PERSONNEL PSYCHOLOGY
THE STUDY OF PEOPLE AT WORK

PERSONNEL PSYCHOLOGY
2012, 65, 79–119

THE BEST AND THE REST: REVISITING THE NORM OF NORMALITY OF INDIVIDUAL PERFORMANCE

ERNEST O'BOYLE JR.
Department of Management
College of Business and Economics
Longwood University

HERMAN AGUINIS
Department of Management and Entrepreneurship
Kelley School of Business
Indiana University

We revisit a long-held assumption in human resource management, organizational behavior, and industrial and organizational psychology that individual performance follows a Gaussian (normal) distribution. We conducted 5 studies involving 198 samples including 633,263 researchers, entertainers, politicians, and amateur and professional athletes. Results are remarkably consistent across industries, types of jobs, types of performance measures, and time frames and indicate that individual performance is not normally distributed—instead, it follows a Paretian (power law) distribution. Assuming normality of individual performance can lead to misspecified theories and misleading practices. Thus, our results have implications for all theories and applications that directly or indirectly address the performance of individual workers including performance measurement and management, utility analysis in preemployment testing and training and development, personnel selection, leadership, and the prediction of performance, among others.

Research and practice in organizational behavior and human resource management (OBHRM), industrial and organizational (I-O) psychology, and other fields including strategic management and entrepreneurship ultimately build upon, directly or indirectly, the output of the individual worker. In fact, a central goal of OBHRM is to understand and predict the performance of individual workers. There is a long-held assumption in OBHRM that individual performance clusters around a mean and then fans out into symmetrical tails. That is, individual performance is assumed to follow a normal distribution (Hull, 1928; Schmidt & Hunter, 1983;

A previous version of this paper was presented at the Annual Meetings of the Academy of Management, San Antonio, Texas, August 2011. We thank Wayne L. Winston for his insightful comments regarding data analysis issues.

Correspondence and requests for reprints should be addressed to Herman Aguinis, Department of Management and Entrepreneurship, Kelley School of Business, Indiana University, 1309 E. 10th Street, Suite 630D, Bloomington, IN 47405-1701; haguinis@indiana.edu.

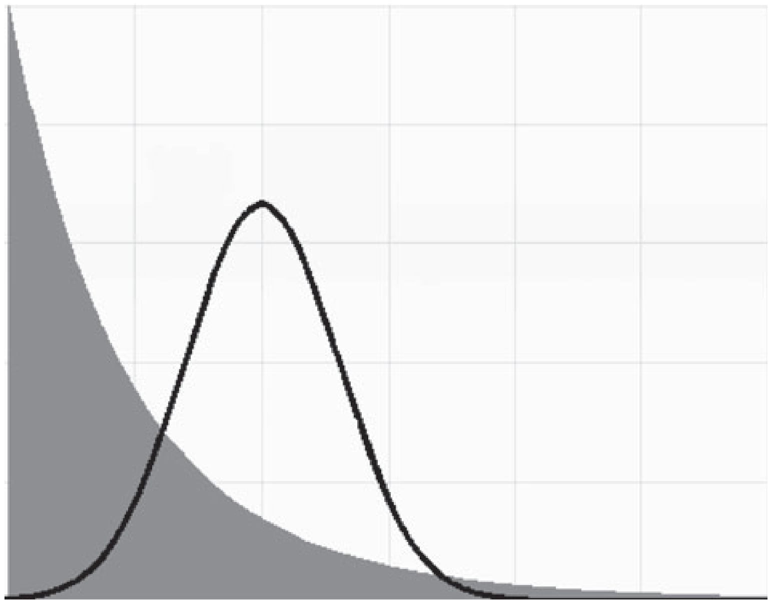

Figure 1: **A Normal Distribution (Black) Overlaying a Paretian Distribution (Grey).**

Tiffin, 1947). When performance data do not conform to the normal distribution, then the conclusion is that the error "must" lie within the sample not the population. Subsequent adjustments are made (e.g., dropping outliers) in order to make the sample "better reflect" the "true" underlying normal curve. Gaussian distributions are in stark contrast to Paretian or power law distributions, which are typified by unstable means, infinite variance, and a greater proportion of extreme events. Figure 1 shows a Paretian distribution overlaid with a normal curve.

The goal of our research is to revisit the norm of normality of individual performance and discuss implications for OBHRM theory and research; methodology; and practice, policy making, and society. Our manuscript is organized as follows. First, we describe the origins and document the presence of the norm of normality regarding individual performance. Second, we discuss the Gaussian (i.e., normal) and Paretian (i.e., power law) distributions and key differences between them. Third, we describe five separate studies involving 198 samples including 633,263 researchers, entertainers, politicians, and amateur and professional athletes. Results of each of these five studies are remarkably consistent and indicate that individual performance does not follow a normal distribution and, instead, it follows a power law distribution. Finally, we discuss implications of our results, including directions for future research.

The Norm of Normality of Individual Performance

The normal distribution has been used to model a variety of phenomena including human traits such as height (Yule, 1912) and intelligence (Galton, 1889), as well as probability distributions (Hull, 1928), economic trends such as stock pricing (Bronzin, 1908), and the laws of thermodynamics (Reif, 1965). Based on the normal distribution's prevalence across scientific disciplines and phenomena, it has seemed reasonable to assume that normality would also be the distribution of individual performance.

Although the assumption of individual performance normality is common across most research domains in OBHRM, it seems to have originated in the performance appraisal literature. More than half a century ago, Ferguson (1947) noted that "ratings for a large and representative group of assistant managers should be distributed in accordance with the percentages predicted for a normal distribution" (p. 308). The normality assumption persisted through the years, and researchers began to not only assume job performance normality but forced it upon the observed distributions regardless of the actual observed distributional properties. For example, in developing a performance appraisal system, Canter (1953) used "a forced normal distribution of judgments" (p. 456) for evaluating open-ended responses. Likewise, Schultz and Siegel (1961) "forced the [performance] rater to respond on a seven-point scale and to normalize approximately the distribution of his responses" (p. 138). Thus, if a supervisor rated the performance of her subordinates and placed most of them into a single category while placing only a small minority in the top ranking, it was assumed that there was a severity bias in need of a correction to normality (Motowidlo & Borman, 1977; Schneier, 1977). Moreover, the advice is that if an employee contributes a disproportionate amount of sales in a firm, he should be dropped from the data set or have his sales statistically adjusted to a more "reasonable" value (e.g., three standard deviations within the mean) before moving forward with a traditional analysis that assumes an underlying normal distribution. Both design practices (i.e., forced-response formats) and statistical analyses (i.e., deletion or "correction" of outliers) in performance evaluation create a normal distribution in samples regardless of the shape of the underlying population distributions.

We readily acknowledge that some researchers and practitioners may not believe that individual performance is normally distributed (e.g., Bernardin & Beatty, 1984; Micceri, 1989; Murphy & Cleveland, 1995;

Saal, Downey, & Lahey, 1980; Schmidt & Johnson, 1973). However, the normality assumption is a convenient way of studying individual performance—just like economists also make assumptions so that their theoretical models can be simplified. As noted by an anonymous reviewer, some may not put too much thought into the shape of the performance distribution whereas others may believe that, with a sufficiently large number of cases, individual performance is normally distributed. Regardless of the actual beliefs, researchers and practitioners assume performance is normally distributed and alter the distribution of scores through the design, training, and analysis of raters' judgments. Specifically, when performance scores deviate from normality, the cause is attributed to leniency bias, severity bias, and/or a halo error (Aguinis, 2009; Schneier, 1977). Rating systems where most employees occupy the same category with only a few at the highest category are assumed to be indicative of range restriction and other "statistical artifacts" (Motowidlo & Borman, 1977). In fact, Reilly and Smither (1985) provided an extensive critique of individual performance research that violates the normality assumption and provided guidance on how to reestablish the normal and presumably correct distribution of performance.

The norm of normality of individual performance is also evident in many other research domains in OBHRM. Consider the case of personnel selection that, similar to the prediction of performance and performance measurement/work outcomes category, is among the top five most popular research domains based on articles published in *Journal of Applied Psychology* and *Personnel Psychology* over the past 45 years (Cascio & Aguinis, 2008a). Utility analysis has allowed researchers and practitioners to establish the financial value added of implementing valid personnel selection procedures, and all utility analysis approaches operate under the normality assumption. For example, Schmidt, Hunter, McKenzie, and Muldrow's (1979) linear homoscedastic model of work productivity "includes the following three assumptions: (a) linearity, (b) equality of variances of conditional distributions, and (c) normality of conditional distributions" (p. 615). In the same manner, Cascio and Ramos (1986) stated that "[a]ssuming a normal distribution of performance, 55% equates to a Fisher *z*-value of .13, which translates back to a validity coefficient of .13 for the old selection procedure" (p. 25). More recently, Sackett and Yang (2000) concluded that "[o]n the basis of the ordinate of the normal curve at the point of selection, it is possible to infer the mean and variance of the unrestricted distribution. Clearly, the use of such approaches relies heavily on the normality assumption" (p. 115). The validity and accuracy of all utility analyses that rely on the assumed normality of individual performance would be put into question if this assumption is actually not tenable.

Summary

Although some have argued that performance may not be normally distributed (Bernardin & Beatty, 1984; Murphy & Cleveland, 1995; Saal et al., 1980; Schmidt & Johnson, 1973), theory and application regarding individual performance are built around the assumption of normality. For example, theories on performance targeting the average worker, the implementation of performance appraisal systems that include dropping outliers, and the choice of analytic techniques point to an underlying assumption of normality. Moreover, we are not aware of influential theoretical developments or applications that explicitly assume that performance follows a nonnormal distribution (Aguinis, 2009; Smither & London, 2009b). So, although some may not believe in the normal distribution of individual performance, there is little evidence that this belief has affected theoretical developments or application in an influential way. Possible reasons for why skepticism about normality has not affected theory and practice may be the lack of empirical evidence negating the normal distribution and the lack of proposed alternative distributions. But, what if performance does not conform to the Gaussian distribution? What if the means are unstable and the variance of these distributions infinite? Quite simply, if performance is not normally distributed, theories that directly or indirectly build upon individual job performance and its prediction may need to be revisited. In addition, popular practices (e.g., utility analysis of preemployment tests and training and development interventions), which also rely on the assumption of individual performance normality, would also need to be revisited.

Next, we provide an alternative perspective, which draws mainly from the economics, mathematics, and statistics literatures, that challenges the norm of normality and posits that individual performance conforms to a power law or Paretian distribution, which is typified by an unstable mean, infinite variance, and a greater number of extreme events (West & Deering, 1995).

The Paretian Distribution and Individual Performance

The possibility of a nonnormal performance distribution of individual performance has been proposed in the past but the normality assumption has remained the dominant approach. Jacobs (1974) argued that in sales industries (automotive, insurance, stock) performance is not normal because a small group of incumbents who possess the expertise and salesmanship dominate activity. If performance output does not conform to a bell-shaped, normal distribution, then power law distributions may apply (West & Deering, 1995). Power laws such as Pareto's

(1897) Law produce fatter tails than those seen in a normal curve. Stated differently, Paretian probability distributions allow more extreme values to be present (see Figure 1). Whereas a value exceeding three standard deviations from the mean is often thought to be an outlier in the context of a normal curve (e.g., Orr, Sackett, & Dubois, 1991), a Paretian distribution would predict that these values are far more common and that their elimination or transformation is a questionable practice. Paretian distributions are sometimes referred to as the 80/20 principle, which has been shown to apply to many contexts and research domains outside of OBHRM. For example, marketing researchers have reported that about 80% of a brand's volume is purchased by about 20% of its buyers (Anschuetz, 1997) and sociology researchers have reported that about 80% of land is owned by about 20% of the population (Pareto, 1897).

