentiation does allow for an explanation based on *limited* instances of nonincremental increases in individual output. Some individuals may experience unpredictable and extremely large output increases, or output shocks. Some other individuals may disproportionately benefit from increasingly larger output increases based on positive feedback between past and future output, or output loops. For example, more people may consume an artist's work by chance and, as a result, the artist may enjoy additional patrons who generally prefer more popular artists. Similarly, as another example, coaches can provide a weightlifter perceived as Olympics-worthy with superior training and other valuable resources, which may lead to some explosive, nonincremental increases in individual output.

Nonetheless, because of the likely dominant role of incremental differentiation, such output shocks or loops would be short-lived and weak enough so that individuals accumulate output in a primarily incremental manner (i.e., at a linear rate). In other words, the process of reaching star-level individual output appears largely incremental instead of being largely characterized by output shocks or loops. Going back to the example involving artists, an artist must have first created and improved her art piece by piece for chance, publicity, or other types of events to create a positive feedback loop between past and future output. Further, the role of any feedback processes would ultimately be limited because an artist can only create or improve her art at an incremental rate (e.g., Ericsson & Charness, 1994). Similarly, going back to the example involving weightlifters, the role of any output shock or loop would also be short-lived because, soon afterward, an Olympic weightlifter must return to painstakingly adding 1, 2, or 3 kg at a time to her personal record as well as making sure to take time off or at least "de-load" on the amount of weight she lifts to prevent injury, among other things. This interpretation of incremental differentiation is consistent with prior studies where allocation of additional resources (e.g., opportunities) per se led to some, but ultimately limited, positive feedback loops between past and future individual output (e.g., McNatt & Judge, 2004; van de Rijt et al., 2014). Thus, incremental differentiation means that feedback loops and similar mechanisms do not play a prominent role in the individual output context.

**Implications for theory development.** Our results also have implications for developing theory regarding the link between past and future individual output. On one hand, the "samples" approach to predicting future individual output suggests that past individual output would be the best predictor (Aguinis et al., 2016; Wernimont & Campbell, 1968). The underlying rationale is that the closer the point-to-point correspondence between the samples used for the predictor and criterion, the greater is the predictive power. On the other hand, the "signs" approach suggests that knowledge, skills, abilities, and other individual characteristics (KSAOs)—or signs of future individual output—would be better predictors of future individual output than is past individual output (Bangerter, Roulin, & König, 2012; Callinan & Robertson, 2000). Though prior evidence suggests that KSAOs are generally superior or equally effective compared to past individual output as predictors of future individual output (Kuncel, Hezlett, & Ones, 2001), much of the recent successes in the model focused on KSAOs and using validity coefficients as a metric are not because of substantive improvements to our understanding of what predicts performance but instead because of the use of statistical correction techniques (Bosco, Aguinis, Singh, Field, & Pierce, 2015; Cascio & Aguinis, 2008). Further, the signs approach focused on KSAOs "seems to have reached a ceiling or plateau in terms of its ability to make accurate predictions about future [individual output]" (Cascio & Aguinis, 2008, p. 141). So, there is a need to go beyond the dominant model emphasizing KSAOs and gain a more precise understanding of the link between past and future individual output. In the following, and based on our findings, we develop theoretical premises about the relation between past and future individual output.

First, our results suggest that not all types of past individual output would effectively predict future individual output. Specifically, proportionate differentiation implies that past individual output in terms of both output accumulation rate and initial output predict future individual output. In contrast, incremental differentiation implies that past individual output in terms of output accumulation rate, but not initial output, predicts future individual output. Given the two competing perspectives, we found that incremental differentiation, not proportionate differentiation, likely constitutes the dominant generative mechanism for most of our individual output samples. Thus, results indicate that past individual output in terms of output accumulation rate, but not initial output, would help significantly predict future individual output. This is an important implication because, despite some evidence that past individual output often does predict future individual output, prior research has not been clear as to how different definitions and operationalizations of past individual output may affect the prediction of future individual output. Instead, studies have focused on showing how a particular operationalization of past individual output predicts future individual output. As examples, studies have operationalized past individual output in terms of "the last year of [objectively measured] collegiate performance" (Lyons, Hoffman, Michel, & Williams, 2011, p. 162) or "gross sales commissions averaged across a 3-month period" (Zyphur, Chaturvedi, & Arvey, 2008, p. 220). Given such, our results more precisely clarify that future individual output would be better predicted by past individual output in terms of output accumulation rate rather than initial output.

Second, our results suggest that high variability in individual output would be followed by even higher, not lower, variability in individual output in the future. That is, (potentially) symmetric distributions are associated with homogenization, which implies that variability in past individual output would be followed by lower variability in future individual output. However, our results showed that (potentially) symmetric distributions are less likely explanations for most of our individual output samples. Further, we found that the dominant generative mechanism for most of our individual output samples was incremental differentiation, which suggests that high variability in individual output would be followed by even higher variability in individual output in the future, especially if there is high variability across individuals in terms of output accumulation rate.

## Implications for Future Research

Moving forward, it would be interesting to use distribution pitting along with computational modeling to triangulate on theories regarding the individual output distributions. For example, future research may simulate distributions of publications based on a number of models that incorporate different sets of individual characteristics (e.g., proactivity in terms of seeking out collaborators). The same study may then infer the most likely simulation model by using distribution pitting to identify distributions that follow exponential tail distributions—given that the majority of our researcher publication samples followed exponential tail distributions (associated with the generative mechanism of incremental differentiation).

In addition, future studies can examine the boundary conditions of incremental differentiation as the generative mechanism of individual output distributions. We described incremental differentiation as a model where differences among individuals in terms of output accumulation rate, but not initial output, predict their future output levels. So, one future research direction is to operationalize individuals' output accumulation rates in many different ways, and check whether their relations with future individual output differ from one another. Another direction is to examine moderators of the relation between output accumulation rate and future individual output. For example, is the relation between output accumulation rate and future individual output weaker among newcomers because their output-related information is more ambiguous than that of incumbents (Terviö, 2009)?

Future research may also identify contexts that are better characterized by generative mechanisms other than incremental differentiation. One possibility is that incremental differentiation applies to individual output measures referring to raw output, but not to measures referring to value generated from an individual's raw output. As one type of value-based measures of individual output, financial measures have more upside potential (e.g., tripling of a company's stock value) than nonfinancial measures (e.g., number of competitions won; Aoki & Yoshikawa, 2006). For example, top performers may obtain opportunities (e.g., a CEO's movement to a large corporation) that allow them to increase financial output by multiple folds within a relatively short period of time (e.g., rapid increases in market capitalization attributed to a CEO). So, individual output based on financial measures may instead follow pure power law or lognormal distributions, thus indicating self-organized criticality or proportionate differentiation as the likely dominant generative mechanism, respectively. Other examples of value-based measures of individual output include revenue generated by entrepreneurs, managers, inventors (e.g., of the next major drug), high-tech employees (e.g., programmers), and salespeople dealing with big contracts (Aguinis et al., 2016; Aguinis et al., 2017; Crawford et al., 2015)—that our data set did not include. In fact, our individual output data mainly consisted of raw output rather than value generated from an individual's raw output, as shown in Table 3 (exceptions are samples #96, #98, #100, #102, #104, #106, #198–199, #206–207, #209, #211, and #214, which are financial measures). In short, future research can adopt a deductive approach to identify contexts where incremental differentiation may not constitute the dominant explanation of individual output distributions.

Finally, Dpit can be used for assessing the distribution of performance defined as not only output but also behavior. For example, future research can assess the distribution of organizational citizenship behaviors, counterproductive work behaviors, or aggression toward others (Liu et al., 2015). The use of Dpit may lead to the discovery that, for example, a certain type of performance behavior follows an exponential distribution, indicating incremental differentiation as the generative mechanism. In contrast, the use of Dpit may show that other types of performance behavior follow (potentially) symmetric distributions, suggesting the presence of homogenization as the generative mechanism. Dpit can also be used for assessing the distribution of other events in organizational science research such as accidents and errors (made during error management training), which would contribute to the safety as well as training and development literatures. Closely examining the shape of performance behavior distributions will, in turn, lead to a better understanding of the underlying generative mechanisms.

## Implications for Practice

Given incremental differentiation as the likely dominant explanation of individual output distributions, our results suggest that higher variability in output accumulation rates will be associated with a greater proportion of top performers. To facilitate even greater overall output and production of top performers, an organization could heavily select for individuals with the highest levels of output accumulation rates. The organization can also disproportionately invest in the training and development of individuals with already the highest levels of output accumulation rates so that they reach even higher levels of future output. We acknowledge that the shapes of individual output distributions may change (e.g., increased length of the right tail) as a result of following our practical advice. One reason is that our practical recommendations may lead to nonincremental increases in individual output among some individuals (e.g., feedback processes). However, as previously discussed in the Implications for Theory section, we also clarify that such nonincremental increases would be limited in duration and strength because of the likely dominant role of incremental differentiation. Our practical advice would thus likely affect the parameter values of exponential tail distributions, rather than changing the type of the distribution (e.g., into a lognormal or pure power law distribution).

