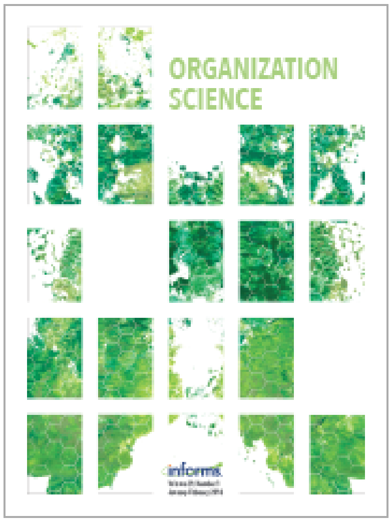


Team Performance: Nature and Antecedents of Nonnormal Distributions

Kyle J. Bradley, Herman Aguinis

informs
http://pubsonline.informs.org/journal/orsc

ORGANIZATION SCIENCE
Articles in Advance, pp. 1–21
ISSN 1047-7039 (print), ISSN 1526-5455 (online)

Team Performance: Nature and Antecedents of Nonnormal Distributions

Kyle J. Bradley,[a,*] Herman Aguinis[b]

[a] Department of Management, College of Business Administration, Kansas State University, Manhattan, Kansas 66506; [b] Department of Management, School of Business, The George Washington University, Washington, District of Columbia 20052
*Corresponding author

Contact: kyljbrad@ksu.edu, https://orcid.org/0000-0003-1970-5505 (KJB); haguinis@gwu.edu, https://orcid.org/0000-0002-3485-9484 (HA)

Received: April 28, 2020
Revised: February 5, 2021; August 18, 2021; June 7, 2022
Accepted: July 11, 2022
Published Online in Articles in Advance: September 8, 2022

https://doi.org/10.1287/orsc.2022.1619

Abstract. Team research typically assumes that team performance is normally distributed: teams cluster around average performance, performance variability is not substantial, and few teams inhabit the upper range of the distribution. Ironically, although most team research and methodological practices rely on the normality assumption, many theories actually imply nonnormality (e.g., performance spirals, team composition, team learning, punctuated equilibrium). Accordingly, we investigated the nature and antecedents of team performance distributions by relying on 274 performance distributions including 200,825 teams (e.g., sports, politics, firefighters, information technology, customer service) and more than 500,000 workers. First, regarding their overall nature, only 11% of the distributions were normal, star teams are much more prevalent than predicted by normality, the power law with an exponential cutoff is the most dominant distribution among nonnormal distributions (i.e., 73%), and incremental differentiation (i.e., differential performance trajectories across teams) is the best explanation for the emergence of these distributions. Second, this conclusion remained unchanged after examining theory-based boundary conditions (i.e., tournament versus nontournament contexts, performance as aggregation of individual-level performance versus performance as a team-level construct, performance assessed with versus without a hard left-tail zero, and more versus less sample homogeneity). Third, we used the team learning curve literature as a conceptual framework to test hypotheses and found that authority differentiation and lower temporal stability are associated with distributions with larger performance variability (i.e., a greater proportion of star teams). We discuss implications for existing theory, future research directions, and methodological practices (e.g., need to check for nonnormality, Bayesian analysis, outlier management).

Supplemental Material: The online appendix is available at https://doi.org/10.1287/orsc.2022.1619.

Keywords: team performance • incremental differentiation • performance distribution

Introduction

Team research relies on the assumption that team performance is normally distributed. If this is true, the majority of teams would cluster around the average level of team performance, performance variability is not substantial, and relatively few teams inhabit the upper range of the distribution. In contrast to this typical normality assumption, several theories currently used in teams research refer to mechanisms that lead to the formation of nonnormal distributions (e.g., performance spirals (Lindsley et al. 1995); team composition theories (Mathieu et al. 2014); team learning (Argote and Epple 1990)). In fact, foundational theories of team performance such as the input-mediator-output-input (IMOI) model (Ilgen et al. 2005) suggest that team performance is highly influenced by processes that lead to major advantages for some teams that continue to build and would lead to the formation of nonnormal distributions. An empirical finding challenging the normality assumption would indicate a need to understand the generative mechanisms that create these distributions that would lead to adjustments in several team theories. For example, theories regarding team learning (Luan et al. 2016) currently do not explain how the possibility of a large proportion of star teams would impact team decisions regarding the choice of external referents (Argote and Ingram 2000). To clarify, we use a prevalent definition for teams as "small groups of interdependent individuals who share responsibility for outcomes" (Hollenbeck et al. 2012, p. 82).