There are important differences between Gaussian and Paretian distributions. First, Gaussian distributions underpredict the likelihood of extreme events. For instance, when stock market performance is predicted using the normal curve, a single-day 10% drop in the financial markets should occur once every 500 years (Buchanan, 2004). In reality, it occurs about once every 5 years (Mandelbrot, Hudson, & Grunwald, 2005). Second, Gaussian distributions assume that the mean and standard deviation, so central to tests of statistical significance and computation of effect sizes, are stable. However, if the underlying distribution is Paretian instead of normal, means and standard deviations are not stable and Gaussian-based point estimates as well as confidence intervals are biased (Andriani & McKelvey, 2009). Third, a key difference between normal and Paretian distributions is scale invariance. In OBHRM, scale invariance usually refers to the extent to which a measurement instrument generalizes across different cultures or populations. A less common operationalization of the concept of scale invariance refers to isomorphism in the shape of score distributions regardless of whether one is examining an individual, a small work group, a department, an organization, or all organizations (Fiol, O'Connor, & Aguinis, 2001). Scale invariance also refers to the distribution remaining constant whether one is looking at the whole distribution or only the top performers. For example, the shape of the wealth distribution is the same whether examining the entire population or just the top 10% of wealthy individuals (Gabaix, 1999). Related to the issue of scale invariance, Gabaix, Gopikrishnan, Plerou, and Stanley (2003) investigated financial market fluctuations across multiple time points and markets and found that data conformed to a power law distribution. The same distribution shape was found in both United States (U.S.) and French markets, and the power law correctly predicted both the crashes of 1929 and 1987.

Germane to OBHRM in particular is that if performance operates under power laws, then the distribution should be the same regardless of the level of analysis. That is, the distribution of individual performance should closely mimic the distribution of firm performance. Researchers who study performance at the firm level of analysis do not necessarily assume that the underlying distribution is normal (e.g., Stanley et al., 1995). However, as noted earlier, researchers who study performance at the individual level of analysis do follow the norm of normality in their theoretical development, research design, and choices regarding data analysis. These conflicting views, which may be indicative of a micro–macro divide in OBHRM and related fields (e.g., Aguinis, Boyd, Pierce, & Short, 2011), could be reconciled if individual performance is found to also follow a power law distribution, as it is the case for firm performance (Bonardi, 2004; Powell, 2003; Stanley et al., 1995).

The Present Studies

We conducted five separate studies to determine whether the distribution of individual performance more closely follows a Paretian curve than a Gaussian curve. In all studies, the primary hypothesis was that the distribution of performance is better modeled with a Paretian curve than a normal curve. For each of the five studies, we used the chi-square (χ^2) statistic to determine whether individual performance more closely follows a Paretian versus a Gaussian distribution. The chi-square is a "badness of fit" statistic because higher values indicate worse fit (Aguinis & Harden, 2009). That is, the greater the degree of divergence of an empirically derived performance distribution from a Gaussian or Paretian distribution, the higher the chi-square. Accordingly, for each of the samples we studied we first forced the data to conform to a normal distribution and then forced the same data to conform to a Paretian distribution. For each comparison, a smaller chi-square value indicates which of the two theoretical distributions describes the data better. To calculate the chi-square for each distribution, we used Decision Tools Suite add-on @Risk 5.5 (Palisades Corporation, 2009). This program operates within Microsoft Excel and provides estimates of fit for a variety of distributions, including normal and Paretian distributions.

The deficiency and contamination problems associated with performance measurement (collectively known as the criterion problem) are still unresolved (Cascio & Aguinis, 2011; Murphy, 2008). Accordingly, given our ambitious goal to challenge a long-held assumption in the field, we deliberately conducted five separate studies including heterogeneous industries and used a large number of performance operationalizations including qualitative evaluations (e.g., award nominations in which raters propose the names of nominees), quantitative outcomes (e.g., number of publications by researchers), and observed behavior based on specific events (e.g., career homeruns of baseball players) or overall reputation (e.g., votes for politicians). Taken together, our five studies involve 198 samples and include 633,263 researchers, entertainers, politicians, and amateur and professional athletes.

Although we deliberately chose an eclectic and heterogeneous set of performance measures, they did share several features. First and most important, each operationalization contained behaviors that directly affected important outcomes such as compensation, bonuses, and promotions. In other words, the performance measures included in our study are particularly meaningful for the individual workers because they have important consequences. Second, the outcomes we chose primarily reflect individual behavior that is largely under an individual's control (Aguinis, 2009). The determinants of an individual's performance are individual characteristics, the context of work, and the interaction between the two. Thus, even in individual sports such as golf, other players influence each golfer's behavior. However, we made efforts to minimize the effects of contextual factors on individual performance. For example, the examination of baseball errors in Study 5 was based on grouping players into their respective positions. This prevented right fielders that typically have fewer fielding opportunities to be grouped with shortstops that have more fielding opportunities. In addition, when possible, we scaled the performance of each individual to a measurement period that was time bound. For example, in Study 4, we included career performance measures as well as single-season measures and examined both single-season performance and career performance.

Our research deliberately excludes samples of individuals whose performance has been exclusively rated by a supervisor because such performance operationalizations include biases that would render addressing our research question impossible. Specifically, instructing raters to follow a bell curve, implementing rater error training programs, normalizing performance ratings, and correcting for leniency or severity bias all help to create a normal curve in a sample regardless of the true underlying distribution.

Study 1

Method

Overview. In Study 1, we tested whether a Paretian or Gaussian distribution better fit the distribution of performance of 490,185 researchers who have produced 943,224 publications across 54 academic disciplines between January 2000 and June 2009.

Procedure

We categorized academic disciplines using Journal Citation Reports (JCR), which provide impact factors in specific subject categories across the physical and social sciences. In some cases, there were multiple subfields included within one JCR category. For instance, there are eight entries for material sciences (e.g., ceramics, paper and wood, composites), but we identified authors across all material sciences so that authors publishing in more than one area would have all their publications included. Our analyses included 54 academic fields (see Table 1). We used impact factors (also reported in JCR) from 2007 to identify the top five journals within each of the 54 fields. We chose field-specific journals to avoid having the search contaminated by authors from other sciences. For instance, *Accounts of Chemical Research* most likely only includes articles related to chemistry, but this assumption cannot be made with an interdisciplinary journal such as *Nature*, which publishes chemistry research alongside other scientific research. We next used the *Publish or Perish* program (Harzing, 2008), which relies on Google Scholar, to identify all authors who had published at least one article in one of these journals between January 2000 and June 2009. These procedures resulted in a total of 490,185 researchers who have produced 943,224 scholarly journal publications.

Operationalization of individual performance. Publication in top-tier journals is the most important antecedent of meaningful outcomes for faculty including salary and tenure status (Gomez-Mejia & Balkin, 1992). Thus, in Study 1 we operationalized performance as research productivity, specifically as the number of articles published by an author in one of the top five journals over the 9.5-year observation period. All authors of each article were recorded, and no differentiation was made based on authorship order.

Results

Results reported in Table 1 show that the Paretian distribution yielded a superior fit than the Gaussian distribution in every one of the 54 scientific fields. Recall that a larger chi-square value indicates worse fit and, thus, can be considered an index of badness of fit. As Table 1 shows, the average misfit for the Paretian distribution was 23,888 whereas the misfit of the normal distribution was larger than forty-four trillion (i.e., 44,199,201,241,681)—a difference in favor of the Paretian distribution in the order of 1:1.9 billion. Figure 2a displays a histogram of the empirically observed performance distribution of researchers. To interpret these results further, consider the field of Agriculture (see Table 1). A normal

TABLE 1
Distribution of Individual Performance of Researchers: Fit With Gaussian vs. Paretian Distributions

Sample	*N*	Mean	*SD*	Gaussian (χ^2)	Paretian (χ^2)
Agriculture	25,006	1.91	2.54	3.15E+14	2.02E+04
Agronomy	8,923	1.42	1.16	2.56E+13	5.07E+04
Anthropology	5,755	1.87	1.95	7.37E+12	6,460
Astronomy	13,101	3.10	3.99	3.17E+12	2,200
Biological psychology	8,332	1.40	1.11	1.96E+12	6.56E+04
Clinical psychology	10,418	1.89	2.38	8.24E+12	2,321
Computer science	3,597	1.45	1.11	1.53E+11	9,523
Criminology	678	1.29	.77	3.09E+11	1.10E+04
Demography	737	1.58	2.91	6.16E+11	3.65E+05
Dentistry	12,345	2.26	2.98	1.64E+13	2,329
Dermatology	30,531	2.25	3.38	4.22E+13	7,801
Developmental psychology	7,303	1.75	1.90	1.39E+12	3,588
Ecology	5,730	1.71	1.68	6.89E+12	2,605
Economics	3,048	1.62	1.67	6.71E+11	4,693
Education	1,201	1.26	.84	3.75E+10	1.11E+06
Educational psychology	3,032	1.70	1.55	5.97E+12	1,668
Environmental science	2,447	1.42	1.17	3.44E+11	3.25E+04
Ergonomics	3,309	1.34	.90	2.81E+12	3.20E+04
Ethics	1,073	1.65	1.78	2.41E+12	1,571
Ethnic studies	2003	1.47	1.38	2.04E+12	2.01E+04
Finance	3,019	2.14	2.52	6.05E+12	1,663
Forestry	12,211	1.82	1.80	4.47E+06	6,098
Genetics	16,574	1.71	2.18	1.83E+13	1.45E+04
History	6,708	1.54	.97	9.04E+09	1.78E+04
Hospitality	1,684	1.38	1.00	1.85E+11	2.23E+04
Industrial relations	1,504	1.34	.83	6.00E+12	7,136
Intl. relations	1,483	1.65	3.09	2.21E+11	2.19E+05
Law	1,350	1.55	1.24	6.09E+11	2,570
Linguistics	3,600	1.73	1.78	3.35E+10	3,058
Material sciences	24,723	1.76	2.42	3.71E+13	2.24E+04
Mathematics	3,972	1.45	1.02	3.64E+13	1.06E+04
Medical ethics	2,928	1.92	3.21	2.10E+12	4,982
Parasitology	11,667	1.78	2.12	1.27E+13	7,650
Pharmacology	11,654	1.54	1.68	3.44E+13	6.76E+04
Physics	1,373	1.18	.73	7.25E+09	4.74E+05
Public administration	3,473	1.73	1.73	2.12E+13	2,408
Radiology	27,146	2.25	2.88	1.59E+13	5,184
Rehabilitation	5,661	1.50	1.52	5.76E+13	3.78E+04
Rheumatology	6,665	1.48	1.25	2.89E+13	2.86E+04
Robotics	5,021	1.92	2.17	7.39E+12	2,953
Social psychology	4,425	2.35	3.04	3.29E+12	1,171
Social work	2,357	1.45	1.16	1.84E+11	7,851

continued

TABLE 1 (continued)

Sample	N	Mean	SD	Gaussian (χ^2)	Paretian (χ^2)
Sociology	2,417	1.81	1.49	5.60E+12	4,024
Sports medicine	16,412	1.79	2.08	1.25E+14	7,819
Statistics	10,679	2.08	2.52	4.32E+13	3,012
Substance abuse	9,513	1.78	1.95	2.45E+13	7,274
Thermodynamics	9,856	2.45	3.31	1.14E+13	1,882
Urban studies	3,548	1.33	.83	5.39E+11	2.73E+04
Urology	37,761	2.25	2.99	9.90E+13	3.34E+04
Vet. science	31,224	1.90	2.13	3.34E+12	3.34E+04
Virology	17,480	2.25	2.88	4.94E+12	9,851
Water resources	25,757	2.43	3.79	7.28E+13	5,043
Women studies	2,982	1.26	1.00	5.39E+12	1.63E+05
Zoology	14,789	1.46	1.13	5.17E+12	6.17E+04
Weighted average				*44,199,201,241,681*	*23,888*

distribution and a sample size of 25,006 would lead to approximately 35 scholars with more than 9.5 publications (three standard deviations above the mean). In contrast, our data include 460 scholars with 10 or more publications. In other words, the normal distribution underestimates the number of extreme events and does not describe the actual distribution well.