We also derive practical implications directly from the shape of exponential tail distributions. For example, based on fairness theory, allocation of pay and other types of valued resources should reflect the shape of individual output distributions (Aguinis & O'Boyle, 2014). We more precisely recommend that compensation practices reflect exponential tail distributions of individual output. So, differences among individuals in terms of compensation should be smaller than those implied in more heavily right-tailed distributions (e.g., pure power law distribution, lognormal distribution). An organization may also benefit from ensuring relatively small compensation differences among top performers, as indicated by the often, and not always, low variability in an exponential tail distribution's right tail. Otherwise, large differences in compensation among top performers may lead to perceptions of unfairness, which can create or aggravate harmful competition and workplace conflict among those top performers (Groysberg, Polzer, & Elfenbein, 2011). This does not mean that organizations should maintain small compensation differences among all individuals. Though small differences in compensation among top performers may be warranted in an exponential tail distribution of individual output, the distribution still allows for large differentiation between top and ordinary performers.

## Limitations and Additional Future Research Directions

We found that the two exponential tail distributions (i.e., indicators of incremental differentiation) likely constitute the dominant distribution shape for the majority of our samples—that is, for each of the first four sample groups (i.e., researchers, entertainers, politicians, and athletes) as shown in Table 3. However, for most of the samples in the last sample group, "additional occupations" (e.g., manufacturing, service, and clerical jobs), neither of the two exponential tail distributions was the likely dominant distribution. That is, our methodological procedures led to the conclusion that exponential tail distributions are the likely dominant distribution in 8 out of the 32 samples in the additional occupations sample group. This raises the question of whether and how much our study's findings and implications are generalizable to the additional occupations sample group. So, a future research direction is to assess whether and how our results generalize to other occupations, jobs, and organizational contexts.

Further, our results and implications apply to interindividual distributions of output, but not necessarily to the intraindividual level of analysis. For instance, even though we found that incremental differentiation constitutes the likely dominant generative mechanism at the interindividual level of output, follow-up studies may find that individuals' output accumulation rates mask theoretically and practically significant punctuations and bursts in output accumulation within short periods of time. The reason is that processes that generate differences among individuals are not necessarily the same as processes that generate differences within an individual over time (Dalal et al., 2014, 2009; Molenaar, 2004). Using our methodological procedures, future research can examine the shape of intraindividual distributions as well as their underlying generative mechanism(s).

Finally, the three decision rules that we used for implementing distribution pitting should not be interpreted as leading to clear-cut, black-and-white results. Instead, our decision rules are designed to help the user choose the most likely dominant distribution for a given dataset, given that the shape of an individual output distribution may be the result of multiple mechanisms operating simultaneously. In the future, methodological advances may allow the user to identify and weigh the importance of each mechanism contributing to the shape of an individual output distribution.

## Concluding Comments

In retrospect, perhaps it was inevitable that we would derive novel findings that to date have been masked by the normal versus nonnormal dichotomy. In fields such as physics, research has found that "a much more common distribution than the [pure] power law is the exponential [distribution], which arises in many circumstances" (Newman, 2005, p. 336). From a historical perspective, research in the natural sciences and mathematics offered some of the earliest discussions about nonnormality in terms of the pure power law distribution (e.g., Bak, 1996; Mandelbrot, 1963), which was later introduced to the social sciences (e.g., Andriani & McKelvey, 2009; Boisot & McKelvey, 2011), which in turn influenced organizational science research (e.g., Aguinis et al., 2016; O'Boyle & Aguinis, 2012). Meanwhile, questioning the ubiquity of pure power law distributions, research in other fields has also accumulated evidence that many phenomena conform better to other types of nonnormal and heavily right-tailed distributions (e.g., Downey, 2001; Mitzenmacher, 2004; Newman, 2005; Stumpf & Porter, 2012). Yet, this body of work had not been incorporated into organizational science research regarding individual output distributions—that is, until now in our study.

Our taxonomy, distribution pitting methodology along with the Dpit package, and empirical results derived from implementing Dpit provided novel insights about individual output distributions and their generative mechanisms. We found that exponential tail distributions and their generative mechanism, incremental differentiation, likely constitute the dominant distribution shape and explanation of most individual output distributions. Thus, our results challenge past conclusions indicating the pervasiveness of other distributions and their generative mechanisms discussed in our taxonomy. In particular, our results challenge the pervasiveness of the pure power law distribution and its generative mechanism (i.e., self-organized criticality). We hope our taxonomy and the Dpit package will serve as catalysts for future theory advancements regarding the distributions and generative mechanisms of individual output and many other variables in organizational science.

## References


Aguinis, H. (2013). *Performance management* (3rd ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Aguinis, H., Martin, G. P., Gomez-Mejia, L. R., O'Boyle, E. H., & Joo, H. (2017). The two sides of CEO pay injustice: A power law conceptualization of CEO over and underpayment. *Management Research: Journal of the Iberoamerican Academy of Management*. Advance online publication. http://dx.doi.org/10.1108/MRJIAM-02-2017-0731

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

Aguinis, H., & Vandenberg, R. J. (2014). An ounce of prevention is worth a pound of cure: Improving research quality before data collection. *Annual Review of Organizational Psychology and Organizational Behavior, 1,* 569–595. http://dx.doi.org/10.1146/annurev-orgpsych-031413-091231

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

Andriani, P., & McKelvey, B. (2009). From Gaussian to Paretian thinking: Causes and implications of power laws in organizations. *Organization Science, 20,* 1053–1071. http://dx.doi.org/10.1287/orsc.1090.0481

Aoki, M., & Yoshikawa, H. (2006). Stock prices and the real economy: Power law versus exponential distributions. *Journal of Economic Interaction and Coordination, 1,* 45–73. http://dx.doi.org/10.1007/s11403-006-0003-x

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

Bailey, J. (2007). Rainbows, polarization, and the search for habitable planets. *Astrobiology, 7,* 320–332. http://dx.doi.org/10.1089/ast.2006.0039

Bailey, K. D. (1994). *Typologies and taxonomies: An introduction to classification techniques*. Thousand Oaks, CA: Sage. http://dx.doi.org/10.4135/9781412986397

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

Balasooriya, U., & Abeysinghe, T. (1994). Selecting between gamma and Weibull distributions: An approach based on predictions of order statistics. *Journal of Applied Statistics, 21,* 17–27. http://dx.doi.org/10.1080/757583646

Banerjee, A., & Yakovenko, V. M. (2010). Universal patterns of inequality. *New Journal of Physics, 12,* 075032. http://dx.doi.org/10.1088/1367-2630/12/7/075032

Bangerter, A., Roulin, N., & König, C. J. (2012). Personnel selection as a signaling game. *Journal of Applied Psychology, 97,* 719–738. http://dx.doi.org/10.1037/a0026078

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

Barabási, A. L., & Bonabeau, E. (2003). Scale-free networks. *Scientific American, 288,* 60–69. http://dx.doi.org/10.1038/scientificamerican0503-60

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

Berk, A., & Kaše, R. (2010). Establishing the value of flexibility created by training: Applying real options methodology to a single HR practice. *Organization Science, 21,* 765–780. http://dx.doi.org/10.1287/orsc.1090.0476

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

Boisot, M., & McKelvey, B. (2011). Connectivity, extremes, and adaptation: A power-law perspective of organizational effectiveness. *Journal of Management Inquiry, 20,* 119–133. http://dx.doi.org/10.1177/1056492610385564

Bommer, W. H., Johnson, J. L., Rich, G. A., Podsakoff, P. M., & MacKenzie, S. B. (1995). On the interchangeability of objective and subjective measures of employee performance: A meta-analysis. *Personnel Psychology, 48,* 587–605. http://dx.doi.org/10.1111/j.1744-6570.1995.tb01772.x

Bosco, F. A., Aguinis, H., Singh, K., Field, J. G., & Pierce, C. A. (2015). Correlational effect size benchmarks. *Journal of Applied Psychology, 100,* 431–449. http://dx.doi.org/10.1037/a0038047

Box Office Mojo. (2013). *People index*. Retrieved from http://www.boxofficemojo.com/people/

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

Callinan, M., & Robertson, I. T. (2000). Work sample testing. *International Journal of Selection and Assessment, 8,* 248–260. http://dx.doi.org/10.1111/1468-2389.00154

Camerer, C., Babcock, L., Loewenstein, G., & Thaler, R. (1997). Labor supply of New York City cabdrivers: One day at a time. *Quarterly Journal of Economics, 112,* 407–441. http://dx.doi.org/10.1162/003355397555244

Campbell, J. P. (1990). Modeling the performance prediction problem in industrial and organizational psychology. In M. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 1, pp. 687–731). Palo Alto, CA: Consulting Psychologists Press.