From a methodological standpoint, findings demonstrating the nonnormal nature of team performance distributions can change how past team research is interpreted and future team research is conducted. Specifically, under the assumption of normality, extreme data points are considered uncommon anomalies and can therefore be treated as undesirable errors. Thus, outliers (i.e., star teams) are often deleted or the entire data set is transformed to be able to better fit the normal distribution to comply with general linear model (GLM; ordinary least squares (OLS) regression, structural equation modeling) assumptions such as residual homogeneity (Aguinis et al. 2013, Becker et al. 2019). Thus, it is common practice to ignore (by deleting them) or minimize (by using "robust" approaches that transform and trim data) the impact that extreme observations have on substantive conclusions (Aguinis et al. 2013, Becker et al. 2019). In other words, "squeezing" heavy-tailed distributions through data transformations and manipulations to avoid violating statistical assumptions artificially reduces observed variability in team performance scores and consequently artificially changes the nature of the relation between team performance and other variables. Importantly, the focus of our paper is on the distribution level of analysis, not the team level of analysis, and we therefore address the distribution as a whole by focusing on generative mechanisms that lead to different distribution shapes.

Theoretical Background, Research Questions, and Hypotheses

Ironically, much of the theory underlying empirical team research implies that team performance is *not* normally distributed but instead follows a heavy-tailed distribution. Next, we highlight how heavy-tailed distributions differ from one another and why team theory predicts these distributions.

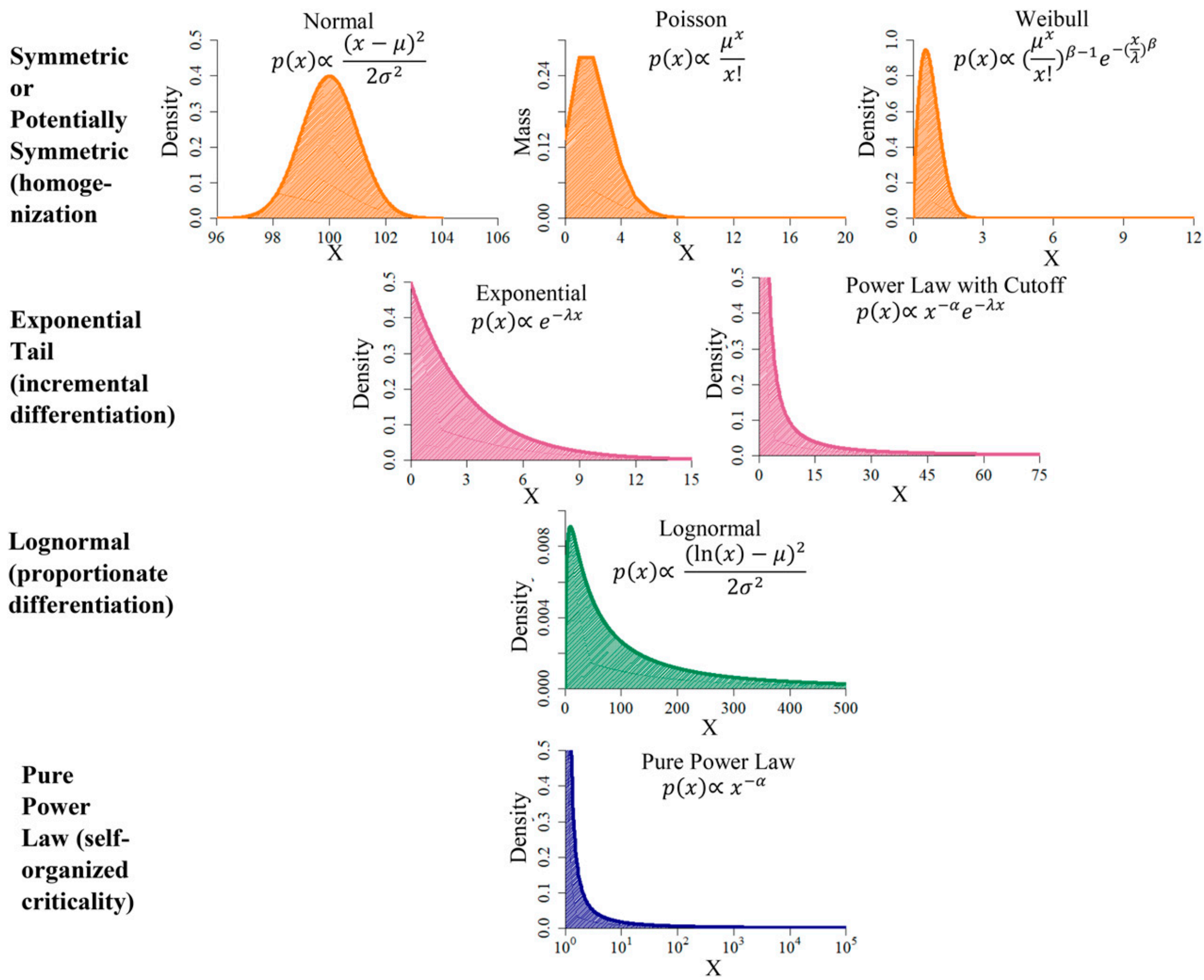
Generating Mechanisms of Team Performance Distributions