Discussion

We tested whether the distribution of research performance best fits a Gaussian distribution or a Paretian distribution. Results based on chi-square statistics and a comparison of Figure 2a with Figure 1 provide evidence that the performance of researchers follows a Paretian distribution.

Study 2

Method

Overview. A potential limitation of Study 1 is that performance (i.e., successful publication) was assessed by a single individual: a journal editor. Even if editors rely heavily on the opinion of associate editors and reviewers, the decision to publish each article has been made by a very small number of individuals in each case (i.e., typically an action editor and two or three reviewers). Moreover, there is evidence regarding the low reliability (i.e., consistency across reviewers) in the peer review process (Gilliland & Cortina, 1997). Therefore, it is possible that the job performance distribution of researchers is idiosyncratic enough to challenge

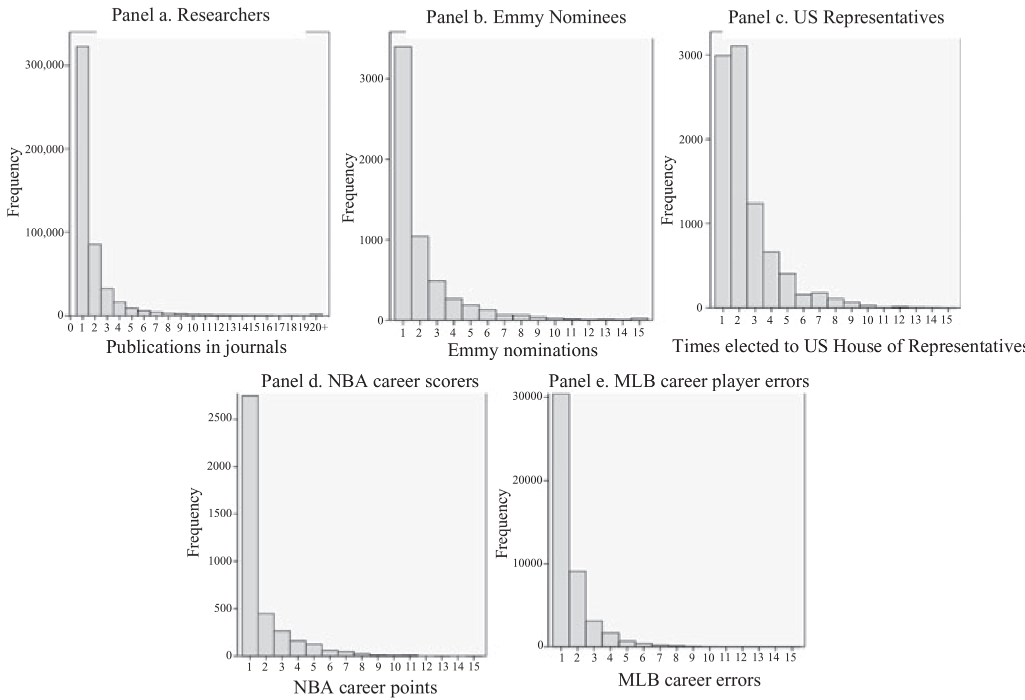

Figure 2: **Distribution of Individual Performance for Researchers (*N* = 490,185), Emmy Nominees (*N* = 5,826), United States Representatives (*N* = 8,976), NBA Career Scorers (*N* = 3,932), and Major League Baseball (MLB) Career Errors (*N* = 45,885).**

Note. For all Y axes, "Frequency" refers to number of individuals. For clarity, individuals with more than 20 publications (Panel a) and more than 15 Emmy nominations (Panel b) were included in the last bins. For panels c–e, participants were divided into 15 equally spaced bins.

the notion that it generalizes to other types of workers. Accordingly, we conducted Study 2 to expand the generalizability of these findings and test whether a different industry with different performance metrics and larger number of raters involved would confirm results from Study 1. In Study 2, we examined the performance of 17,750 individuals in the entertainment industry, with performance rated by a large voting body or more objective performance measures such as the number of times an entertainer received an award, nomination, or some other indicator (e.g., Grammy nominations, *New York Times* best-selling list).

Procedure. To capture the multifaceted nature of performance of entertainers, we completed the following steps. First, we generated a list of the different forms of entertainment and determined that motion pictures, music, literature, and theater would serve as the population of interest given the availability of data. Second, we consulted several individuals within the film and music industries to help identify well-known (e.g., Oscars, Grammys) as well as lesser-known and more industry-specific awards (e.g., Edgar Allen Poe Award for mystery writing). Third, we proceeded with data collection by searching Web sites for relevant data. When data were not available online, we contacted the organization that distributes the awards for more detailed information. We identified more than 100 potential entertainment awards, but incomplete records and award limits (i.e., certain awards limit the number of times a recipient can win) reduced the number of qualified groups. Because a small group of awardees could diverge from normality due to either sampling error or true divergence from normality, we stipulated that a group must consist of at least 100 individuals in order to qualify for inclusion. Forty-two (42) awards and honors met our inclusion criteria (see Table 2).

Operationalization of Individual Performance. Award nominations (Oscars, Emmys), expert rankings (*Rolling Stone*), and appearances on a best seller list such as the *New York Times* Best Seller List all served as measures of individual performance. These types of performance measures either are based on expert opinions (e.g., music critics) or peer voting (e.g., Oscars). Although the number of nominations a performer receives is a count variable, these counts encapsulate ratings and may better conform to traditional subjective ratings such as those most typically found in traditional OBHRM research (Landy & Farr, 1980).

Results

Table 2 shows results that are very similar to those in Table 1: The distribution of individual performance is substantially closer to a Paretian distribution than to a Gaussian distribution for each of the 42 samples. The average misfit of the Gaussian distribution was more than

TABLE 2

Distribution of Individual Performance of Entertainers: Fit With Gaussian vs. Paretian Distributions

Sample	*N*	Mean	*SD*	Data collection time frame	Gaussian (χ^2)	Paretian (χ^2)	Performance operationalization and comments
AVN nom. actor	132	1.83	1.36	2008	480	160	AVN nominations across a wide variety of
AVN nom. actress	245	1.77	1.38	2008	1.03E+04	251	categories counted towards the
AVN nom. actor	135	1.82	1.66	2009	4.29E+04	78	performance total
AVN nom. actress	302	1.82	1.50	2009	3.61E+07	153	
AVN nom. director	108	1.52	1.20	2009	1.42E+08	187	
Cable ACE nom. actor	115	1.23	.55	1978–1997	4,212	945	Nominees for best director or acting role
Cable ACE nom. actress	104	1.21	.59	1978–1997	6,392	4,685	
Country Music Awards nom.	106	1.84	1.49	1967–2009	7.89E+05	52	Ratings for Best Male or Female Vocalist
Edgar Allen Poe Awards nom.	121	15.07	10.40	1954–2009	147	53	Expert rankings in Best Novel category
Emmy nom. Actor	685	2.86	2.74	1949–2009	5.49E+07	173	Nomination to any category and an artist
Emmy nom. Actress	442	2.49	2.33	1949–2009	1.41E+06	110	can obtain multiple nominations in the
Emmy nom. Art direction	866	1.82	2.37	1949–2009	4.26E+10	1,097	same year. The nomination process
Emmy nom. Casting	193	2.16	2.01	1949–2009	2.46E+05	51	combines a popular vote with volunteer
Emmy nom. Choreography	127	1.71	1.54	1949–2009	4.93E+05	140	judging panels
Emmy nom. Cinematography	588	1.68	1.43	1949–2009	1.57E+07	387	
Emmy nom. Direction	395	1.95	2.07	1949–2009	1.01E+08	92	
Emmy nom. Editing	942	1.89	1.77	1949–2009	3.83E+13	614	
Emmy nom. Lighting	131	3.02	3.55	1949–2009	2.51E+05	29	
Emmy nom. Writing	1,457	2.46	2.72	1949–2009	3.10E+09	356	
Golden Globe nom. actor	392	2.07	2.02	1944–2009	2.38E+09	111	Nomination to any category and an artist
Golden Globe nom. actress	415	2.05	2.06	1944–2009	2.14E+08	123	can obtain multiple nominations in the

continued

TABLE 2 (continued)

Sample	N	Mean	SD	Data Collection Time Frame	Gaussian (χ^2)	Paretian (χ^2)	Performance Operationalization and Comments
Golden Globe nom. direction	156	1.94	1.53	1944–2009	1.19E+04	95	same year. The Hollywood Foreign
Golden Globe nom. TV actor	375	2.12	1.78	1944–2009	7.25E+04	203	Press Association rates and votes on the
Golden Globe nom. TV actress	354	2.19	1.79	1944–2009	3.27E+04	234	nominees
Grammy nom.	3313	2.02	2.78	1959–2009	3.82E+12	1,307	Nomination to any category
Man Booker Prize Fiction nom.	283	1.35	.84	1969–2009	3.62E+05	2,018	Expert rankings in Best Novel category
MTV VMA nom.	561	3.98	5.58	1984–2009	1.37E+07	78	Fan voting and industry ratings
NYT Best Seller fiction	222	2.42	3.85	1950–2009	4.54E+08	219	Each appearance on the *New York Times*
NYT Best Seller nonfiction	419	1.19	.65	1950–2009	2.71E+06	6.20E+04	Bestseller list
Oscar nom. actor	421	1.84	1.62	1927–2009	3.34E+11	177	Nominations as determined by Academy
Oscar nom. art direction	531	2.64	3.75	1927–2009	8.10E+08	94	members using a preferential-voting
Oscar nom. direction	199	1.97	1.60	1927–2009	1.57E+06	76	system for best director and nominees
Oscar nom. Actress	432	1.80	1.50	1927–2009	4.07E+07	289	in a primary or supporting acting role
Oscar nom. cinematography	159	1.91	1.56	1927–2009	5.67E+06	84	
PEN award voting	125	14.55	11.45	1976–2009	2295	21	Nomination in any category (e.g., drama)
Pulitzer Prize nom. drama	121	1.26	.75	1917–2009	2.00E+06	3711	Selection to finalist for the drama category
Rolling Stone Top 500 albums	261	1.90	1.52	1940–2009	1.78E+06	137	Number of appearances on the Top 500
Rolling Stone Top 500 songs	247	2.02	2.19	1940–2009	1.27E+08	75	list as rated by contributors and writers
Tony nom. actress	583	1.59	1.18	1947–2009	3.91E+08	817	Nominations determined by a panel of
Tony nom. choreography	108	2.10	2.10	1947–2009	445	93	judges from entertainment industry
Tony nom. actor	642	1.43	.94	1947–2009	1.77E+08	2,056	
Tony nom. director	237	1.86	1.70	1947–2009	1.00E+12	133	
Weighted average					*2,769,315,505,476*	*2,092*	

Note. ACE = Award for Cable Excellence, AVN = Adult Video News, MTV = Music Television, PEN = poets, playwrights, essayists, editors, and novelists, *NYT* = *New York Times*, TV = television, nom = nominations.