Cascio, W. F., & Aguinis, H. (2008). Staffing twenty-first-century organizations. *Academy of Management Annals, 2,* 133–165. http://dx.doi.org/10.1080/19416520802211461

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

Crawford, G. C. (2012). Disobeying power laws: Perils for theory and method. *Journal of Organization Design, 1,* 75–81. http://dx.doi.org/10.7146/jod.6419

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

Czirók, A., Schlett, K., Madarász, E., & Vicsek, T. (1998). Exponential distribution of locomotion activity in cell cultures. *Physical Review Letters, 81,* 3038–3041. http://dx.doi.org/10.1103/PhysRevLett.81.3038

Dalal, R. S., Bhave, D. P., & Fiset, J. (2014). Within-person variability in job performance: A theoretical review and research agenda. *Journal of Management, 40,* 1396–1436. http://dx.doi.org/10.1177/0149206314532691

Dalal, R. S., Lam, H., Weiss, H. M., Welch, E. R., & Hulin, C. L. (2009). A within-person approach to work behavior and performance: Concurrent and lagged citizenship-counterproductivity associations, and dynamic relationships with affect and overall job performance. *Academy of Management Journal, 52,* 1051–1066. http://dx.doi.org/10.5465/AMJ.2009.44636148

Das, M., Das, H., & Mackenzie, F. (1996). Satisfaction with town services: Do user experience and satisfaction count? *Service Industries Journal, 16,* 191–205. http://dx.doi.org/10.1080/02642069600000021

Dierickx, I., & Cool, K. (1989). Asset stock accumulation and sustainability of competitive advantage. *Management Science, 35,* 1504–1511. http://dx.doi.org/10.1287/mnsc.35.12.1504

Downey, A. B. (2001). *The structural cause of file size distributions*. Paper presented at MASCOTS '01: The Ninth International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems, August 15–18, Cincinnati, OH, Menlo Park: IEEE Press. http://dx.doi.org/10.1109/MASCOT.2001.948888

Erdheim, J., Wang, M., & Zickar, M. J. (2006). Linking the Big Five personality constructs to organizational commitment. *Personality and Individual Differences, 41,* 959–970. http://dx.doi.org/10.1016/j.paid.2006.04.005

Erdogan, B., & Bauer, T. N. (2009). Perceived overqualification and its outcomes: The moderating role of empowerment. *Journal of Applied Psychology, 94,* 557–565. http://dx.doi.org/10.1037/a0013528

Ereshefsky, M. (1997). The evolution of the Linnaean hierarchy. *Biology & Philosophy, 12,* 493–519. http://dx.doi.org/10.1023/A:1006556627052

Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist, 49,* 725–747. http://dx.doi.org/10.1037/0003-066X.49.8.725

Gabaix, X. (1999). Zipf's Law for cities: An explanation. *Quarterly Journal of Economics, 114,* 739–767. http://dx.doi.org/10.1162/003355399556133

Gibrat, R. (1931). *Les inégalités économiques*. Paris: Recueil Sirey.

Gillespie, C. S. (2014). Fitting heavy tailed distributions: The poweRlaw package. *arXiv: 1407.3492*. Retrieved from http://arxiv.org/abs/1407.3492

Gong, Y., Wang, M., Huang, J. C., & Cheung, S. Y. (2014). Toward a goal orientation-based feedback-seeking typology: Implications for employee performance outcomes. *Journal of Management*. Advance online publication. http://dx.doi.org/10.1177/0149206314551797

Grant, A. M. (2012). Leading with meaning: Beneficiary contact, prosocial impact, and the performance effects of transformational leadership. *Academy of Management Journal, 55,* 458–476. http://dx.doi.org/10.5465/amj.2010.0588

Grant, A. M., Nurmohamed, S., Ashford, S. J., & Dekas, K. (2011). The performance implications of ambivalent initiative: The interplay of autonomous and controlled motivations. *Organizational Behavior and*

*Human Decision Processes, 116,* 241–251. http://dx.doi.org/10.1016/j.obhdp.2011.03.004

Grant, A. M., & Sumanth, J. J. (2009). Mission possible? The performance of prosocially motivated employees depends on manager trustworthiness. *Journal of Applied Psychology, 94,* 927–944. http://dx.doi.org/10.1037/a0014391

Grant, A. M., & Wrzesniewski, A. (2010). I won't let you down . . . or will I? Core self-evaluations, other-orientation, anticipated guilt and gratitude, and job performance. *Journal of Applied Psychology, 95,* 108–121. http://dx.doi.org/10.1037/a0017974

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

Groysberg, B., Polzer, J. T., & Elfenbein, H. A. (2011). Too many cooks spoil the broth: How high-status individuals decrease group effectiveness. *Organization Science, 22,* 722–737. http://dx.doi.org/10.1287/orsc.1100.0547

Gupta, R. D., & Kundu, D. (1999). Generalized exponential distributions. *Australian & New Zealand Journal of Statistics, 41,* 173–188. http://dx.doi.org/10.1111/1467-842X.00072

Gupta, R. D., & Kundu, D. (2001). Exponentiated exponential family: An alternative to gamma and Weibull distributions. *Biometrical Journal Biometrische Zeitschrift, 43,* 117–130. http://dx.doi.org/10.1002/1521-4036(200102)43:1<117::AID-BIMJ117>3.0.CO;2-R

Hambrick, D. C. (2007). The field of management's devotion to theory: Too much of a good thing? *Academy of Management Journal, 50,* 1346–1352. http://dx.doi.org/10.5465/AMJ.2007.28166119

Hearnshaw, L. S. (1937). Selection tests for paper sorters. *Journal of Occupational Psychology, 11,* 145–153.

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

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

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the graduate record examinations: Implications for graduate student selection and performance. *Psychological Bulletin, 127,* 162–181. http://dx.doi.org/10.1037/0033-2909.127.1.162

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. 205–259). New York, NY: Cambridge University Press. http://dx.doi.org/10.1007/978-94-010-1863-0_14

Lawshe, C. H. (1948). *Principles of personnel testing*. New York, NY: McGraw-Hill.

Leather, S. R., & Quicke, D. J. L. (2009). Where would Darwin have been without taxonomy? *Journal of Biological Education, 43,* 51–52. http://dx.doi.org/10.1080/00219266.2009.9656151

Leavitt, K., Mitchell, T. R., & Peterson, J. (2010). Theory pruning: Strategies to reduce our dense theoretical landscape. *Organizational Research Methods, 13,* 644–667. http://dx.doi.org/10.1177/1094428109345156

Lenoir, A., Hefetz, A., Simon, T., & Soroker, V. (2001). Comparative dynamics of gestalt odour formation in two ant species: *Camponotus fellah* and *Aphaenogaster senilis* (Hymenoptera: Formicidae). *Physiological Entomology, 26,* 275–283. http://dx.doi.org/10.1046/j.0307-6962.2001.00244.x

Leroy, H., Dierynck, B., Anseel, F., Simons, T., Halbesleben, J. R. B., McCaughey, D., . . . Sels, L. (2012). Behavioral integrity for safety, priority of safety, psychological safety, and patient safety: A team-level study. *Journal of Applied Psychology, 97,* 1273–1281. http://dx.doi.org/10.1037/a0030076

Limpert, E., Stahel, W. A., & Abbt, M. (2001). Log-normal distributions across the sciences: Keys and clues. *Bioscience, 51,* 341–352. http://dx.doi.org/10.1641/0006-3568(2001)051[0341:LNDATS]2.0.CO;2

Liu, Y., Wang, M., Chang, C.-H., Shi, J., Zhou, L., & Shao, R. (2015). Work-family conflict, emotional exhaustion, and displaced aggression toward others: The moderating roles of workplace interpersonal conflict and perceived managerial family support. *Journal of Applied Psychology, 100,* 793–808. http://dx.doi.org/10.1037/a0038387

Lyons, B. D., Hoffman, B. J., Michel, J. W., & Williams, K. J. (2011). On the predictive efficiency of past performance and physical ability: The case of the National Football League. *Human Performance, 24,* 158–172. http://dx.doi.org/10.1080/08959285.2011.555218

Maier, N. R. F., & Verser, G. C. (1982). *Psychology in industrial organizations* (5th ed.). Boston, MA: Houghton Mifflin.