We focus on the seven distributions most commonly observed in natural phenomena, which are grouped into four categories (Sornette 2006, Joo et al. 2017): (a) exponential tail (i.e., exponential and power law with an exponential cutoff), (b) lognormal, (c) pure power law, and (d) symmetric or potentially symmetric (i.e., normal or Gaussian, Poisson, or Weibull). Each distribution category results from a distinct, unique, and specific generating mechanism. As a visual aid, Figure 1 includes representations of each of the seven distributions comprising the four categories. From a more technical standpoint, Figure 1 also includes the equations and parameters defining each distribution.

First, *incremental differentiation* is the generating mechanism that results in *exponential tail distributions* (i.e., exponential and power law with exponential cutoff) due to processes that lead to output increments. Under these distributions, star teams are common, but diminishing returns lead to smaller variability between star teams. With incremental differentiation, teams with higher performance trajectories are predicted to eventually rise to stardom, whereas those on lower trajectories do not. For example, teams that can consistently learn and adapt to changing conditions will continue to compound their advantages over other teams. These performance trajectories represent the linear increase in the average amount of output a team is able to produce in a specified time period, resulting in this specific type of distribution. Several team theories support the idea that incremental differentiation takes place such as team learning, which emphasizes the importance of speed of learning on performance (Edmondson et al. 2007). For instance, past research has addressed the rate of learning and its differential impact on team performance depending on factors such as the level of experience of team members (Pisano et al. 2001) and team stability (Reagans et al. 2005). Variance in team learning leads to teams with higher performance trajectories than others, which in turn would lead to the emergence of an exponential tail distribution.

Second, in *proportionate differentiation*, both the initial level of performance for a team and the performance trajectory of the team lead to the generation of *lognormal distributions* due to processes that lead to output loops. Under these types of distributions, star teams are common but variability between star teams is not as great as under other heavy-tailed distributions. With proportionate differentiation, some teams initially have higher levels of performance and continue with a high improvement rate, which results in a high proportion of star teams. Much like self-fueling performance spirals, teams that start off with a high level of performance can leverage their prior performance to continue to build on that success (Lindsley et al. 1995). For example, Banker et al. (1996) reported that teams starting with a high initial level of performance continued to improve over time at a much higher rate than others. Similarly, because of the existence of negative performance spirals (Lindsley et al. 1995), many teams set in motion processes that lead to a very large pile of poorly performing teams that in turn results in more star teams by comparison. Additionally, theories regarding team composition note that the makeup of a team can lead to differences in initial levels of performance and improvement in performance trajectories moving forward (Mathieu et al.

Figure 1. Visual Representation of Seven Types of Distributions Within Four Categories with Generating Mechanisms



Notes. Normal (μ= 100, σ = 1), power law with an exponential cutoff (α = 1.5, λ = 0.37), exponential (λ = 0.5), lognormal (μ = 4.5, σ = 1.5), Weibull (β = 1.8, λ = 0.85), Poisson (μ = 2), and pure power law (α = 1.5). In each of the panels, except the one containing the Poisson distribution, the *x* axis represents values of a continuous variable, whereas the *y* axis ("Density") represents the likelihood of the continuous variable taking on a given value or range of values. In the panel containing the Poisson distribution, the *x* axis represents values of a discrete variable, whereas the *y* axis ("Mass") represents the likelihood of the discrete variable taking on a given discrete value.