1 billion times larger than the average misfit of a Paretian distribution (i.e., 2,769,315,505,476 vs. 2,092). To understand the nature of these results better, consider the Grammy nominations under an assumption of normality. Of the 3,313 individuals nominated for a Grammy, only 5 should be three standard deviations above the mean with more than 10 nominations. In contrast, our data include 64 entertainers with more than 10 nominations. As in Study 1, the normal curve does not describe the actual distribution well.

Discussion

Results of Study 2 closely matched those of Study 1. Entertainers' performance better fits a Paretian distribution than a Gaussian distribution. These findings across a variety of entertainment industries and with multiple performance operationalizations provides further evidence regarding the robustness of the Paretian distribution as the better model of individual performance compared to a normal distribution. As an illustration, Figure 2b displays the overall distribution of the largest sample of entertainers, Emmy nominees. This histogram illustrates that the empirically derived distribution aligns with a Paretian distribution (cf. Figure 1).

Study 3

Method

Overview. In Study 3, we examined the distribution of performance of politicians. Study 3 includes a set of performance raters that is even more inclusive compared to those in Study 2: All citizens eligible to vote in a given political jurisdiction. In Study 3 we examined the performance of candidates (i.e., being elected or reelected) running for elective offices at the state (e.g., legislature of the state of Oregon in the U.S.) and national levels (e.g., Dutch Parliament). We included the performance of 42,745 candidates running for office in 42 types of elections in Australia, Canada, Denmark, Estonia, Finland, Holland, Ireland, New Zealand, the United Kingdom, and the U.S.

Procedure. We identified elected officials through national and state Web sites. We first constructed a list of 195 nations, 50 U.S. states, 10 Canadian provinces, and 6 Australian territories to serve as potential sources of data. The search began at the national level, and we eliminated nations without a democratic form of government such as absolute monarchies (e.g., Saudi Arabia), theocracies (e.g., Vatican City), and one-party nations (e.g., Cuba). Next, offices with term limits or lifetime appointments were excluded as the results would be artificially truncated at the maximum number of terms an individual could serve. For this reason, we eliminated most executive and judicial branches of government, thus leaving legislatures as the primary source of data. For the remaining potential data sources, we searched for a complete list of current and past members in each country, state, and province. Lists of current members were available for nearly all governments, but relatively few governments made past members or the dates that current members were first elected available. For example, the reporting of past members for the Australian legislature was intermittent, therefore a complete list of members was not available. However, a complete list of present members and their original election to office, as well as a complete list of the most recent legislature that contained no current members (1969), were also available. Because these two groups had no overlap, we included them separately in the database. As in Study 2, we limited our search to groups that contained at least 100 individuals. We identified 42 samples from state and national governing bodies. Table 3 includes descriptive information about each of these samples.

Operationalization of individual performance. As stated earlier, elected official performance was operationalized by an individual's election to office. In most cases, this was established as the number of times a person's name appeared on each new session of the legislature. In cases where only the current legislature was available, the length of service was recorded (either days or years in office) and used as a measure of performance. Thus, this type of performance measure is essentially based on ratings—those provided by eligible voters in any given political district.

Results

Results included in Table 3 indicate that the data fit a Paretian distribution better than a Gaussian distribution for 31 of the 42 samples. The average fit strongly favored the Paretian (misfit of 8,692) versus the Gaussian distribution (misfit of over one trillion). Using the U.S. House of Representatives as an example, the normal distribution suggests that of the 8,976 individuals to have served in the House, no more than 13 representatives should be three standard deviations above the mean (i.e., serve more than 13 terms). Contrary to the expectation based on the normality assumption, 173 U.S. Representatives have served more than 13 terms.

Discussion

Results suggest that the individual performance of politicians follows a Paretian distribution more closely than a normal distribution. Specifically, the data fit a Paretian distribution more closely than a normal

TABLE 3

Distribution of Individual Performance of Politicians: Fit With Gaussian vs. Paretian Distributions

Sample	*N*	Mean	*SD*	Data collection time frame	Gaussian (χ^2)	Paretian (χ^2)
Alabama Leg.	104	11.43	8.47	2009	79	157
Australia House	128	8.57	8.10	1969	22	164
Australia House	153	10.46	6.84	2009	122	119
Canadian Leg.	4,059	2.65	1.87	1867–2009	1.06E+07	1.05E+04
Connecticut Leg.	151	9.89	6.31	2009	28	249
Denmark Par.	177	10.41	7.39	2009	167	354
Dutch Par.	150	5.32	3.90	2009	1.01E+05	184
Estonia Par.	100	2.00	1.11	2009	225	167
Finland Par.	200	9.39	7.74	2009	293	229
Georgia House	179	4.80	3.89	2009	333	96
Illinois Leg.	120	9.96	6.48	2009	60	220
Iowa Leg.	100	6.74	4.89	2009	42	11
Ireland Par.	1,147	3.99	3.15	1919–2009	2970	1,443
Ireland Senate	716	2.40	1.95	1919–2009	2.06E+04	809
Kansas House	5,675	2.72	2.94	2009	6.53E+12	5,636
Kansas Senate	1,209	4.01	3.34	1812–2009	3.76E+07	1,171
Kentucky Leg.	100	5.06	4.04	2009	82	57
Louisiana House	3,468	1.93	1.97	1812–2009	2.09E+12	6,128
Maine Leg.	153	2.58	2.06	2009	2.44E+09	24
Maryland Leg.	141	9.42	7.63	2009	212	165
Mass. House	160	9.82	6.88	2009	113	205
Minnesota House	134	4.31	3.66	2009	387	47
Missouri Leg.	163	4.70	2.19	2009	921	571
New Bruns. Par.	1,136	2.24	1.52	1882–2009	7,933	2,855

continued

TABLE 3 (continued)

Sample	N	Mean	SD	Data collection time frame	Gaussian (χ^2)	Paretian (χ^2)
New York Association	148	11.61	8.94	2009	76	193
New Zealand Leg.	122	8.05	7.49	2009	120	126
N. Carolina Association	124	4.63	3.38	2009	68	141
Nova Scotia Leg.	414	3.01	1.26	1867–2009	2,096	539
Oklahoma Leg.	101	4.70	2.67	2009	98	254
Ontario Leg.	1,879	4.56	3.30	1867–2009	1.62E+05	5,539
Oregon Leg.	377	4.47	3.81	1858–2009	2654	108
Oregon Senate	161	5.45	4.44	1858–2009	1597	94
Pennsylvania House	200	10.76	9.18	2009	269	141
Quebec Leg.	399	3.52	2.40	1867–2009	697	583
S. Carolina House	125	8.23	6.44	2009	87	118
Tasmania Association	542	3.11	2.35	1856–2009	4742	442
Tennessee House	100	5.22	4.10	2009	134	63
UK Par.	7,214	3.41	2.59	1801–2009	4.32E+09	1.70E+04
US House	8,976	3.42	3.23	1789–2009	6.39E+08	1.43E+04
US Senate	1,840	9.14	7.79	1789–2009	2.42E+04	3,264
Virginia Association	100	11.09	8.26	2009	497	96
Wisconsin Leg.	100	8.11	6.99	2009	309	89
Weighted average					*1,037,389,925,013*	*8,692*

Note. UK = United Kingdom, US = United States, Bruns = Brunswick, Leg = Legislature, Mass = Massachusetts, Par = Parliament, S = South, N = North.

distribution in 74% of the samples we studied. Figure 2c illustrates the shape of the observed performance distribution for the largest sample in Study 3, the U.S. House of Representatives. One reason why the superior fit for a Paretian distribution was not seen in all samples may be that the established frequency of elections prevents superstar performers to emerge in the same way as they do in other industries like those we examined in Studies 1 and 2. We speculate that this may be a result of the measure of performance not being as sensitive and discriminating regarding various levels of performance compared to the measures of performance we used in Studies 1 and 2. Performance in Study 3 was binary with the legislator either returning to office with a simple majority vote or being ousted by another also by a simple majority vote. Therefore, an incumbent who does just enough to be reelected (e.g., victory margin of 1%) receives the same performance "score" as an extraordinary incumbent (e.g., victory margin of 25%). Nevertheless, on average, the difference in fit favored the Paretian distribution in the order of 1:119 million.

Study 4

Method

Overview. In Study 4 we investigated the performance of athletes in collegiate and professional sports. Study 4 presents additional evidence to supplement Studies 1–3 because many of the measures of performance are more objective, they rely more heavily on measures of physical performance, and depend more strongly on the individual (see Table 4).

Procedure. We compiled a list of individual and team sports at the collegiate and professional levels. We accessed each Web site that hosted the sport (e.g., MLB.com) and downloaded the necessary statistics from each database. In all cases, we were able to find the necessary data for the chosen sports for at least one complete season. In most cases, we collected data from multiple years, and we were able to record all players, but in some cases, only the top players were available (e.g., only the top 1,000 MLB pitchers were available).

Operationalization of individual performance. We attempted to identify the most individual-based measures of performance. For instance, runs batted (RBI) in baseball are both a function of a hitter's prowess and the ability of those batting before him. Therefore, a less contaminated measure for baseball is home runs. For individual sports such as golf and tennis, we chose the total numbers of wins, but team sports required a different operationalization. For sports such as soccer and hockey, we used goals or points as the performance metric, and for position-oriented sports like U.S. football we used receptions, rushing yards, and touchdowns.