Malevergne, Y., Pisarenko, V., & Sornette, D. (2005). Empirical distributions of stock returns: Between the stretched exponential and the power law? *Quantitative Finance, 5,* 379–401. http://dx.doi.org/10.1080/14697680500151343

Mandelbrot, B. B. (1963). The variation of certain speculative prices. *Journal of Business, 36,* 394–419. http://dx.doi.org/10.1086/294632

Mandelbrot, B. B., & Taleb, N. N. (2010). Mild vs. wild randomness: Focusing on those risks that matter. In F. X. Diebold, N. A. Doherty, & R. J. Herring (Eds.), *The known, the unknown and the unknowable in financial institutions* (pp. 47–58). Princeton, NJ: Princeton University Press. http://dx.doi.org/10.1515/9781400835287-004

Mankiw, N. G. (2013). Defending the one percent. *Journal of Economic Perspectives, 27,* 21–34. http://dx.doi.org/10.1257/jep.27.3.21

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

McNatt, D. B., & Judge, T. A. (2004). Boundary conditions of the Galatea effect: A field experiment and constructive replication. *Academy of Management Journal, 47,* 550–565. http://dx.doi.org/10.2307/20159601

Meng, J., & Burris, S. (2013). The role of the Chinese police in methadone maintenance therapy: A literature review. *International Journal on Drug Policy, 24,* e25–e34. http://dx.doi.org/10.1016/j.drugpo.2013.03.010

Miller, M. E. (2000). Troubled times for med school teachers. *Hopkins Medical News, Winter Issue*. Retrieved from http://www.hopkinsmedicine.org/hmn/w00/top.html

Minbashian, A., & Luppino, D. (2014). Short-term and long-term within-person variability in performance: An integrative model. *Journal of Applied Psychology, 99,* 898–914. http://dx.doi.org/10.1037/a0037402

Mitzenmacher, M. (2004). A brief history of generative models for power law and lognormal distributions. *Internet Mathematics, 1,* 226–251. http://dx.doi.org/10.1080/15427951.2004.10129088

Molenaar, P. C. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement, 2,* 201–218.

Mossa, S., Barthélémy, M., Stanley, H. E., & Nunes Amaral, L. A. (2002). Truncation of power law behavior in "scale-free" network models due to information filtering. *Physical Review Letters, 88,* 138701. http://dx.doi.org/10.1103/PhysRevLett.88.138701

Newby, M., & Winterton, J. (1983). The duration of industrial stoppages. *Journal of the Royal Statistical Society Series A (General), 146,* 62–70. http://dx.doi.org/10.2307/2981488

Newman, M. E. J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics, 46,* 323–351. http://dx.doi.org/10.1080/00107510500052444

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

O'Boyle, E., Jr., & Aguinis, H. (2012). The best and the rest: Revisiting the norm of normality of individual performance. *Personnel Psychology, 65,* 79–119. http://dx.doi.org/10.1111/j.1744-6570.2011.01239.x

Pichon, G., Awono-Ambene, H. P., & Robert, V. (2000). High heterogeneity in the number of Plasmodium falciparum gametocytes in the bloodmeal of mosquitoes fed on the same host. *Parasitology, 121,* 115–120. http://dx.doi.org/10.1017/S0031182099006277

Pierce, J. R., & Aguinis, H. (2013). The too-much-of-a-good-thing effect in management. *Journal of Management, 39,* 313–338. http://dx.doi.org/10.1177/0149206311410060

Platzer, M., & Reutterer, T. (2016). Ticking away the moments: Timing the regularity helps to better predict customer activity. *Marketing Science, 35,* 779–799. http://dx.doi.org/10.1287/mksc.2015.0963

Ployhart, R. E., Weekley, J. A., & Baughman, K. (2006). The structure and function of human capital emergence: A multilevel examination of the attraction-selection-attrition model. *Academy of Management Journal, 49,* 661–677. http://dx.doi.org/10.5465/AMJ.2006.22083023

Poincaré, H. (1921). *The foundations of science: Science and hypothesis, the value of science, and science and method* (G. B. Halstead, Trans.). New York, NY: Science Press.

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

Rinne, H. (2008). *The Weibull distribution: A handbook*. Boca Raton, FL: CRC Press. http://dx.doi.org/10.1201/9781420087444

Sackett, P. R., Gruys, M. L., & Ellingson, J. E. (1998). Ability-personality interactions when predicting job performance. *Journal of Applied Psychology, 83,* 545–556. http://dx.doi.org/10.1037/0021-9010.83.4.545

Salganik, M. J., Dodds, P. S., & Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science, 311,* 854–856. http://dx.doi.org/10.1126/science.1121066

Schneider, B. (1987). The people make the place. *Personnel Psychology, 40,* 437–453. http://dx.doi.org/10.1111/j.1744-6570.1987.tb00609.x

Seminogov, V. N., Semchishen, V. A., Panchenko, V. Y., Seiler, T., & Mrochen, M. (2002). Scattering of light beams propagating through a dielectric surface with a large-scale roughness: II. Gaussian and non-Gaussian homogenization of scattered beams. *Laser Physics, 12,* 1333–1348.

Shen, W., Kiger, T. B., Davies, S. E., Rasch, R. L., Simon, K. M., & Ones, D. S. (2011). Samples in applied psychology: Over a decade of research in review. *Journal of Applied Psychology, 96,* 1055–1064. http://dx.doi.org/10.1037/a0023322

Simonton, D. K. (2003). Scientific creativity as constrained stochastic behavior: The integration of product, person, and process perspectives. *Psychological Bulletin, 129,* 475–494. http://dx.doi.org/10.1037/0033-2909.129.4.475

Sitzmann, T., & Yeo, G. (2013). A meta-analytic investigation of the within-person self-efficacy domain: Is self-efficacy a product of past performance or a driver of future performance? *Personnel Psychology, 66,* 531–568. http://dx.doi.org/10.1111/peps.12035

Sliter, M., Sliter, K., & Jex, S. (2012). The employee as a punching bag: The effect of multiple sources of incivility on employee withdrawal behavior and sales performance. *Journal of Organizational Behavior, 33,* 121–139. http://dx.doi.org/10.1002/job.767

Sornette, D. (2006). *Critical phenomena in natural sciences* (2nd ed.). Stuttgart, Germany: Springer.

Sornette, D., & Ouillon, G. (2012). Dragon-kings: Mechanisms, statistical methods and empirical evidence. *European Physical Journal Special Topics, 205,* 1–26. http://dx.doi.org/10.1140/epjst/e2012-01559-5

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

Stead, W. H., & Shartle, C. L. (1940). *Occupational counseling techniques: Their development and application*. New York, NY: American Book.

Stumpf, M. P. H., & Porter, M. A. (2012). Mathematics: Critical truths about power laws. *Science, 335,* 665–666. http://dx.doi.org/10.1126/science.1216142

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

Taleb, N. N. (2012). *Antifragile: Things that gain from disorder*. New York, NY: Random House.

Tepeci, M. (1999). Increasing brand loyalty in the hospitality industry. *International Journal of Contemporary Hospitality Management, 11,* 223–230. http://dx.doi.org/10.1108/09596119910272757

Terviö, M. (2009). Superstars and mediocrities: Market failure in the discovery of talent. *Review of Economic Studies, 76,* 829–850. http://dx.doi.org/10.1111/j.1467-937X.2008.00522.x

Tiffin, J., & McCormick, E. J. (1965). *Industrial psychology* (5th ed.). Englewood Cliffs, NJ: Prentice Hall.

Vancouver, J. B., Li, X., Weinhardt, J. M., Steel, P., & Purl, J. D. (2016). Using a computational model to understand possible sources of skews in distributions of job performance. *Personnel Psychology, 69,* 931–974. http://dx.doi.org/10.1111/peps.12141

van Dalen, H. P., Henkens, K., & Wang, M. (2015). Recharging or retiring older workers? Uncovering the age-based strategies of European employers. *Gerontologist, 55,* 814–824. http://dx.doi.org/10.1093/geront/gnu048

van de Rijt, A., Kang, S. M., Restivo, M., & Patil, A. (2014). Field experiments of success-breeds-success dynamics. *Proceedings of the National Academy of Sciences of the United States of America, 111,* 6934–6939. http://dx.doi.org/10.1073/pnas.1316836111

Viswesvaran, C., & Ones, D. S. (2000). Perspectives on models of job performance. *International Journal of Selection and Assessment, 8,* 216–226. http://dx.doi.org/10.1111/1468-2389.00151

Wang, M., Kammeyer-Mueller, J., Liu, Y., & Li, Y. (2015). Context, socialization, and newcomer learning. *Organizational Psychology Review, 5,* 3–25. http://dx.doi.org/10.1177/2041386614528832

Wernimont, P. F., & Campbell, J. P. (1968). Signs, samples, and criteria. *Journal of Applied Psychology, 52,* 372–376. http://dx.doi.org/10.1037/h0026244

Whetten, D. A., Felin, T., & King, B. G. (2009). The practice of theory borrowing in organizational studies: Current issues and future directions. *Journal of Management, 35,* 537–563. http://dx.doi.org/10.1177/0149206308330556

Whetzel, D. L., McDaniel, M. A., Yost, A. P., & Kim, N. (2010). Linearity of personality-performance relationships: A large-scale examination. *International Journal of Selection and Assessment, 18,* 310–320. http://dx.doi.org/10.1111/j.1468-2389.2010.00514.x

Yamada, K., Shinmoto, H., Oshio, K., Ito, S., Kumagai, H., & Kaji, T. (2016). Diffusion-weighted MR imaging for the assessment of renal function: Analysis using statistical models based on truncated Gaussian and gamma distributions. *Magnetic Resonance in Medical Sciences, 15,* 237–245. http://dx.doi.org/10.2463/mrms.mp.2015-0067

Zyphur, M. J., Chaturvedi, S., & Arvey, R. D. (2008). Job performance over time is a function of latent trajectories and previous performance. *Journal of Applied Psychology, 93,* 217–224. http://dx.doi.org/10.1037/0021-9010.93.1.217

# Appendix A

## Simulation Study to Investigate the Accuracy of Distribution Pitting and Decision Rules

### Overview

We used R to simulate both discrete and continuous distributions. The purpose of this simulation study was to assess the accuracy of our methodological procedures for classifying distributions (i.e., distribution pitting and the three decision rules). Our simulated distributions were based on the parameter values of the seven generic distributions depicted in Figure 1 (i.e., the pure power law, lognormal, exponential, power law with an exponential cutoff, normal, Poisson, and Weibull distribution). In particular, we simulated each of the seven generic distributions repeatedly because R generates data randomly within the specified parameter value(s). In turn, each simulated distribution included 1,000 observations.