2014). Moreover, due to variability in general ability factors that exist both at the individual level (Devine and Philips 2001, Bell 2007) and at the team level (Woolley et al. 2010, Riedl et al. 2021), we would expect that differences in team composition will lead to some teams starting at higher initial levels of performance and being able to learn faster (Aggarwal et al. 2019), which together lead to highly heterogenous levels of performance that would result in a lognormal distribution.

Third, *self-organized criticality* drives the emergence of *pure power law distributions*, such as the pareto distribution (Pareto 1897), because of processes that lead to unpredictable and extremely large output shocks. In the presence of self-organized criticality, some teams produce output until they reach a critical state, where minor events trigger performance improvements that range from small to very large. Under these distributions, star teams are common and variability even among star teams can be extremely large. This mechanism is likely present when critical states are achieved as a team accumulates components that are interconnected. For instance, if a team is concurrently working on multiple projects (i.e., components) that are closely related to each other (i.e., interconnected), a small breakthrough on one project can lead to breakthroughs on multiple projects simultaneously (Simonton 2003). This type of mechanism is part of the punctuated equilibrium model (Gersick 1988) given that teams are hypothesized to progress in a nonlinear fashion due to periods of stasis followed by sudden drastic changes (Chang et al. 2003). In addition, exogenous events may also trigger a shock that impacts learning trajectories. For instance, the global COVID-19 pandemic shifted teamwork to be completed remotely, and some teams were able to capitalize on this shock, whereas others could not. In the presence of the generating mechanism of self-organized criticality, we would expect power law distributions to emerge.

Finally, *homogenization* drives the emergence of *symmetric and potentially symmetric distributions* (i.e., normal, Weibull, and Poisson) due to processes that reduce differences among teams' output. Under these distributions, star teams would be uncommon because the majority would cluster around the average. For example, imitation learning or institutionalized norms dictated by a governing body (e.g., drafting rules in the National Football League) can reduce differences between teams over time. However, there seem to be few team theories that point to this mechanism as being dominant. For homogenization to occur, the initial starting point of teams, their performance trajectory, and the critical states discussed earlier would exert a smaller force in creating team performance distributions compared with the homogenization effect. In our study, we offer empirical evidence that homogenization seems to be the exception rather than the rule regarding the generation of team performance distributions.

As mentioned earlier, the heavy-tailed distributions summarized in Figure 1 are uniquely associated with specific mechanisms that are responsible for the emergence of each category type (Mitzenmacher 2004, Kim and Yum 2008, Andriani and McKelvey 2009, Clauset et al. 2009, Amitrano 2012, Joo et al. 2017, Aguinis et al. 2018). For example, regarding proportionate differentiation, there is a critical role of the initial team performance value (Banerjee and Yakovenko 2010), whereas incremental differentiation generates heavy-tailed distributions where the initial value of performance is unimportant in determining the proportion of teams that rise to stardom level. Additionally, although both proportionate and incremental differentiation generate heavy-tailed distributions through differentiation in team performance trajectories, incremental differentiation recognizes diminishing returns that exist as teams continue to improve their performance. Thus, the presence of a specific distribution shape provides evidence that the associated generative mechanism is the dominant one. However, these generative mechanisms do not necessarily operate in isolation. But, because of their distinct nature, a dominant mechanism overrides the impact of others, leading to the emergence of a specific distribution shape. In sum, multiple mechanisms may exist simultaneously to influence the formation of the team performance distribution shape, but it is possible to identify the most dominant empirically (Clauset et al. 2009, Joo et al. 2017).

It is useful to understand how different structural characteristics can potentially impact the emergence of the varying distributions based on generative mechanisms through an illustration based on team learning. If, for instance, an environment exists where teams can easily learn from one another, we would likely expect homogenization to impact the shape of the distribution. With few constraints impacting a team's external learning behavior, teams would naturally replicate the effective processes and behaviors of the successful teams around them leading to a clustering of teams around an average score—a normal distribution. In contrast, other structural constraints such as the existence of patents or copyrights that limit the access of external learning to other teams are more likely to lead to heavy-tailed distributions. For example, if a team is able to innovate a new process or technology that radically changes an industry while also being protected by a patent, we would expect a pure power law distribution to emerge due to the mechanism of self-organized criticality. Likewise, we would expect another type of heavy-tailed distributions (i.e., lognormal) in environments where external learning is resource intensive, as teams that have the resources necessary to engage in external learning behaviors would experience the feedback loop theorized in proportionate differentiation. Essentially, teams that start with adequate resources available to engage in external learning would be able to improve performance that would provide them with access to more resources to engage in further external learning behaviors, and so on.