TABLE 4

Distribution of Positive Individual Performance of Athletes: Fit With Gaussian vs. Paretian Distributions

Sport	*N*	Mean	*SD*	Data collection time frame	Gaussian (χ^2)	Paretian (χ^2)	Performance operationalization and comments
MLB career strike outs	1,001	1103.67	563.26	1900–2006	6.00E+11	91	A variety of metrics of positive
MLB career HR	1,004	174.00	109.44	1900–2006	4850	89	behaviors for players and managers
MLB career mgr wins	647	301.02	450.90	1900–2006	3630	480	
NCAA Div 1-1/ERA	516	.25	.07	2009	5.26E+09	59	A variety of metrics of positive
NCAA Div 1-HR	548	11.57	3.76	2009	959	211	behaviors for baseball players only
NCAA Div 2-1/ERA	300	.31	.11	2009	2770	17	
NCAA Div 2-HR	383	10.26	3.66	2009	5.08E+03	92	
NCAA Div 3-1/ERA	500	.31	.12	2009	9.69E+07	25	
NCAA Div 3-HR	424	7.27	2.91	2009	5.08E+06	87	
NCAA pass. TDs	193	11.65	10.41	2009	279	171	A variety of metrics of positive
NCAA rushing	529	407.56	444.68	2009	788	958	behaviors for football players only
NCAA WR yd.	798	299.09	294.03	2009	1100	649	
NCAA TE yd.	297	146.84	190.73	2009	1.20E+06	301	
NCAA sacks	992	2.55	2.23	2009	5.85E+04	1081	
Cricket runs	252	4279.48	2205.77	1909–2009	239	52	Top 200 cricketers in runs/wickets
Cricket wickets	150	201.55	117.84	1909–2009	1830	15	
EPL goals	1,521	10.90	19.72	1992–2009	1.17E+12	96	Number of goals scored
NBA coaches wins	258	183.15	263.73	1946–2009	9.57E+05	147	A variety of metrics of positive
NBA career points	3,932	2670.91	4308.44	1946–2009	2.62E+12	3517	behaviors for players and managers
PGA career wins	200	14.05	12.62	1916–2009	2.08E+04	34	All time tournament wins
Men's swimming	654	1.78	1.38	1896–2009	8.08E+07	681	Gold, silver, or bronze medal across an
Women's swimming	538	1.75	1.39	1896–2009	1.19E+06	773	entire career
Men's track	981	1.34	.76	1896–2009	4.42E+09	5058	

continued

TABLE 4 (continued)

Sport	*N*	Mean	*SD*	Data collection time frame	Gaussian (χ^2)	Paretian (χ^2)	Performance operationalization and comments
Women's track	437	1.45	.94	1896–2009	1.41E+08	943	
Men's alpine	167	1.46	.94	1896–2009	1.10E+08	265	
Women's alpine	148	1.64	.98	1896–2009	791	198	
PBA titles	200	4.95	6.70	1959–2009	1.43E+06	47	All time tournament wins
NASCAR points	125	1138.41	1410.47	2009	262	119	Points earned in the Sprint Cup series
NFL coaches wins	413	31.25	46.64	1999–2009	1.19E+07	104	A variety of metrics of positive behaviors for players and managers
NFL kick return yd.	250	3238.43	1698.95	1999–2009	2.67E+06	29	
NFL TD receptions	253	50.92	20.59	1999–2009	2.29E+08	41	
NFL field goals	252	110.61	108.69	1999–2009	683	118	
NFL sacks	251	59.38	28.43	1999–2009	9240	37	
NFL rushing yd.	250	5611.01	2708.46	1999–2009	127	20	
NFL passing yd.	250	16897.20	11431.10	1999–2009	312	112	
NHL defense assists	1,533	107.12	165.48	1917–2009	4.96E+09	909	Points scored for all NHL players across their careers
NHL centers points	1,213	191.55	300.87	1917–2009	3.64E+05	708	
NHL right wing points	1,073	162.36	246.82	1917–2009	5.11E+07	537	
NHL left wing points	1,102	141.81	210.36	1917–2009	2.32E+06	640	
NHL goalies saves	392	3497.99	4848.02	1917–2009	1650	526	
Men's tennis	146	2.94	2.80	1877–2009	1080	41	Grand Slam tournament wins across an entire career
Women's tennis	110	3.68	4.37	1877–2009	1140	29	
NCAA basketball	100	615.21	75.02	2009	142	8	Points scored for a single season
Weighted average					*502,193,974,189*	*1,076*	

Note. EPL = English Premier League, ERA = Earned run average, HR = Homeruns, MLB = Major League Baseball, NASCAR = National Association for Stock Car Auto Racing, NBA = National Basketball Association, NCAA = National Collegiate Athletic Association. NFL = National Football League, NHL = National Hockey League, PBA = Professional Bowling League, PGA = Professional Golf Association, TD = touchdowns, yd = yards.

Results

Results summarized in Table 4 indicate that the distribution of individual performance follows a Paretian distribution more closely than a Gaussian distribution. In addition, we examined the distribution of performance within teams and for a more limited time period rather than across organizations and careers. This helps rule out the possibility that our results are due to analyses across organizations and careers that may accentuate the preponderance of outliers. We examined the most recent season (2009–2010) for teams in the English Premier League (goals), Major League Baseball (homeruns and strikeouts), and National Hockey League (points). The 94 samples ($N = 1{,}797$) for these analyses are considerably smaller, ranging from 9 to 30 athletes, thus making sampling error a limitation. However, in 84 of these 94 comparisons, chi-square statistics provided evidence regarding the superior fit of the Paretian distribution.

Discussion

Across a variety of sports, nationalities, and levels of play, the Paretian distribution yielded a significantly better fit than the Gaussian distribution. In all but one sample, the performance distribution favored a power law more strongly than a normal distribution. Moreover, the sports and types of athletic performance included in Study 4 vary regarding the extent to which an individual performer participated in a team or individual sport, and we also conducted within-team and within-season analyses. The overall weighted mean chi-square statistic for the Paretian distribution is about 466 million times smaller (i.e., indicating better fit) compared to a Gaussian distribution. Figure 2d illustrates these results in a histogram of Study 4's largest sample: National Basketball Association (NBA) career points.

Study 5

Method

Overview. Studies 1–4 focused on positive performance (e.g., publications, awards). However, the Paretian distribution possesses a greater frequency of extreme values at both ends of the distribution. If individual performance is truly Paretian, then both sides of the performance distribution should contain a disproportionate number of individuals. Our decision to conduct Study 5 focusing on negative performance was also motivated by an increased interest in counterproductive work behaviors (CWBs), such as abusive supervision (Tepper, Henle, Lambert, Giacalone,

& Duffy, 2008), employee theft (Greenberg, 2002), and workplace bullying (Salin, 2003).

The challenge to investigating the negative performance distribution is that, unlike positive performance, negative performance is often covert. Even in industries where performers are under constant scrutiny, a considerable number of negative behaviors occur without the knowledge of others. A researcher that forges his data, an actor hiding a drug addiction, and a politician accepting bribes are all behaviors that certainly qualify as negative, but quantifying their frequency is difficult. Complicating matters further, negative performance can be done intentionally (e.g., consuming alcohol while working) or unintentionally (e.g., accidentally breaking the office printer).

Procedure. To identify the distribution of negative performance, we used samples from collegiate and professional sports. We examined the distribution of negative performance such as English Premier League yellow cards, NBA career turnovers, and Major League Baseball first-base errors. The negative performance behaviors vary in their intent, severity, and frequency, but they are similar in that they are all detrimental to the organization. Using samples from sports allows for an objective examination of negative performance that would be virtually impossible to capture in industries in which performance is not measured so systematically and with more ambiguous definitions of what constitutes positive and negative performance. Study 5 included 17 types of negative performance of 57,300 individual athletes in six sporting disciplines (see Table 5).

Operationalization of individual performance. We attempted to identify negative behaviors that were primarily attributable to an individual performer. For instance, an incomplete pass in U.S. football can be either a failure of the quarterback or a failure of the receiver, but an interception is primarily viewed as a quarterback error. We identified four sports for which negative performance can largely be attributed to one individual. We contacted several individuals that currently or formerly participated in these sports at an elite level (received an NCAA scholarship or played professionally) to serve as subject matter experts and ensure that our classifications were acceptable. When possible, we divided the sport into individual positions. This was done because some negative performance behaviors are more likely due to the role the player has on the team. For example, shortstops are most likely to have a baseball hit to them, thus they have a greater chance of being charged an error. In keeping with our focus on performance measures that lead to important outcomes, we chose negative performance that is most likely to result in meaningful consequences for the athlete (e.g., fines, suspensions, losing sponsorships, being cut from a team).

TABLE 5

Distribution of Negative Individual Performance of Athletes: Fit With Gaussian vs. Paretian Distributions

Sample	N	Mean	SD	Data collection time frame	Gaussian (χ^2)	Paretian (χ^2)	Performance operationalization and comments
MLB hit batters in a single season	1,007	51.54	26.46	1900–2006	1.02E+05	65	Errors assigned for MLB players
MLB 1B errors in a single season	5,933	5.66	5.75	1900–2006	2.41E+06	4,097	
MLB 2B errors in a single season	6,400	8.05	8.74	1900–2006	2.32E+06	1,534	
MLB 3B errors in a single season	7,099	8.01	8.55	1900–2006	1.53E+07	3345	
MLB C errors in a single season	6,276	5.46	4.73	1900–2006	4.59E+09	6973	
MLB OF errors in a single season	13,721	4.27	3.81	1900–2006	6.57E+11	2.36E+04	
MLB SS errors in a single season	6,456	11.98	13.55	1900–2006	5.19E+06	3043	
EPL yellow cards	1,876	9.78	12.26	1992–2009	4.73E+06	442	Yellow cards administered
NBA career turnover	2,691	427.94	604.02	1946–2009	3.15E+07	2487	A variety of metrics of negative behaviors for players
NBA career tech, flag, and ejections	432	8.84	10.33	2009	1.96E+09	109	

continued

TABLE 5 (continued)

Sample	*N*	Mean	*SD*	Data collection time frame	Gaussian (χ^2)	Paretian (χ^2)	Performance operationalization and comments
NFL fumbles	251	55.28	23.19	1999–2009	2030	26	A variety of metrics of negative
NFL interceptions	253	102.74	60.76	1999–2009	141	72	behaviors for players
NHL defense penalty minutes	1,505	368.67	460.31	1917–2009	7.49E+06	1580	Penalty minutes received for all
NHL centers penalty minutes	1,129	216.84	329.99	1917–2009	8.38E+06	546	NHL players across their careers
NHL right wing penalty minutes	1,015	288.89	466.39	1917–2009	9.02E+07	567	
NHL left wing penalty minutes	1,053	286.61	449.41	1917–2009	7.13E+11	605	
NCAA thrown interceptions	202	7.02	4.61	2009	97	45	Quarterbacks only
Weighted average					*170,951,462,785*	*7,975*	

Note. 1B = first basemen, 2B = second basemen, 3B = third basemen, C = catchers, EPL = English Premier League, OF = outfielders, MLB = Major League Baseball, NBA = National Basketball Association, NFL = National Football League, NCAA = National Collegiate Athletic Association, tech/flag = technical or flagrant foul, NHL = National Hockey League, PGA = Professional Golf Association, SS = shortstops.

Results

Results regarding the distribution of negative performance of athletes are included in Table 5. Each of the samples fitted a Paretian distribution better than a Gaussian distribution. The average misfit of the Paretian distribution was only 7,975 whereas the average misfit of the normal distribution exceeded 170 billion. Similar to Study 4, we conducted within-team, within-season analyses to rule out a potential longevity confounding effect. We used EPL yellow cards, MLB hit batters, and NHL penalty minutes for the 2009–2010 season. This included 67 teams ($N = 1419$), of which 52 better conformed to a Paretian distribution.

Discussion

Results of Study 5 closely match those regarding positive performance distributions investigated in Studies 1–4. The distribution of negative performance more closely resembles a power law compared to a normal distribution. We found this same result regarding every one of the 17 samples including a variety of negative performance behaviors and across several types of sports including baseball, hockey, basketball, and non-U.S. football. As an illustration, Figure 2e includes a histogram for Study 5's largest sample: MLB errors. The within-team, within-season analyses further supported the Paretian distribution of performance output.

General Discussion

Theories and practices about performance, personnel selection, training, leadership, and many other domains in OBHRM are firmly rooted in the "norm of normality" where individual performance follows a normal distribution and deviations from normality are seen as "data problems" that must be "fixed." Although some may be skeptical regarding this norm of normality, there is little evidence that this belief has affected theoretical developments or application in an influential way. This norm has developed many decades ago and, to our knowledge, has not been tested systematically and comprehensibly. Individual performance serves as a building block not only for OBHRM and I-O psychology but for most organizational science theories and applications. Thus, understanding the distribution of individual performance is a key issue for organizational science research and practice.