### Method

First, we simulated 20 discrete distributions for each of the seven generic distributions. As a result, we simulated a total of 140 distributions, which we subsequently used to test the accuracy of our methodological procedures for samples containing discrete data. Second, we simulated 20 continuous distributions for each of the generic distributions except the Poisson distribution. Our simulation of continuous distributions did not include Poisson distributions, which can only model discrete data. Thus, we simulated a total of 120 distributions, which we subsequently used to test the accuracy of our methodological procedures for samples containing continuous data. We then applied distribution pitting and the three decision rules to each of the simulated distributions. To check how frequently our methodological procedures may fail to correctly identify the true underlying distribution, we calculated Type 1 and 2 error rates based on the simulated discrete and continuous data, as described next.

### Results and Discussion

A summary of results from our simulated data analyses can be found in Table A1 for discrete data and Table A2 for continuous data. Regarding the discrete data, our methodological procedures led to correctly identifying the dominant distribution 100% of the time for the pure power law, lognormal, exponential, normal, and Weibull distributions. Among the 20 true distributions of the power law with an exponential cutoff, we correctly identified the dominant distribution in 19 instances, while in-

Table A1
*Distribution Identified as Dominant per Simulated Discrete Distribution Using Dpit*

| ID | True distribution | Distribution identified as dominant | ID | True distribution | Distribution identified as dominant |
|---|---|---|---|---|---|
| 1 | Pure power law | Pure power law | 28 | Lognormal | Lognormal |
| 2 | Pure power law | Pure power law | 29 | Lognormal | Lognormal |
| 3 | Pure power law | Pure power law | 30 | Lognormal | Lognormal |
| 4 | Pure power law | Pure power law | 31 | Lognormal | Lognormal |
| 5 | Pure power law | Pure power law | 32 | Lognormal | Lognormal |
| 6 | Pure power law | Pure power law | 33 | Lognormal | Lognormal |
| 7 | Pure power law | Pure power law | 34 | Lognormal | Lognormal |
| 8 | Pure power law | Pure power law | 35 | Lognormal | Lognormal |
| 9 | Pure power law | Pure power law | 36 | Lognormal | Lognormal |
| 10 | Pure power law | Pure power law | 37 | Lognormal | Lognormal |
| 11 | Pure power law | Pure power law | 38 | Lognormal | Lognormal |
| 12 | Pure power law | Pure power law | 39 | Lognormal | Lognormal |
| 13 | Pure power law | Pure power law | 40 | Lognormal | Lognormal |
| 14 | Pure power law | Pure power law | 41 | Exponential | Exponential |
| 15 | Pure power law | Pure power law | 42 | Exponential | Exponential |
| 16 | Pure power law | Pure power law | 43 | Exponential | Exponential |
| 17 | Pure power law | Pure power law | 44 | Exponential | Exponential |
| 18 | Pure power law | Pure power law | 45 | Exponential | Exponential |
| 19 | Pure power law | Pure power law | 46 | Exponential | Exponential |
| 20 | Pure power law | Pure power law | 47 | Exponential | Exponential |
| 21 | Lognormal | Lognormal | 48 | Exponential | Exponential |
| 22 | Lognormal | Lognormal | 49 | Exponential | Exponential |
| 23 | Lognormal | Lognormal | 50 | Exponential | Exponential |
| 24 | Lognormal | Lognormal | 51 | Exponential | Exponential |
| 25 | Lognormal | Lognormal | 52 | Exponential | Exponential |
| 26 | Lognormal | Lognormal | 53 | Exponential | Exponential |
| 27 | Lognormal | Lognormal | 54 | Exponential | Exponential |

(*Appendices continue*)

Table A1 (*continued*)

| ID | True distribution | Distribution identified as dominant | ID | True distribution | Distribution identified as dominant |
|---|---|---|---|---|---|
| 55 | Exponential | Exponential | 98 | Normal | Normal |
| 56 | Exponential | Exponential | 99 | Normal | Normal |
| 57 | Exponential | Exponential | 100 | Normal | Normal |
| 58 | Exponential | Exponential | 101 | Poisson | Poisson |
| 59 | Exponential | Exponential | 102 | Poisson | Undetermined[a,b] |
| 60 | Exponential | Exponential | 103 | Poisson | Poisson |
| 61 | Power law w/ cutoff | Power law w/ cutoff | 104 | Poisson | Poisson |
| 62 | Power law w/ cutoff | Power law w/ cutoff | 105 | Poisson | Normal |
| 63 | Power law w/ cutoff | Power law w/ cutoff | 106 | Poisson | Poisson |
| 64 | Power law w/ cutoff | Power law w/ cutoff | 107 | Poisson | Poisson |
| 65 | Power law w/ cutoff | Power law w/ cutoff | 108 | Poisson | Normal |
| 66 | Power law w/ cutoff | Power law w/ cutoff | 109 | Poisson | Normal |
| 67 | Power law w/ cutoff | Power law w/ cutoff | 110 | Poisson | Normal |
| 68 | Power law w/ cutoff | Pure power law | 111 | Poisson | Poisson |
| 69 | Power law w/ cutoff | Power law w/ cutoff | 112 | Poisson | Poisson |
| 70 | Power law w/ cutoff | Power law w/ cutoff | 113 | Poisson | Poisson |
| 71 | Power law w/ cutoff | Power law w/ cutoff | 114 | Poisson | Poisson |
| 72 | Power law w/ cutoff | Power law w/ cutoff | 115 | Poisson | Normal |
| 73 | Power law w/ cutoff | Power law w/ cutoff | 116 | Poisson | Normal |
| 74 | Power law w/ cutoff | Power law w/ cutoff | 117 | Poisson | Poisson |
| 75 | Power law w/ cutoff | Power law w/ cutoff | 118 | Poisson | Normal |
| 76 | Power law w/ cutoff | Power law w/ cutoff | 119 | Poisson | Poisson |
| 77 | Power law w/ cutoff | Power law w/ cutoff | 120 | Poisson | Poisson |
| 78 | Power law w/ cutoff | Power law w/ cutoff | 121 | Weibull | Weibull |
| 79 | Power law w/ cutoff | Power law w/ cutoff | 122 | Weibull | Weibull |
| 80 | Power law w/ cutoff | Power law w/ cutoff | 123 | Weibull | Weibull |
| 81 | Normal | Normal | 124 | Weibull | Weibull |
| 82 | Normal | Normal | 125 | Weibull | Weibull |
| 83 | Normal | Normal | 126 | Weibull | Weibull |
| 84 | Normal | Normal | 127 | Weibull | Weibull |
| 85 | Normal | Normal | 128 | Weibull | Weibull |
| 86 | Normal | Normal | 129 | Weibull | Weibull |
| 87 | Normal | Normal | 130 | Weibull | Weibull |
| 88 | Normal | Normal | 131 | Weibull | Weibull |
| 89 | Normal | Normal | 132 | Weibull | Weibull |
| 90 | Normal | Normal | 133 | Weibull | Weibull |
| 91 | Normal | Normal | 134 | Weibull | Weibull |
| 92 | Normal | Normal | 135 | Weibull | Weibull |
| 93 | Normal | Normal | 136 | Weibull | Weibull |
| 94 | Normal | Normal | 137 | Weibull | Weibull |
| 95 | Normal | Normal | 138 | Weibull | Weibull |
| 96 | Normal | Normal | 139 | Weibull | Weibull |
| 97 | Normal | Normal | 140 | Weibull | Weibull |

*Note.* Superscripts represent the remaining distributions when the results were undetermined: a = Poisson; b = Weibull; Dpit = distribution pitting methodological approach for identifying the likely dominant distribution, as described in the Method section.



correctly identifying the pure power law distribution as the dominant distribution in 1 instance. Among the 20 true Poisson distributions, we correctly identified the dominant distribution in 12 instances, incorrectly identified the normal distribution as the dominant distribution in 7 instances, and encountered an undetermined finding in 1 instance (i.e., Weibull and Poisson remained).