Given the aforementioned considerations, our first goal is to assess empirically which distributions emerge as dominant across a broad range of teams and context types. Specifically, is team performance overall characterized by a normal or a heavy-tailed distribution? Moreover, going beyond a simple normal versus nonnormal characterization, is there a best-fitting distribution that arises as dominant for describing team performance? In short, we pose the following question.

Research Question 1. *Which distribution types best describe the overall nature of team performance?*

Theory-Based Boundary Conditions for the Shape of Team Performance Distributions

Even if team performance is overall not normally distributed, there is a need to examine potential boundary conditions (Busse et al. 2017). Next, we consider three potential boundary conditions specifically derived from existing team theory.

Tournament-Like Environments. Although teams generally operate with some level of competition with others, there are certain situations where collaboration between teams is extremely low. This situation is typical in the context of winner-take-all tournaments, where opportunities for collaboration can be limited. However, when teams are not in tournament-like conditions, they often function in a context where between-team collaboration can be high (Le Roy and Fernandez 2015). It is possible that the competitive nature of the team context may serve as a condition that dictates the generative mechanisms present. Specifically, in situations where team competition is low, homogenization may actually be the dominant generative mechanism as teams are able to work more collaboratively, leading to more normal distributions. If, however, teams are in a competitive environment, factors such as the starting level of performance for the teams may create a heavy-tailed distribution due to the proportionate differentiation mechanism. These contexts rely heavily on a ranking of teams to determine a "winner," which puts teams in a dichotomous choice between winning (and others losing) or losing (and others winning) (Sundaresan and Zhang 2012). Because of this, we believe that the competitive environment where teams function may serve as a boundary condition for the existence of heavy-tailed performance distributions. Thus, we ask the following research question.

Research Question 2. *Are heavy-tailed team performance distributions prevalent regardless of whether teams exist in a tournament-like environment?*

Performance Aggregation. Based on the taxonomy of Steiner (1972), there are multiple ways to consider how team tasks are structured, which may also serve as a boundary condition for the team performance distribution shape based on the generative mechanisms present. Specifically, tasks can be additive (i.e., performance is the sum of individual performance), conjunctive (i.e., performance is constrained by the least competent member of the group), disjunctive (i.e., performance depends on the most competent member of the group), or complementary (i.e., performance is determined through the interactions of teams to complete the task). These situations can also be understood as team performance conceptualized and measured as an aggregation of individual-level performance (e.g., additive tasks) or as a team-level construct (e.g., complementary tasks). A question then is whether these two different conceptualizations may serve as a boundary condition. Although situations characterized by an aggregation of individual-level data could result in normal distributions due to the presence of homogenization that occurs when averaging out the performance of individuals on a team, there is also the possibility that individual stars may also drive the formation of heavy-tailed team distributions given their prevalence across teams (Aguinis and O'Boyle 2014). Thus, we ask the following.

Research Question 3. *Are heavy-tailed distributions prevalent regardless of whether team performance is an aggregation of individual-level performance or a team-level construct?*

Performance Constrained by a Hard Left-Tail Zero (i.e., Floor Effect). The presence of a hard left-tail of zero when assessing performance may also be a boundary condition of the shape of the performance distribution. In these situations, homogenization may exert an additional influence over the shape of the team performance distribution. In essence, the majority of teams might cluster around the average and very few teams would be able to inhabit the tails of the distribution. We therefore ask the following.

Research Question 4. *Are heavy-tailed team performance distributions prevalent regardless of whether team performance is constrained by a hard left-tail zero?*

Next, going beyond our examination of (a) the overall pervasiveness of different types of team performance distributions (i.e., Research Question 1) and (b) possible boundary conditions (i.e., Research Questions 2–4), we investigate structural characteristics of teams hypothesized to covary with the heaviness of the distributions' tails (i.e., relative proportion of star teams). In other words, we examine theory-based reasons why some team performance distributions may include greater performance variability and a greater proportion of star teams compared with others.