Our central finding is that the distribution of individual performance does not follow a Gaussian distribution but a Paretian distribution. Our results based on five separate studies and involving 198 samples including 633,263 researchers, entertainers, politicians, and amateur and professional athletes are remarkably consistent. Of a total of 198 samples of performers, 186 (93.94%) follow a Paretian distribution more closely than a Gaussian distribution. If, as our results suggest, most performance outcomes are attributable to a small group of elite performers, then both theory and practice must adjust to the substantial role played by these individuals. Next, we discuss implications of our findings for theory and substantive research; research methodology; and practice, policy making, and society.

Implications for Theory and Substantive Research

Our results have important implications for past and future research in substantive domains in OBRHM, I-O psychology, and other fields concerned with individual performance (e.g., strategy, entrepreneurship). We limit our discussion to key areas identified as being among the most popular in OBHRM over the past 45 years (Cascio & Aguinis, 2008a): performance measurement and management, utility analysis in preemployment testing and training and development, leadership and teamwork, and the understanding and prediction of performance.

Regarding performance measurement and management, the current *zeitgeist* is that the median worker should be at the mean level of performance and thus should be placed in the middle of the performance appraisal instrument. If most of those rated are in the lowest category, then the rater, measurement instrument, or both are seen as biased (i.e., affected by severity bias; Cascio & Aguinis, 2011 chapter 5). Performance appraisal instruments that place most employees in the lowest category are seen as psychometrically unsound. These basic tenets have spawned decades of research related to performance appraisal that might "improve" the measurement of performance because such measurement would result in normally distributed scores given that a deviation from a normal distribution is supposedly indicative of rater bias (cf. Landy & Farr, 1980; Smither & London, 2009a). Our results suggest that the distribution of individual performance is such that most performers are in the lowest category. Based on Study 1, we discovered that nearly two thirds (65.8%) of researchers fall below the mean number of publications. Based on the Emmy-nominated entertainers in Study 2, 83.3% fall below the mean in terms of number of nominations. Based on Study 3, for U.S. representatives, 67.9% fall below the mean in terms of times elected. Based on Study 4, for NBA players, 71.1% are below the mean in terms of points scored. Based on Study 5, for MLB players, 66.3% of performers are below the mean in terms of career errors. Moving from a Gaussian to a Paretian perspective, future research regarding performance measurement would benefit from the development of measurement instruments that, contrary to past efforts, allow for the identification of those top performers who account for the majority of results. Moreover, such improved measurement instruments should not focus on distinguishing between slight performance differences of non-elite workers. Instead, more effort should be placed on creating performance measurement instruments that are able to identify the small cohort of top performers.

As a second illustration of the implications of our results, consider the research domain of utility analysis in preemployment testing and training and development. Utility analysis is built upon the assumption of normality, most notably with regard to the standard deviation of individual performance (SD_y), which is a key component of all utility analysis equations. In their seminal article, Schmidt et al. (1979) defined SD_y as follows: "If job performance in dollar terms is normally distributed, then the difference between the value to the organization of the products and services produced by the average employee and those produced by an employee at the 85th percentile in performance is equal to SD_y" (p. 619). The result was an estimate of \$11,327. What difference would a Paretian distribution of job performance make in the calculation of SD_y? Consider the distribution found across all 54 samples in Study 1 and the productivity levels in this group at (a) the median, (b) 84.13th percentile, (c) 97.73rd percentile, and (d) 99.86th percentile. Under a normal distribution, these values correspond to standardized scores (z) of 0, 1, 2, and 3. The difference in productivity between the 84.13th percentile and the median was two, thus a utility analysis assuming normality would use $SD_y = 2.0$. A researcher at the 84th percentile should produce \$11,327 more output than the median researcher (adjusted for inflation). Extending to the second standard deviation, the difference in productivity between the 97.73rd percentile and median researcher should be four, and this additional output is valued at \$22,652. However, the difference between the two points is actually seven. Thus, if SD_y is two, then the additional output of these workers is \$39,645 more than the median worker. Even greater disparity is found at the 99.86th percentile. Productivity difference between the 99.86th percentile and median worker should be 6.0 according to the normal distribution; instead the difference is more than quadruple that (i.e., 25.0). With a normality assumption, productivity among these elite workers is estimated at \$33,981 (\$11,327 $\times$ 3) above the median, but the productivity of these workers is actually \$141,588 above the median. We chose Study 1 because of its large overall sample size, but these same patterns of productivity are found across all five studies. In light of our results, the value-added created by new preemployment tests and the dollar value of training programs should be reinterpreted from a Paretian point of view that acknowledges that the differences between workers at the tails and workers at the median are considerably wider than previously thought. These are large and meaningful differences suggesting important implications of shifting from a normal to a Paretian distribution. In the future, utility analysis should be conducted using a Paretian point of view that acknowledges that differences between workers at the tails and workers at the median are considerably wider than previously thought.

Our results also have implications for OB domains. For example, consider the case of leadership research that, similar to other OB domains (e.g., work motivation, job satisfaction, organizational commitment), has traditionally focused on the "average worker" and how best to improve a group's mean performance. Leadership theories grounded in Gaussian thinking focus on the productivity of the majority of workers rather than the workers responsible for the majority of productivity. Given a normal distribution, 68% of output should derive from the individuals located between the 16th percentile and 84th percentile. With so much of the total output produced by workers around the median, it makes sense for leaders to focus most of their energy on this group. However, if performance follows a Paretian distribution similar to that found in Study 1, only 46% (vs. 68%) is produced by this group. If we extend this illustration further, we expect approximately 95% of the output to be produced by workers between the 2.5th and 97.5th percentiles in a normal distribution. However, using the distribution found in Study 1, only 81% of output comes from this group of workers. With less output from the center of the distribution, more output is found in the tails. Ten percent of productivity comes from the top percentile and 26% of output derives from the top 5% of workers. Consequently, a shift from a normal to a Paretian distribution points to the need to revise leadership theories to address the exchanges and influence of the extreme performers because our results demonstrate that a small set of followers produces the majority of the output. Leadership theories that avoid how best to manage elite workers will likely fail to influence the total productivity of the followers in a meaningful way. Thus, greater attention should be paid to the tremendous impact of the few vital individuals. Despite their small numbers, slight percentage increases in the output of top performers far outweigh moderate increases of the many. New theory is needed to address the identification and motivation of elite performers.

In addition to the study of leadership, our results also affect research on work teams (e.g., group empowerment, shared efficacy, team diversity). Once again, our current understanding of the team and how groups influence performance is grounded in an assumption of normality. The common belief is that teamwork improves performance through increased creativity, synergies, and a variety of other processes (Mathieu, Maynard, Rapp, & Gilson, 2008). If performance follows a Paretian distribution, then these existing theories are insufficient because they fail to address how the presence of an elite worker influences group productivity. We may expect the group productivity to increase in the presence of an elite worker, but is the increase in group output negated by the loss of individual output of the elite worker being slowed by non-elites? It may also be that elites only develop in interactive, dynamic environments, and the isolation of elite workers or grouping multiple elites together could hamper their abnormal productivity. Once again, the finding of a Paretian distribution of performance requires new theory and research to address the elite nested within the group. Specifically, human performance research should adopt a new view regarding what human performance looks like at the tails. Researchers should address the social networks of superstars within groups in terms of identifying how the superstar emerges, communicates with others, interacts with other groups, and what role non-elites play in the facilitating of overall performance.

At a more fundamental level, our understanding of job performance itself needs revisiting. Typically, job performance is conceptualized as consisting of three dimensions: in-role or task behavior, organizational citizenship behavior (OCB), and CWB (Rotundo & Sackett, 2002). CWB (i.e., harmful behaviors targeted at the organization or its members) has always been assumed to have a strong, negative relation with the other two components, but it is unclear if this relationship remains strong, or even negative, among elite performers. For example, the superstars of Study 4 often appeared as supervillains in Study 5. Do the most productive workers also engage in the most destructive behavior? If so, future research should examine if this is due to managers' fear of reprimanding a superstar, the superstar's sense of entitlement, non-elites covering for the superstar's misbehavior out of hero worship, or some interaction of all three.

Finally, going beyond any individual research domain, a Paretian distribution of performance may help explain why despite more than a century of research on the antecedents of job performance and the countless theoretical models proposed, explained variance estimates (R^2) rarely exceed .50 (Cascio & Aguinis, 2008b). It is possible that research conducted over the past century has not made important improvements in the ability to predict individual performance because prediction techniques rely on means and variances assumed to derive from normal distributions, leading to gross errors in the prediction of performance. As a result, even models including theoretically sound predictors and administered to a large sample will most often fail to account for even half of the variability in workers' performance. Viewing individual performance from a Paretian perspective and testing theories with techniques that do not require the normality assumptions will allow us to improve our understanding of factors that account for and predict individual performance. Thus, research addressing the prediction of performance should be conducted with techniques that do not require the normality assumption.

Implications for Research Methodology

What are the consequences of using traditional Gaussian-based techniques with individual performance data that follow a Paretian distribution? A basic example is a test of differences in means (e.g., independent group *t*-test) for some intervention where individuals are randomly assigned to groups. The assumption is that, given sufficient group sizes, no one individual will deviate from the mean enough to cause a significant difference when there is none or vice versa (Type I and Type II errors). However, random assignment will only balance the groups when the distribution of the outcome is normally distributed (when the prevalence of outliers is low). In the case of Paretian distributions, the prevalence of outliers is much higher. As a result, a single high performer has an important impact on the mean of the group and ultimately on the significance or nonsignificance of the test statistic. Likewise, the residual created by extreme performers' distance from a regression line widens standard errors to create Type II errors. Interestingly, the wide standard errors and unpredictable means caused by extreme performers should result in great variability in findings in terms of both statistical significance and direction. This may explain so many "inconsistent findings" in the OBHRM literature (Schmidt, 2008). Based on the problems of applying Gaussian techniques to Paretian distribution, our first recommendation for researchers examining individual performance is to test for normality. Paretian distributions will often appear highly skewed and leptokurtic. In addition to basic tests of skew and kurtosis, additional diagnostics such as the chi-square test used in the present studies should be incorporated in the data screening stage of individual performance.

Along with testing for normality, our results also suggest that the methodological practice of forcing normality through outlier manipulation or deletion may be misguided. Dropping influential cases excludes the top performers responsible for the majority of the output, and doing so creates a sample distribution that does not mirror the underlying population distribution. As such, sample statistics will bear little resemblance to population parameters. Samples that exclude outliers generalize only to those individuals around the median of the distribution. Therefore, our second recommendation for research methodology is to shift the burden of proof from outlier retention to outlier deletion/transformation. That is, influential cases should be retained in the data set unless there is clear evidence that their value is incorrect (e.g., typographical error) or belong to a population to which the researcher does not wish to generalize. Regardless, the handling of influential cases should always be reported.

An additional implication our findings is that ordinary least squares regression, ANOVA, structural equation modeling, meta-analysis, and all techniques that provide accurate estimates only under a normal distribution assumption should not be used when the research question involves individual performance output. If researchers find that their data are not normally distributed, and they do not artificially truncate the distribution through outlier deletion, then this leaves the question of how to proceed with analysis. Paretian distributions require analytic techniques that are not common in OBHRM but nonetheless are readily available. Techniques exist that properly and accurately estimate models where the outcome is Paretian. Poisson processes are one such solution, and although not well established in OBHRM research, they do have a history in the natural sciences (e.g., Eliazar & Klafter, 2008) and finance (e.g., Embrechts, Kluppelberg, & Mikosch, 1997). In addition, agent-based modeling (ABM) is an inductive analytic tool that operates without the theoretical assumptions that our results debunk (see Macy & Willer, 2002 for a review of ABM). ABM can be used to develop and test theories of superstars in the more fundamental context of autonomous agents independently and in conjunction with others, making decisions based on very simple rules (Bonabeau, 2002). The result is an understanding of performance based on dynamism instead of equilibrium, interdependent agents instead of independent automatons, and nonlinear change instead static linearity.