In total, out of the 140 simulated distributions of discrete data, our procedures correctly identified the dominant distribution in 131 instances (i.e., 93.6% accurate). On the other hand, our procedures led to 8 instances of Type 1 error (i.e., 1 instance of incorrectly identifying the pure power law distribution as the dominant distribution and 7 instances of incorrectly identifying the normal distribution as the dominant distribution) as well as 9 instances of Type 2 error (i.e., 1 instance of incorrectly identifying the pure power law distribution as the dominant distribution, 7 instances of incorrectly identifying the normal distribution as the dominant distribution, and 1 instance of an undetermined finding). In other words, results based on the simulated discrete data showed Type 1 and 2 error rates of 5.7% (= 8/140) and 6.4% (= 9/140), respectively.

(*Appendices continue*)

Table A2
*Distribution Identified as Dominant per Simulated Continuous Distribution Using Dpit*

| ID | True distribution | Distribution identified as dominant | ID | True distribution | Distribution identified as dominant |
|---|---|---|---|---|---|
| 1 | Pure power law | Pure power law | 61 | Power law w/ cutoff | Power law w/ cutoff |
| 2 | Pure power law | Power law w/ cutoff | 62 | Power law w/ cutoff | Power law w/ cutoff |
| 3 | Pure power law | Power law w/ cutoff | 63 | Power law w/ cutoff | Power law w/ cutoff |
| 4 | Pure power law | Power law w/ cutoff | 64 | Power law w/ cutoff | Power law w/ cutoff |
| 5 | Pure power law | Power law w/ cutoff | 65 | Power law w/ cutoff | Power law w/ cutoff |
| 6 | Pure power law | Power law w/ cutoff | 66 | Power law w/ cutoff | Power law w/ cutoff |
| 7 | Pure power law | Power law w/ cutoff | 67 | Power law w/ cutoff | Power law w/ cutoff |
| 8 | Pure power law | Power law w/ cutoff | 68 | Power law w/ cutoff | Power law w/ cutoff |
| 9 | Pure power law | Power law w/ cutoff | 69 | Power law w/ cutoff | Power law w/ cutoff |
| 10 | Pure power law | Power law w/ cutoff | 70 | Power law w/ cutoff | Power law w/ cutoff |
| 11 | Pure power law | Power law w/ cutoff | 71 | Power law w/ cutoff | Power law w/ cutoff |
| 12 | Pure power law | Power law w/ cutoff | 72 | Power law w/ cutoff | Power law w/ cutoff |
| 13 | Pure power law | Power law w/ cutoff | 73 | Power law w/ cutoff | Power law w/ cutoff |
| 14 | Pure power law | Power law w/ cutoff | 74 | Power law w/ cutoff | Power law w/ cutoff |
| 15 | Pure power law | Power law w/ cutoff | 75 | Power law w/ cutoff | Power law w/ cutoff |
| 16 | Pure power law | Power law w/ cutoff | 76 | Power law w/ cutoff | Power law w/ cutoff |
| 17 | Pure power law | Power law w/ cutoff | 77 | Power law w/ cutoff | Power law w/ cutoff |
| 18 | Pure power law | Power law w/ cutoff | 78 | Power law w/ cutoff | Power law w/ cutoff |
| 19 | Pure power law | Power law w/ cutoff | 79 | Power law w/ cutoff | Power law w/ cutoff |
| 20 | Pure power law | Power law w/ cutoff | 80 | Power law w/ cutoff | Power law w/ cutoff |
| 21 | Lognormal | Lognormal | 81 | Normal | Normal |
| 22 | Lognormal | Lognormal | 82 | Normal | Lognormal |
| 23 | Lognormal | Lognormal | 83 | Normal | Normal |
| 24 | Lognormal | Lognormal | 84 | Normal | Normal |
| 25 | Lognormal | Lognormal | 85 | Normal | Normal |
| 26 | Lognormal | Lognormal | 86 | Normal | Normal |
| 27 | Lognormal | Lognormal | 87 | Normal | Normal |
| 28 | Lognormal | Lognormal | 88 | Normal | Normal |
| 29 | Lognormal | Lognormal | 89 | Normal | Lognormal |
| 30 | Lognormal | Lognormal | 90 | Normal | Normal |
| 31 | Lognormal | Lognormal | 91 | Normal | Normal |
| 32 | Lognormal | Lognormal | 92 | Normal | Normal |
| 33 | Lognormal | Lognormal | 93 | Normal | Normal |
| 34 | Lognormal | Lognormal | 94 | Normal | Normal |
| 35 | Lognormal | Lognormal | 95 | Normal | Normal |
| 36 | Lognormal | Lognormal | 96 | Normal | Normal |
| 37 | Lognormal | Lognormal | 97 | Normal | Normal |
| 38 | Lognormal | Lognormal | 98 | Normal | Normal |
| 39 | Lognormal | Lognormal | 99 | Normal | Power law w/ cutoff |
| 40 | Lognormal | Lognormal | 100 | Normal | Normal |
| 41 | Exponential | Power law w/ cutoff | 101 | Weibull | Weibull |
| 42 | Exponential | Power law w/ cutoff | 102 | Weibull | Weibull |
| 43 | Exponential | Power law w/ cutoff | 103 | Weibull | Weibull |
| 44 | Exponential | Power law w/ cutoff | 104 | Weibull | Weibull |
| 45 | Exponential | Power law w/ cutoff | 105 | Weibull | Weibull |
| 46 | Exponential | Exponential | 106 | Weibull | Weibull |
| 47 | Exponential | Exponential | 107 | Weibull | Weibull |
| 48 | Exponential | Exponential | 108 | Weibull | Weibull |
| 49 | Exponential | Exponential | 109 | Weibull | Weibull |
| 50 | Exponential | Exponential | 110 | Weibull | Weibull |
| 51 | Exponential | Power law w/ cutoff | 111 | Weibull | Weibull |
| 52 | Exponential | Power law w/ cutoff | 112 | Weibull | Weibull |
| 53 | Exponential | Power law w/ cutoff | 113 | Weibull | Weibull |
| 54 | Exponential | Exponential | 114 | Weibull | Weibull |
| 55 | Exponential | Exponential | 115 | Weibull | Weibull |
| 56 | Exponential | Power law w/ cutoff | 116 | Weibull | Weibull |
| 57 | Exponential | Power law w/ cutoff | 117 | Weibull | Weibull |
| 58 | Exponential | Power law w/ cutoff | 118 | Weibull | Weibull |
| 59 | Exponential | Exponential | 119 | Weibull | Weibull |
| 60 | Exponential | Exponential | 120 | Weibull | Weibull |

*Note.* Our analyses on the simulated distributions of continuous data did not result in any instance of undetermined finding. Dpit = distribution pitting methodological approach for identifying the likely dominant distribution, as described in the Method section.

(*Appendices continue*)

Regarding the continuous data, our methodological procedures led to correctly identifying the dominant distribution 100% of the time for the pure power law, lognormal, power law with an exponential cutoff, and Weibull distributions. Among the 20 true exponential distributions, we correctly identified the dominant distribution in 9 instances, while incorrectly identifying the power law with an exponential cutoff as the dominant distribution in 11 instances. Among the 20 true normal distributions, we correctly identified the dominant distribution in 17 instances, incorrectly identified the lognormal distribution as the dominant distribution in 2 instances, and incorrectly identified the power law with an exponential cutoff as the dominant distribution in 1 instance. There were no instances of undetermined findings ("undetermined" refers to samples where multiple distributions remained even after implementing all three decision rules).

In total, out of the 120 simulated distributions of continuous data, our procedures correctly identified the dominant distribution in 106 instances (i.e., 88.3% accurate). On the other hand, our procedures led to 14 instances of Type 1 or 2 error (i.e., 12 instances of incorrectly identifying the power law with an exponential cutoff as the dominant distribution and 2 instances of incorrectly identifying the lognormal distribution as the dominant distribution). That is, results based on the simulated continuous data showed a Type 1 or 2 error rate of 11.7%. Across both discrete and continuous data, our procedures correctly identified the dominant distribution 91.2% of the time (= [131 + 106]/260), while Type 1 and 2 error rates were 8.5% (= [8 + 14]/260) and 8.8% (= [9 + 14]/260), respectively.