Structural Characteristics of Teams as Predictors of Heaviness of Distributions' Tails

We used the team learning curve literature as an overarching conceptual framework to examine structural characteristics of teams hypothesized to predict the heaviness of the distributions' tails (i.e., heavier tails include a greater proportion of star teams). This is a useful theoretical framework because it focuses on the impact of team learning rate on outcomes (Edmondson et al. 2007). Specifically, differences in learning rates across teams can lead to extreme performance differences and therefore help us understand the emergence of different distribution types (Bell et al. 2012). In fact, the recursive nature of team learning, which is based on incremental improvements in team performance, is directly related to incremental differentiation as a mechanism that generates heavier tails (Knapp 2010). Additionally, models of team learning recognize the impact of the structure of the teams in influencing team performance (Knapp 2010, Bell et al. 2012).

To accomplish our goal of identifying key structural characteristics within a team learning curve framework, we reviewed the literature on team typologies (Cohen and Bailey 1997, Hollenbeck et al. 2012, Foster et al. 2015). In their review, Hollenbeck et al. (2012) described the three main characteristics of teams that have played a prominent role in the creation of the various classifications: authority differentiation, temporal stability, and skill differentiation. This is a particularly relevant and appropriate taxonomy because each of the three characteristics is an integral part of the team learning process and therefore consistent with our use of team learning curve research as an overarching conceptual framework (Kane et al. 2005, Greer et al. 2011, Ren and Argote 2011). Given team learning curve theorizing, incremental differentiation implies that the performance trajectories differ such that some teams linearly improve their performance at a greater rate than others. Thus, it is the differential trajectories of performance across teams that lead to the generation of large variability and heavy tails. We expect the structural characteristics in team contexts to play an important role in the extent to which those trajectories lead to large variability and a greater proportion of star teams, as described next.

Authority Differentiation. Authority differentiation relates to the centrality of decision-making power within a team (Hollenbeck et al. 2012). Contexts that are low in authority differentiation require teams without a single leader with final authority in decision making; rather, authority is spread out to multiple team members. An example of this type of team context are labor management committees, for which decisions are only made after achieving a unanimous vote of team members (Romme 2004). At the opposite end of the spectrum, contexts high in authority differentiation include teams with a single member who wields control over decision making. Research on team learning curves suggests that differential rates of team learning take place based on how well teams are managed (Edmondson et al. 2007). Also, strong leadership on teams can enhance coordination and can amplify the advantages that some teams have over others (Greer et al. 2018). Because the leader is tasked with making strategic decisions for the team, the rate of team learning and the performance trajectory should be greatly impacted by the decisions of this single individual. Therefore, when a single star performer in terms of both task and team functions is on the team (Volmer and Sonnentag 2011), team performance is positively impacted. As such, when authority differentiation is high, teams that learn to use these star performers in leadership roles are likely to show higher performance trajectories.

Hypothesis 1. *Greater authority differentiation is associated with greater team performance variability and distributions with a greater proportion of star teams.*

Temporal Stability. Temporal stability refers to the stability of teams over time, both in the short term (e.g., team membership changes during the course of a single project) and in the long term (e.g., team stability over the course of many projects; Hollenbeck et al. 2012). In team contexts characterized by high levels of temporal stability, team membership remains fairly constant over time. For example, at the National Aeronautics and Space Administration (NASA), the core engineering team remained largely intact over the course of several of the Apollo missions (Fries 1992). In contrast, construction crews are low in temporal stability because many workers rotate on and off projects (Baiden et al. 2006). Although teams that stay together longer typically demonstrate higher average performance (Gibson and Gibbs 2006), research has not yet addressed the impact of temporal stability on the variability of team performance. The external team learning literature (Bresman 2010) offers insights regarding this issue. Specifically, the rate of team learning and performance can be negatively impacted by frequent changes in team membership (Kane et al. 2005, Edmondson et al. 2007). However, the negative impact on learning and performance can often be mitigated through a number of means (e.g., incoming member knowledge; Kane et al. 2005). Accordingly, lower levels of team stability (i.e., more frequent team member turnover) offer greater opportunities for teams to generate varying levels of learning and performance. In addition, low temporal stability has a negative impact on learning and subsequent performance (Reagans et al. 2005). This is especially true for learning that takes place externally to a team. When teams are higher in temporal stability, they use time and resources to learn from the successes and mistakes of other teams (Bresman 2010), a likely situation in contexts where teams have routinized their efforts due to spending longer periods of time together (Katz 1982). On the other hand, teams in contexts characterized by low temporal stability also have avenues through which they can effectively learn from other teams even when faced with the challenges of shorter team tenures (Vashdi et al. 2013), but not all teams will effectively pursue those avenues, leading to differential performance trajectories over time. It is the increase in variability of team performance that we expect will lead to a greater proportion of star teams in the performance distribution.