In addition, Bayesian techniques are likely to provide the greatest applicability to the study of superstars. Beyond moving away from null hypothesis significance testing, Bayesian techniques provide the additional benefit of dealing with the nonlinearity introduced by influential cases because they allow the underlying distribution to be specified a priori (Beyth-Marom & Fischhoff, 1983). Thus, a researcher can test hypotheses without having to assume normality or force it upon the data (Kruschke, Aguinis, & Joo, unpublished data). For example, one can specify that performance follows a Paretian distribution. Bayesian techniques are slowly being adopted in OBHRM and related disciplines (Detienne, Detienne, & Joshi, 2003; Nystrom, Soofi, & Yasai-Ardekani, 2010; Somers, 2001). Regardless of the specific data-analytic approach, our final methodological recommendation is the use of techniques that do not rely on the normality assumption.

Implications for Practice, Policy Making, and Society

Our results lead to some difficult questions and challenges in terms of practice, policy making, and societal issues because they have implications for discussions around equality and merit (Ceci & Papierno, 2005). There are several areas within OBHRM such as employee training and development and compensation that rely on the assumption that individual performance is normally distributed, and any intervention or program that changes this distribution is seen as unnatural, unfair, or biased (Schleicher, & Day, 1998). In evaluation, interventions are deemed successful to the extent that all those who go through them experience improved performance. But, if training makes the already good better and leaves the mediocre and poor performers behind, then this is usually seen as an indication of program faultiness. The Matthew effect (Ceci & Papierno, 2005; Merton, 1968) states that those already in an advantageous position are able to leverage their position to gain disproportionate rewards. It is disproportionate because the perception is that their inputs into a system do not equal the outputs they receive. Training programs that especially benefit elite performers are seen as unfair because they artificially alter the normal curve that is the "natural" distribution of performance. The Matthew effect has been found in a variety of settings (e.g., Chapman & McCauley, 1993; Judge & Hurst, 2008; Sorenson & Waguespack, 2006). Likewise, compensation systems such as pay for performance and CEO compensation are an especially divisive issue, with many claiming that disproportionate pay is an indicator of unfair practices (Walsh, 2008). Such differences are seen as unfair because if performance is normally distributed then pay should be normally distributed as well.

Our results put the usual conceptions and definitions of fairness and bias, which are based on the norm of normality, into question and lead to some thorny and complicated questions from an ethical standpoint. How can organizations balance their dual goals of improving firm performance and also employee performance and well-being (Aguinis, 2011)? Is it ethical for organizations to allocate most of their resources to an elite group of top performers in order to maximize firm performance? Should separate policies be created for top performers given that they add greater value to the organization than the rest? Our results suggest that practitioners must revisit how to balance the dual goals of improving firm performance and employee performance and well-being as well as determine the proper allocation of resources for both elites and nonelites.

Beyond concepts of ethics and fairness, a Paretian distribution of performance has many practical implications for how business is done. As we described earlier, a Pareto curve demonstrates scale invariance, and thus whether looking at the entire population or just the top percentile, the same distribution shape emerges. For selection, this means that there are real and important differences between the best candidate and the second best candidate. Superstars make or break an organization, and the ability to identify these elite performers will become even more of a necessity as the nature of work changes in the 21st century (Cascio & Aguinis, 2008b). Our results suggest that practitioners should focus on identification and differentiation at the tails of the distribution so as to best identify elites.

Organizations must also rethink employment arrangements with superstars, as they will likely be very different from traditional norms in terms of starting compensation, perquisites, and idiosyncratic employment arrangements. Superstars perform at such a high level that makes them attractive to outside firms, and thus even in a recession these individuals have a high degree of job mobility. In an age of hypercompetitiveness, organizations that cannot retain their top performers will struggle to survive. At present, we know very little about the motivations, traits, and behaviors of elite performers. Our work indicates that superstars exist but does not address the motivations, behaviors, and individual differences of the superstar. We see the emerging literature on I-Deals (Rousseau, Ho, & Greenberg, 2006) and core-self-evaluations (Judge, Erez, Bono, & Thoresen, 2003) as potentially fruitful areas of managing and retaining superstars and encourage practitioners to incorporate these literature streams into their work.

Potential Limitations and Suggestions for Future Research

We attempted to establish the shape of the individual performance distribution across a variety of settings and industries. Although we analyzed multiple industries and a variety of performance operationalizations, it is still possible that these samples do not generalize to other occupations. In addition, for the reasons described earlier, most of the data we used do not include performance measures as typically operationalized in I-O psychology research. We expand on these issues next.

Our results clearly support the superiority of the Paretian distribution compared to the Gaussian distribution to model individual performance. In Studies 1, 2, 4, and 5, we found only one sample (NCAA rushing) for which individual performance was better modeled with a Gaussian distribution than a Paretian distribution. However, in Study 3 we found 11 samples favoring a Gaussian model. Note that given a total of 42 samples in Study 3, results still heavily favor a Paretian distribution (i.e., 74% of samples favored a Paretian distribution). However, a closer examination of results from Study 3 may provide some insights regarding conditions under which Gaussian distributions are likely to be found. We acknowledge the speculative nature of the material that follows, and we emphasize that, rather than conclusions, these should be seen as hypotheses and research questions that need to be addressed by future research.

Consider two measurement-related reasons for the potential better fit of a Gaussian distribution. First, a measure of performance may be too coarse to capture differences between superstars and the "simply adequate" (Aguinis, Pierce, & Culpepper, 2009). Specifically, in Study 3, performance was measured as whether an official was elected or not, and the measure did not capture differences among performers such as by how many votes an individual won or lost an election. So, industries and types of jobs for which performance is measured with coarse scales may lead to observed distributions that are Gaussian rather than Paretian. Second, consider situations in which there are constraints imposed on the ratings of performance. As described earlier, ratings of performance, particularly supervisory ratings, are one of the most popular ways to operationalize performance in I-O psychology research. These constraints can distort the shape of the underlying performance distribution when normality is introduced by the scale or rater evaluation training. In these cases, normality is likely to emerge in the observed data regardless of the true shape of the performance distribution.

Now, consider three situations and reasons why the underlying performance distribution, not just observed performance scores, may actually fit a Gaussian as opposed to a Paretian model. First, it may be the case that, in certain industries and certain job types, superstars simply do not emerge. For example, the manufacturing economy of the 20th century strove not only for uniformity of product but also uniformity of worker. Quotas, union maximums, assembly lines, and situational and technological constraints all constrained performance to values close to the mean. Even in jobs without these formal constraints, informal barriers (i.e., norms) existed in ways that limited the emergence of superstars. Productivity norms such as those found in the Hawthorne studies (Roethlisberger & Dickson, 1939) support a normal distribution of performance dependent on informal corrective actions of coworkers. Workers violating the established productivity norms were chastised, bullied, and ostracized into conforming to prescribed output levels. Hence, even organizations outside of the manufacturing sector where there are limited formal constraints to productivity may still fail to see the emergence of superstars and Paretian distributions.

In government, as is the case for the Study 3 samples, similar norms may lead to the curtailment of outliers, resulting in an observed distribution that is Gaussian rather than Paretian. Second, we also speculate that compared to the participants in the other studies, government officials as we had in Study 3 have fewer direct ties between performance and compensation. Pay raises in legislatures are generally voted on and applied equally across all members. Therefore, if a representative in the Alabama legislature wished to receive higher rewards (e.g., money, fame, power), she would either need to increase the compensation of all her fellow representatives or run for a more prestigious office. For samples in Study 3, all but one of those favoring a Gaussian distribution (the Danish Folketing) were lower houses of the legislative branch. Thus, superstar representatives may have quickly moved on to occupy higher-level and more prestigious positions such as becoming senators or governors. Finally, given the nature of work and organizations in the 21st century (Cascio & Aguinis, 2008b), we believe that the Paretian distribution will apply to an increasingly high number of industries, occupations, and jobs. However, industries and organizations that rely on manual labor, have limited technology, and place strict standards for both minimum and maximum production are likely to lead to normal distributions of individual performance. As we move into the 21st century, software engineers, consultants, healthcare workers, and educators make up an increasingly large part of the economy; but, for the foreseeable future, farmers, factory workers, and construction crews will continue to play an important role, and these types of jobs may best be modeled with a normal distribution (e.g., Hull, 1928; Tiffin, 1947). In short, we readily acknowledge that our results are circumscribed to the types of industries and performance operationalizations included in our five studies because these factors may serve as boundary conditions for our findings.

Our results point to the influential role of elite performers (i.e., superstars), which opens new research avenues for the future. First, although we know there are more superstars than a normal curve would suggest, exactly what percentage of workers can be considered superstars has not been established. The classification of superstars is a subjective judgment, and there are no norms to indicate what proportion of workers should be considered elite. Second, research is needed on the deleterious effects of superstars. For example, does the presence of a superstar demotivate other workers to such an extent that total organizational output decreases?

Finally, our research provides information on the performance distribution, but it does not examine what individual characteristics top performers possess nor did it investigate the stability of the top performing group. When and how do these individuals reach the elite group? What is the precise composition of this elite group—do individuals rotate in and out of this group, or once in the top group, they remain in the top for most of their career? What individual, group, and cultural factors predict an individual's membership in the top-performing group over time? Ultimately, certain individuals likely possess abilities and skills that increase the probability of extraordinary performance, but the interactive nature of performance and context suggests that environmental factors and other actors in the network also play a role in determining individual performance (Aguinis, 2009). That is, superstars likely cannot emerge in a vacuum. Top researchers can devote the necessary time to their work because there are others who take on some of their teaching and administrative duties.

Hollywood stars emerge in part because of supporting casts on screen as well as off screen (e.g., agents, managers, publicists).

Concluding Remarks

Much like the general population, we, OBHRM researchers and practitioners, are not immune to “received doctrines” and “things we just know to be true” (Lance, 2011). These issues are “taught in undergraduate and graduate classes, enforced by gatekeepers (e.g., grant panels, reviewers, editors, dissertation committee members), discussed among colleagues, and otherwise passed along among pliers of the trade far and wide and from generation to generation” (Lance, 2011: 281). We conclude that the norm of normality regarding individual performance qualifies to be such a received doctrine because, even when not explicitly stated, it permeates the theory, design, and analysis of OBHRM research as well as OBHRM practices. In contrast, based on five separate studies involving 198 samples including 633,263 researchers, entertainers, politicians, and amateur and professional athletes, our results indicate that individual job performance follows a Paretian distribution. Assuming normality of individual performance can lead to misspecified theories and misleading practices. Thus, our results have implications for all theories and applications in OBHRM and related fields (e.g., I-O psychology, strategic management, entrepreneurship) that directly or indirectly rely upon the performance of individual workers.

REFERENCES

Aguinis H. (2009). *Performance management* (2nd ed.). Upper Saddle River, NJ: Pearson/Prentice Hall.