Results indicated that accurate decisions were overwhelmingly more frequent compared to false positive and false negative decisions. Our methodological procedures are even more accurate when results are derived based on how well our procedures identify the correct distribution category. Out of the 140 simulated distributions of discrete data, our procedures correctly identified the dominant distribution category in 139 instances (i.e., 99.3% accurate). Further, our procedures only led to 1 instance of Type 1 or 2 error (i.e., 1 instance of incorrectly identifying the pure power law distribution as the dominant distribution category). Out of the 120 simulated distributions of continuous data, our procedures correctly identified the dominant distribution category in 117 instances (i.e., 97.5% accurate). Moreover, our procedures only led to 3 instances of Type 1 or 2 error (i.e., 2 instances of incorrectly identifying the lognormal distribution as the dominant distribution category and 1 instance of incorrectly identifying exponential tail distributions as the dominant distribution category). In short, across both discrete and continuous data, our procedures correctly identified the dominant distribution category 98.5% of the time (= [139 + 117]/260), while the Type 1 or 2 error rate was only 1.5% (= [1 + 3]/260).

# Appendix B

## Follow-Up Study to Investigate the Necessity of the Third Decision Rule

### Overview and Method

The purpose of the follow-up study was to examine whether using the third decision rule for identifying the likely dominant distribution improves the accuracy of conclusions. We used the same 140 distributions of discrete data and also 120 distributions of continuous data simulated in Appendix A (Simulation Study to Investigate the Accuracy of Distribution Pitting and Decision Rules). However, unlike the study described in Appendix A, this follow-up study only applied the first two decision rules to check the extent to which the absence of the third decision rule reduces the accuracy of results.

### Results and Discussion

A summary of results is in Table B1 for discrete data and Table B2 for continuous data. Regarding the discrete data, using the first two decision rules while not the third decision rule led to correctly identifying the dominant distribution 100% of the time for the lognormal and Weibull distributions. However, among the 20 true distributions of the pure power law, we correctly identified the dominant distribution in 1 instance. Among the 20 true exponential distributions, we correctly identified the dominant distribution in 8 instances. Among the 20 true distributions of the power law with an exponential cutoff, we correctly identified the dominant distribution in 6 instances. Among the 20 true normal distributions, we correctly identified the dominant distribution in 0 instance. Among the 20 true Poisson distributions, we correctly identified the dominant distribution in 12 instances. In total, out of the 140 simulated distributions of discrete data, using the first two decision rules while not the third decision rule only led to correctly identifying the dominant distribution in 67 instances (i.e., 47.9% accurate). This is in stark contrast to the results in Appendix A, where we used all three decision rules and correctly identified the dominant distribution in 131 instances out of the 140 simulated distributions of discrete data (i.e., 93.6% accurate).

(*Appendices continue*)

Table B1
*Distribution Identified as Dominant per Simulated Discrete Distribution Using Dpit's Decision Rules #1 and #2 and Not Using Decision Rule #3*

| ID | True distribution | Distribution identified as dominant | ID | True distribution | Distribution identified as dominant |
|---|---|---|---|---|---|
| 1 | Pure power law | Undetermined[a,b] | 59 | Exponential | Exponential |
| 2 | Pure power law | Undetermined[a,b,f] | 60 | Exponential | Exponential |
| 3 | Pure power law | Undetermined[a,b,f] | 61 | Power law w/ cutoff | Undetermined[b,d,g] |
| 4 | Pure power law | Undetermined[a,b,f] | 62 | Power law w/ cutoff | Undetermined[b,d] |
| 5 | Pure power law | Undetermined[a,b,f] | 63 | Power law w/ cutoff | Undetermined[b,d,g] |
| 6 | Pure power law | Undetermined[a,b] | 64 | Power law w/ cutoff | Power law w/ cutoff |
| 7 | Pure power law | Undetermined[a,b] | 65 | Power law w/ cutoff | Undetermined[b,d,g] |
| 8 | Pure power law | Undetermined[a,b] | 66 | Power law w/ cutoff | Power law w/ cutoff |
| 9 | Pure power law | Undetermined[a,b,f] | 67 | Power law w/ cutoff | Undetermined[b,d,g] |
| 10 | Pure power law | Undetermined[a,b,f] | 68 | Power law w/ cutoff | Undetermined[a,b,g] |
| 11 | Pure power law | Undetermined[a,b,f] | 69 | Power law w/ cutoff | Undetermined[b,d,g] |
| 12 | Pure power law | Pure power law | 70 | Power law w/ cutoff | Undetermined[b,d,g] |
| 13 | Pure power law | Undetermined[a,b,f] | 71 | Power law w/ cutoff | Undetermined[b,d,g] |
| 14 | Pure power law | Undetermined[a,b,f] | 72 | Power law w/ cutoff | Power law w/ cutoff |
| 15 | Pure power law | Undetermined[a,b,f] | 73 | Power law w/ cutoff | Undetermined[b,d,g] |
| 16 | Pure power law | Undetermined[a,b,f] | 74 | Power law w/ cutoff | Power law w/ cutoff |
| 17 | Pure power law | Undetermined[a,b] | 75 | Power law w/ cutoff | Undetermined[b,d] |
| 18 | Pure power law | Undetermined[a,b,f] | 76 | Power law w/ cutoff | Undetermined[b,d,g] |
| 19 | Pure power law | Undetermined[a,b] | 77 | Power law w/ cutoff | Undetermined[b,d,g] |
| 20 | Pure power law | Undetermined[a,b,f] | 78 | Power law w/ cutoff | Power law w/ cutoff |
| 21 | Lognormal | Lognormal | 79 | Power law w/ cutoff | Power law w/ cutoff |
| 22 | Lognormal | Lognormal | 80 | Power law w/ cutoff | Undetermined[d,g] |
| 23 | Lognormal | Lognormal | 81 | Normal | Undetermined[b,e] |
| 24 | Lognormal | Lognormal | 82 | Normal | Undetermined[b,e] |
| 25 | Lognormal | Lognormal | 83 | Normal | Undetermined[b,e] |
| 26 | Lognormal | Lognormal | 84 | Normal | Undetermined[b,e] |
| 27 | Lognormal | Lognormal | 85 | Normal | Undetermined[b,e] |
| 28 | Lognormal | Lognormal | 86 | Normal | Undetermined[b,e] |
| 29 | Lognormal | Lognormal | 87 | Normal | Undetermined[b,e] |
| 30 | Lognormal | Lognormal | 88 | Normal | Undetermined[b,e] |
| 31 | Lognormal | Lognormal | 89 | Normal | Undetermined[b,e] |
| 32 | Lognormal | Lognormal | 90 | Normal | Undetermined[b,e] |
| 33 | Lognormal | Lognormal | 91 | Normal | Undetermined[b,e] |
| 34 | Lognormal | Lognormal | 92 | Normal | Undetermined[b,e] |
| 35 | Lognormal | Lognormal | 93 | Normal | Undetermined[b,e] |
| 36 | Lognormal | Lognormal | 94 | Normal | Undetermined[b,e] |
| 37 | Lognormal | Lognormal | 95 | Normal | Undetermined[b,e] |
| 38 | Lognormal | Lognormal | 96 | Normal | Undetermined[b,e] |
| 39 | Lognormal | Lognormal | 97 | Normal | Undetermined[b,e] |
| 40 | Lognormal | Lognormal | 98 | Normal | Undetermined[b,e] |
| 41 | Exponential | Undetermined[b,c] | 99 | Normal | Undetermined[b,e] |
| 42 | Exponential | Undetermined[b,c] | 100 | Normal | Undetermined[b,e] |
| 43 | Exponential | Exponential | 101 | Poisson | Poisson |
| 44 | Exponential | Undetermined[b,c] | 102 | Poisson | Undetermined[f,g] |
| 45 | Exponential | Undetermined[b,c] | 103 | Poisson | Poisson |
| 46 | Exponential | Exponential | 104 | Poisson | Poisson |
| 47 | Exponential | Undetermined[b,c] | 105 | Poisson | Undetermined[e,f] |
| 48 | Exponential | Undetermined[b,c] | 106 | Poisson | Poisson |
| 49 | Exponential | Undetermined[b,c] | 107 | Poisson | Poisson |
| 50 | Exponential | Exponential | 108 | Poisson | Undetermined[e,f,g] |
| 51 | Exponential | Exponential | 109 | Poisson | Undetermined[e,f,g] |
| 52 | Exponential | Undetermined[b,c] | 110 | Poisson | Undetermined[e,f] |
| 53 | Exponential | Undetermined[b,c] | 111 | Poisson | Poisson |
| 54 | Exponential | Undetermined[b,c] | 112 | Poisson | Poisson |
| 55 | Exponential | Undetermined[b,c] | 113 | Poisson | Poisson |
| 56 | Exponential | Undetermined[b,c] | 114 | Poisson | Poisson |
| 57 | Exponential | Exponential | 115 | Poisson | Undetermined[e,f,g] |
| 58 | Exponential | Exponential | 116 | Poisson | Undetermined[e,f,g] |

(*Appendices continue*)

Table B1 (*continued*)

| ID | True distribution | Distribution identified as dominant | ID | True distribution | Distribution identified as dominant |
|---|---|---|---|---|---|
| 117 | Poisson | Poisson | 129 | Weibull | Weibull |
| 118 | Poisson | Undetermined[e,f,g] | 130 | Weibull | Weibull |
| 119 | Poisson | Poisson | 131 | Weibull | Weibull |
| 120 | Poisson | Poisson | 132 | Weibull | Weibull |
| 121 | Weibull | Weibull | 133 | Weibull | Weibull |
| 122 | Weibull | Weibull | 134 | Weibull | Weibull |
| 123 | Weibull | Weibull | 135 | Weibull | Weibull |
| 124 | Weibull | Weibull | 136 | Weibull | Weibull |
| 125 | Weibull | Weibull | 137 | Weibull | Weibull |
| 126 | Weibull | Weibull | 138 | Weibull | Weibull |
| 127 | Weibull | Weibull | 139 | Weibull | Weibull |
| 128 | Weibull | Weibull | 140 | Weibull | Weibull |

*Note*. Superscripts represent the remaining distributions when the results were undetermined: a = pure power law; b = lognormal; c = exponential; d = power law w/ cutoff; e = normal; f = Poisson; g = Weibull. Dpit = distribution pitting methodological approach for identifying the likely dominant distribution, as described in the Method section.