Hypothesis 2. *Lower temporal stability is associated with greater team performance variability and distributions with a greater proportion of star teams.*

Skill Differentiation. Skill differentiation refers to the substitutability of individuals within a team based on their learned skills and other personal characteristics (Hollenbeck et al. 2012). For example, hospital operating teams are high in skill differentiation because surgeons, anesthesiologists, and nurses bring unique skills that they have learned to help the team function as a whole (Edmondson et al. 2001). In contrast, accounting teams are comprised of individuals who have learned similar skills and therefore work in a context that is low in skill differentiation. We posit that contexts characterized by high skill differentiation are likely to result in distributions with large variability and a greater proportion of star teams than those characterized by low skill differentiation. Specifically, specialization of individual members of teams can lead to performance enhancements due to a reduction in cognitive load, an increase in the available knowledge for a team, and a reduction of redundancies that can hamper performance (Hollingshead 1998; Bell et al. 2012). Thus, we offer the following hypothesis.

Hypothesis 3. *Greater skill differentiation is associated with greater team performance variability and distributions with a greater proportion of star teams.*

Finally, because little evidence exists that suggest which of these characteristics would be most important, we also investigated the relative importance of each and offer the following research question.

Research Question 5. *Are team performance variability and the proportion of star teams in a distribution more strongly associated with authority differentiation, temporal stability, or skill differentiation?*

Method

Data and Measures

Sample. Answering our five research questions and testing our three hypotheses requires large data sets because *our level of analysis is not the team, but the team distribution* (i.e., samples of teams). Therefore, we first identified possible archival sources of data using Internet searches. These included data from academic journal teams, political teams, and a wide range of miscellaneous teams that do not fit into a single category (e.g., pub trivia teams, video game teams, firefighter teams). In addition, we relied on sports teams because data collected over decades provide an excellent source of many different measures of team performance (Day et al. 2012). We gathered data from a wide variety of sports teams that represent different contexts and types of competitions to enhance generalizability (Day et al. 2012). Additionally, we intentionally used more than one type of performance indicator when available (e.g., winning percentages, goal differentials). Also, in some cases we gathered data focusing on a more specific measure of team performance (e.g., National Football League touchdowns). However, we also included indicators of overall team performance as well (e.g., winning percentage). In sum, the samples and variables we chose to study are varied, cover multiple dimensions of team performance, and reflect a wide range of team types which allows for a better understanding of a potentially generalized phenomenon.

To further enhance generalizability, we also collected data from more traditional work environments by reviewing articles published in the last three years. Our goal was not to engage in a comprehensive data collection effort; rather, our purpose was to gather some additional evidence regarding the generalizability and robustness of the results. We focused on three journals that publish team-related research (i.e., *Journal of Management, Journal of Applied Psychology,* and *Academy of Management Journal*). Moreover, we focused on studies measuring team performance using objective measures of team output. We identified 12 articles that could serve as additional data sources and contacted the authors to gather data on the performance score and sample size for each team used in their studies with the guarantee that we would not use the data for any other purpose or share them with anyone else. These efforts led to data sets from five research teams from multiple projects.

Overall, we collected performance data on a total of 274 performance distributions including 200,825 teams and more than 500,000 workers. Table 1 provides a detailed description of the sources and data we used. This table is organized into groups and provides a description of the context of the various teams including sports teams, academic journal teams, politics teams, and miscellaneous teams (e.g., firefighter teams, information technology (IT) virtual teams, customer service teams).[1] Additionally, the samples chosen represent variability regarding procedures used to collect the data (i.e., not just convenience samples). For example, data were sampled mostly randomly for general organizational teams by Rego et al. (2019) and Van Bunderen et al. (2018); sales teams by Ahearne et al. (2010); virtual supply chain teams by Maynard et al. (2012); IT virtual teams by Maynard et al. (2019); IT development teams by Rapp and Mathieu (2019); and customer service teams by Mathieu et al. (2006) and Rapp et al. (2016). Small sample size is unlikely to serve as a competing explanation for observed deviations from normality because only 3 of the 274 distributions have a sample size lower than 30. In the interest of full transparency and replicability, we make all our data files available upon request (except for those that were shared with us by authors of published articles).