Aguinis H. (2011). Organizational responsibility: Doing good and doing well. In Zedeck S. (Ed.), *APA handbook of industrial and organizational psychology* (pp. 855–879). Washington, DC: American Psychological Association.

Aguinis H, Boyd BK, Pierce CA, Short JC. (2011). Walking new avenues in management research methods and theories: Bridging micro and macro domains. *Journal of Management*, *37*, 395–403.

Aguinis H, Harden EE. (2009). Cautionary note on conveniently dismissing χ^2 goodness-of-fit test results: Implications for strategic management research. In Bergh DD, Ketchen DJ (Eds.), *Research methodology in strategy and management* (vol. 5, pp. 111–120). Howard House, UK: Emerald.

Aguinis H, Pierce CA, Culpepper SA. (2009). Scale coarseness as a methodological artifact: Correcting correlation coefficients attenuated from using coarse scales. *Organizational Research Methods*, *12*, 623–652.

Andriani P, McKelvey B. (2009). Perspective—from Gaussian to Paretian thinking: Causes and implications of power laws in organizations. *Organization Science*, *20*, 1053–1071.

Anschuetz N. (1997). Profiting from the 80–20 rule of thumb. *Journal of Advertising Research*, *37*, 51–66.

Bernardin HJ, Beatty RW. (1984). *Performance appraisal: Assessing human behavior at work*. Boston: Kent.

Beyth-Marom R, Fischhoff B. (1983). Diagnosticity and pseudo-diagnosticity. *Journal of Personality and Social Psychology*, *45*, 1185–1195.

Bonabeau E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(Suppl. 3), 7280–7287.

Bonardi J. (2004). Global and political strategies in deregulated industries: The asymmetric behaviors of former monopolies. *Strategic Management Journal*, *25*, 101–120.

Bronzin V. (1908). *Theorie der prämiengeschäfte*. Leipzig and Vienna: Verlag Franz Deticke.

Buchanan M. (2004). Power laws & the new science of complexity management. *Strategy Business*, *34*, 1–8.

Canter RR. (1953). A rating-scoring method for free-response data. *Journal of Applied Psychology*, *37*, 455–457.

Cascio WF, Aguinis H. (2008a). Research in industrial and organizational psychology from 1963 to 2007: Changes, choices, and trends. *Journal of Applied Psychology*, *93*, 1062–1081.

Cascio WF, Aguinis H. (2008b). Staffing twenty-first-century organizations. *Academy of Management Annals*, *2*, 133–165.

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

Cascio WF, Ramos RA. (1986). Development and application of a new method for assessing job performance in behavioral/economic terms. *Journal of Applied Psychology*, *71*, 20–28.

Ceci SJ, Papierno PB. (2005). The rhetoric and reality of gap closing: When the "have-nots" gain but the "haves" gain even more. *American Psychologist*, *60*, 149–160.

Chapman GB, McCauley C, (1993). Early career achievements of National Science Foundation (NSF) graduate applicants: Looking for Pygmalion and Galatea effects on NSF winners. *Journal of Applied Psychology*, *78*, 815–820.

Detienne KB, Detienne DH, Joshi SA. (2003). Neural networks as statistical tools for business researchers. *Organizational Research Methods*, *6*, 236–265.

Eliazar I, Klafter J. (2008). Paretian Poisson processes. *Journal of Statistical Physics*, *131*, 487–504.

Embrechts P, Klüppelberg C, Mikosch T. (1997). *Modeling external events*. Berlin, Germany: Springer.

Ferguson LW. (1947). The development of a method of appraisal for assistant managers. *Journal of Applied Psychology*, *31*, 306–311.

Fiol CM, O'Connor EJ, Aguinis H. (2001). All for one and one for all? The development and transfer of power across organizational levels. *Academy of Management Review*, *26*, 224–242.

Gabaix X. (1999). Zipf's law for cities: An explanation. *The Quarterly Journal of Economics*, *114*, 739–767.

Gabaix X, Gopikrishnan P, Plerou V, Stanley HE. (2003). A theory of power-law distributions in financial market fluctuations. *Nature*, *423*, 267–270.

Galton F. (1889). *Natural inheritance*. New York, NY: Macmillan and Co.

Gilliland SW, Cortina JM. (1997). Reviewer and editor decision making in the journal review process. PERSONNEL PSYCHOLOGY, *50*, 427–452.

Gomez-Mejia LR, Balkin DB. (1992). Determinants of faculty pay: An agency theory perspective. *Academy of Management Journal*, *35*, 921–955.

Greenberg J. (2002). Who stole the money, and when? Individual and situational determinants of employee theft. *Organizational Behavior and Human Decision Processes*, *89*, 985–1003.

Harzing A. (2008). *Publish or Perish: A citation analysis software program*. Available from http://www.harzing.com/resources.htm.

Hull CL. (1928). *Aptitude testing*. Chicago, IL: World Book Company.

Jacobs D. (1974). Dependency and vulnerability: An exchange approach to the control of organizations. *Administrative Science Quarterly*, *19*, 45–59.

Judge TA, Erez A, Bono JE, Thoresen CJ. (2003). The core self-evaluations scale: Development of a measure. PERSONNEL PSYCHOLOGY, *56*, 303–331.

Judge TA, Hurst C. (2008). How the rich (and happy) get richer (and happier): Relationship of core self-evaluations to trajectories in attaining work success. *Journal of Applied Psychology*, *93*, 849–863.

Lance CE. (2011). More statistical and methodological myths and urban legends. *Organizational Research Methods*, *14*, 279–286.

Landy FJ, Farr JL. (1980). Performance rating. *Psychological Bulletin*, *87*, 72–107.

Macy MW, Willer R. (2002). From factors to actors: Computational sociology and agent-based modeling. *Annual Review of Sociology*, *28*, 143–166.

Mandelbrot B, Hudson RL, Grunwald E. (2005). The (mis)behaviour of markets. *The Mathematical Intelligencer*, *27*, 77–79.

Mathieu J, Maynard MT, Rapp T, Gilson L. (2008). Team effectiveness 1997–2007: A review of recent advancements and a glimpse into the future. *Journal of Management*, *34*, 410–476.

Merton RK. (1968). The Matthew effect in science: The reward and communication systems of science are considered. *Science*, *159*, 56–63.

Micceri T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, *105*, 156–166.

Motowidlo SJ, Borman WC. (1977). Behaviorally anchored scales for measuring morale in military units. *Journal of Applied Psychology*, *62*, 177–183.

Murphy KR. (2008). Models and methods for evaluating reliability and validity. In Cartwright S, Cooper CL (Eds.), *The Oxford handbook of personnel psychology* (pp. 263–290). New York, NY: Oxford University Press.

Murphy KR, Cleveland J. (1995). *Understanding performance appraisal: Social, organizational, and goal-based perspectives*. Thousand Oaks, CA: Sage.

Nystrom PC, Soofi ES, Yasai-Ardekani M. (2010). Identifying and analyzing extremes: Illustrated by CEOs' pay and performance. *Organizational Research Methods*, *13*, 782–805.

Orr JM, Sackett PR, Dubois CLZ. (1991). Outlier detection and treatment in I-O psychology: A survey of researchers beliefs and an empirical illustration. PERSONNEL PSYCHOLOGY, *44*, 473–486.

Palisades Corporation. (2009). *@RISK 5.5: Risk analysis and simulation*. Ithaca, NY.

Pareto V. (1897). *Le cours d'economie politique*. London, UK: Macmillan.

Powell TC. (2003). Varieties of competitive parity. *Strategic Management Journal*, *24*, 61–86.

Reif F. (1965). *Fundamentals of statistical and thermal physics*. New York, NY: McGraw-Hill.

Reilly RR, Smither JW. (1985). An examination of two alternative techniques to estimate the standard deviation of job performance in dollars. *Journal of Applied Psychology*, *70*, 651–661.

Roethlisberger FJ, Dickson WJ. (1939) *Management and the worker*. Cambridge, MA: Harvard University Press.

Rotundo M, Sackett PR. (2002). The relative importance of task, citizenship, and counterproductive performance to global ratings of job performance: A policy-capturing approach. *Journal of Applied Psychology*, *87*, 66–80.

Rousseau DM, Ho VT, Greenberg J. (2006). I-Deals: Idiosyncratic terms in employment relationships. *Academy of Management Review*, *31*, 977–994.

Saal FE, Downey RG, Lahey MA. (1980). Rating the ratings: Assessing the quality of rating data. *Psychological Bulletin*, *88*, 413–428.

Sackett PR, Yang H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology*, *85*, 112–118.

Salin D. (2003). Ways of explaining workplace bullying: A review of enabling, motivating and precipitating structures and processes in the work environment. *Human Relations*, *56*, 1213–1232.

Schleicher DJ, Day DV. (1998). A cognitive evaluation of frame-of-reference rater training: Content and process issues. *Organizational Behavior and Human Decision Processes*, *73*, 76–101.

Schmidt FL. (2008). Meta-analysis: A constantly evolving research integration tool. *Organizational Research Methods*, *11*, 96–113.

Schmidt FL, Hunter JE. (1983). Individual differences in productivity: An empirical test of estimates derived from studies of selection procedure utility. *Journal of Applied Psychology*, *68*, 407–114.

Schmidt FL, Hunter JE, McKenzie RC, Muldrow TW. (1979). Impact of valid selection procedures on work-force productivity. *Journal of Applied Psychology*, *64*, 609–626.

Schmidt FL, Johnson RH. (1973). Effect of race on peer ratings in an industrial situation. *Journal of Applied Psychology*, *57*, 237–241.

Schneier CE. (1977). Operational utility and psychometric characteristics of behavioral expectation scales: A cognitive reinterpretation. *Journal of Applied Psychology*, *62*, 541–548.

Schultz DG, Siegel AI. (1961). Generalized Thurstone and Guttman scales for measuring technical skills in job performance. *Journal of Applied Psychology*, *45*, 137–142.

Smither JW, London M. (2009a). Best practices in performance management. In Smither JW, London M (Eds.), *Performance management: Putting research into practice* (pp. 585–626). San Francisco, CA: Wiley.

Smither JW, London M. (Eds.) (2009b). *Performance management: Putting research into practice*. San Francisco, CA: Wiley.

Somers MJ. (2001). Thinking differently: Assessing nonlinearities in the relationship between work attitudes and job performance using a Bayesian neural network. *Journal of Occupational and Organizational Psychology*, *74*, 47–61.

Sorenson O, Waguespack DM. (2006). Social structure and exchange: Self-confirming dynamics in Hollywood. *Administrative Science Quarterly*, *51*, 560–589.

Stanley MHR, Buldyrev SV, Havlin S, Mantegna RN, Salinger MA, Stanley HE. (1995). Zipf plots and the size distribution of firms. *Economics Letters*, *49*, 453–457.

Tepper BJ, Henle CA, Lambert LS, Giacalone RA, Duffy MK. (2008). Abusive supervision and subordinates' organization deviance. *Journal of Applied Psychology*, *93*, 721–732.

Tiffin J. (1947). *Industrial psychology* (2nd ed.). New York, NY: Prentice-Hall.

Walsh JP. (2008). CEO compensation and the responsibilities of the business scholar to society. *Academy of Management Perspectives*, *22*, 26–33.

West BJ, Deering B. (1995). *The lure of modern science: Fractal thinking*. Salem, MA: World Scientific.

Yule GU. (1912). On the methods of measuring association between two attributes. *Journal of the Royal Statistical Society*, *75*, 579–652.