Regarding the continuous data, using the first two decision rules while not the third decision rule led to correctly identifying the dominant distribution 100% of the time for the pure power law, lognormal, power law with an exponential cutoff, and Weibull distributions. However, among the 20 true exponential distributions, we correctly identified the dominant distribution in 1 instance. Among the 20 true normal distributions, we correctly identified the dominant distribution in 3 instances. In total, out of the 120 simulated distributions of continuous data, using the first two decision rules while not the third decision rule only led to correctly identifying the dominant distribution in 84 instances (i.e., 70% accurate). This is in contrast to the results in Appendix A,

Table B2

*Distribution Identified as Dominant per Simulated Continuous Distribution Using Dpit's Decision Rules #1 and #2 and Not Using Decision Rule #3*

| ID | True distribution | Distribution identified as dominant | ID | True distribution | Distribution identified as dominant |
|---|---|---|---|---|---|
| 1 | Pure power law | Pure power law | 26 | Lognormal | Lognormal |
| 2 | Pure power law | Pure power law | 27 | Lognormal | Lognormal |
| 3 | Pure power law | Pure power law | 28 | Lognormal | Lognormal |
| 4 | Pure power law | Pure power law | 29 | Lognormal | Lognormal |
| 5 | Pure power law | Pure power law | 30 | Lognormal | Lognormal |
| 6 | Pure power law | Pure power law | 31 | Lognormal | Lognormal |
| 7 | Pure power law | Pure power law | 32 | Lognormal | Lognormal |
| 8 | Pure power law | Pure power law | 33 | Lognormal | Lognormal |
| 9 | Pure power law | Pure power law | 34 | Lognormal | Lognormal |
| 10 | Pure power law | Pure power law | 35 | Lognormal | Lognormal |
| 11 | Pure power law | Pure power law | 36 | Lognormal | Lognormal |
| 12 | Pure power law | Pure power law | 37 | Lognormal | Lognormal |
| 13 | Pure power law | Pure power law | 38 | Lognormal | Lognormal |
| 14 | Pure power law | Pure power law | 39 | Lognormal | Lognormal |
| 15 | Pure power law | Pure power law | 40 | Lognormal | Lognormal |
| 16 | Pure power law | Pure power law | 41 | Exponential | Power law w/ cutoff |
| 17 | Pure power law | Pure power law | 42 | Exponential | Power law w/ cutoff |
| 18 | Pure power law | Pure power law | 43 | Exponential | Power law w/ cutoff |
| 19 | Pure power law | Pure power law | 44 | Exponential | Power law w/ cutoff |
| 20 | Pure power law | Pure power law | 45 | Exponential | Power law w/ cutoff |
| 21 | Lognormal | Lognormal | 46 | Exponential | Undetermined[c,f] |
| 22 | Lognormal | Lognormal | 47 | Exponential | Undetermined[c,f] |
| 23 | Lognormal | Lognormal | 48 | Exponential | Exponential |
| 24 | Lognormal | Lognormal | 49 | Exponential | Undetermined[c,f] |
| 25 | Lognormal | Lognormal | 50 | Exponential | Undetermined[c,f] |

(*Appendices continue*)

Table B2 (*continued*)

| ID | True distribution | Distribution identified as dominant | ID | True distribution | Distribution identified as dominant |
|---|---|---|---|---|---|
| 51 | Exponential | Power law w/ cutoff | 86 | Normal | Undetermined[b,e] |
| 52 | Exponential | Power law w/ cutoff | 87 | Normal | Undetermined[b,e] |
| 53 | Exponential | Power law w/ cutoff | 88 | Normal | Undetermined[b,e] |
| 54 | Exponential | Undetermined[c,f] | 89 | Normal | Lognormal |
| 55 | Exponential | Undetermined[c,f] | 90 | Normal | Normal |
| 56 | Exponential | Power law w/ cutoff | 91 | Normal | Undetermined[b,e] |
| 57 | Exponential | Power law w/ cutoff | 92 | Normal | Undetermined[b,e] |
| 58 | Exponential | Power law w/ cutoff | 93 | Normal | Undetermined[b,e] |
| 59 | Exponential | Undetermined[c,f] | 94 | Normal | Undetermined[b,e] |
| 60 | Exponential | Undetermined[c,f] | 95 | Normal | Undetermined[b,e] |
| 61 | Power law w/ cutoff | Power law w/ cutoff | 96 | Normal | Undetermined[b,e] |
| 62 | Power law w/ cutoff | Power law w/ cutoff | 97 | Normal | Undetermined[b,e] |
| 63 | Power law w/ cutoff | Power law w/ cutoff | 98 | Normal | Undetermined[b,e] |
| 64 | Power law w/ cutoff | Power law w/ cutoff | 99 | Normal | Undetermined[b,d] |
| 65 | Power law w/ cutoff | Power law w/ cutoff | 100 | Normal | Undetermined[b,e] |
| 66 | Power law w/ cutoff | Power law w/ cutoff | 101 | Weibull | Weibull |
| 67 | Power law w/ cutoff | Power law w/ cutoff | 102 | Weibull | Weibull |
| 68 | Power law w/ cutoff | Power law w/ cutoff | 103 | Weibull | Weibull |
| 69 | Power law w/ cutoff | Power law w/ cutoff | 104 | Weibull | Weibull |
| 70 | Power law w/ cutoff | Power law w/ cutoff | 105 | Weibull | Weibull |
| 71 | Power law w/ cutoff | Power law w/ cutoff | 106 | Weibull | Weibull |
| 72 | Power law w/ cutoff | Power law w/ cutoff | 107 | Weibull | Weibull |
| 73 | Power law w/ cutoff | Power law w/ cutoff | 108 | Weibull | Weibull |
| 74 | Power law w/ cutoff | Power law w/ cutoff | 109 | Weibull | Weibull |
| 75 | Power law w/ cutoff | Power law w/ cutoff | 110 | Weibull | Weibull |
| 76 | Power law w/ cutoff | Power law w/ cutoff | 111 | Weibull | Weibull |
| 77 | Power law w/ cutoff | Power law w/ cutoff | 112 | Weibull | Weibull |
| 78 | Power law w/ cutoff | Power law w/ cutoff | 113 | Weibull | Weibull |
| 79 | Power law w/ cutoff | Power law w/ cutoff | 114 | Weibull | Weibull |
| 80 | Power law w/ cutoff | Power law w/ cutoff | 115 | Weibull | Weibull |
| 81 | Normal | Undetermined[b,e] | 116 | Weibull | Weibull |
| 82 | Normal | Lognormal | 117 | Weibull | Weibull |
| 83 | Normal | Undetermined[b,e] | 118 | Weibull | Weibull |
| 84 | Normal | Normal | 119 | Weibull | Weibull |
| 85 | Normal | Normal | 120 | Weibull | Weibull |

*Note*. Superscripts represent the remaining distributions when the results were undetermined: a = pure power law; b = lognormal; c = exponential; d = power law w/ cutoff; e = normal; f = Weibull; Dpit = distribution pitting methodological approach for identifying the likely dominant distribution, as described in the Method section.

where we used all three decision rules and correctly identified the dominant distribution in 106 instances out of 120 simulated distributions of continuous data (i.e., 88.3% accurate).

Overall, results indicated that using decision rules #1 and #2 while not using the third decision rule reduces the accuracy of conclusions regarding the likely dominant distribution. Across both discrete and continuous data, using the first two decision rules while not the third decision rule led to correctly identifying the dominant distribution 58.1% of the time (= [67 + 84]/260). This accuracy rate is much lower than the accuracy rate of 91.2%, which we obtained in the study described in Appendix A by using all three decision rules.

Received July 4, 2016
Revision received January 30, 2017
Accepted January 30, 2017 ■