Structural Characteristics of Teams. Given our focus on the distribution level of analysis, rather than individually investigating each of the 200,825 teams in our sample, we were interested in defining the structural characteristics that define each type of team based on the context in which they operate. As such, we followed the procedure outlined in the Team Descriptive Index Short Form (Lee et al. 2015) with minor changes to indicate our focus on the context. We created a

Table 1. Team Type, Number of Distributions, Average Number of Teams per Distribution (n), and Performance Measures

Team type	Number of distributions	n	Performance measure	Description of performance measure
Sports				
Soccer teams	5	94	Tournament participation	Number of times teams qualified to play in major tournaments
	1	77	Average goals scored	Average number of goals scored per game over a tournament
	1	130	Goal differential	Number of goals scored for – number of goals scored against
NCAAF football teams	5	120	Rushing TDs	Number of touchdowns scored on runs over a season
	5	120	Passing TDs	Number of touchdowns scored on passes over a season
	1	119	Total Offense	Average number of yards gained per game over a season
NHL teams	5	30	Goals scored	Number of scored goals over a season
	1	30	Goals per game	Average number of goals scored per game over a season
	1	30	Goals for/against ratio	Number of goals scored for – number of goals scored against
	1	30	Team wins	Number of wins achieved over a season
	1	30	Power play percentage	Number of power play goals/total number of power plays
MLB teams	2	30	Home runs	Number of homeruns a team makes over a season
	2	30	RBIs	Number of runs-batted-in over a season
	1	30	Team slugging percentage	(Total number of bases reached/total number of at bats) over a season
	1	30	Double plays	Number of defensive plays of two put outs over a season
	1	30	Team batting average	(Total number of hits/total number of at bats) over a season
	1	30	Defensive efficiency ratio	Rate at which balls put into play are converted into outs by a defense
ATP doubles tennis teams	1	178	Total points	Total points gained through advancing in competitions
Ragnar relay teams	18	312	Race time	Total time to complete 200-mile race
ProCycling teams	3	18	World Tour stage wins	Number of stage wins over a season
NCAA bowling teams	1	140	Team pin totals	Number of pins knocked down over a season
Journals				
Journal editorial teams	2	378	SAGE impact factor	(Number of citations/number of articles published) over two years
	27	1,676	Scimago journal rank	Average number of weighted citations received in a given year by articles published over previous three years
	27	1,676	Scimago H index	Number of articles (h) that have received at least h citations
	27	1,676	Scimago citations/document	Average number of citations per document in the journal
Politics				
State campaign teams	51	27	Campaign donations	Total dollars donated over the course of a campaign
U.S. Congress	2	22	Reported bills %	Number of bills passed/total bills introduced in committees
Miscellaneous				
Pub trivia teams	29	1,590	Team points	Total points scored in a single night of trivia
Auto engineering teams	2	113	Competition score	Total team points gained in auto building competition
Bridge engineering teams	4	46	Competition score	Total team points gained in bridge building competition
Video game teams	4	269	Team points	Total team points gained through competitions
Firefighter teams	10	214	Race times	Total time to complete firefighter skills course
	3	103	Turnout time	Total time to leave the station after receiving a call
	3	103	Travel time	Total time to travel from the station to the site of an emergency
Movie production teams	6	100	Gross earnings ($)	Total dollars earned for the length of a movie in theaters
Indiegogo teams	1	266	Total donations	Total money pledged for Indiegogo campaign
	1	266	Percentage of goal met	Percent of goal met for Indiegogo campaign

Table 1. (Continued)

Team type	Number of distributions	n	Performance measure	Description of performance measure
General organizational teams	4	88	Leader reported performance	Leader rated response regarding team outcomes
Sales teams	1	230	Percentage of goal reached	Percentage of sales goal reached
Virtual supply chain teams	1	61	Effectiveness rating	Leader rated ability to meet effectiveness targets
IT virtual teams	1	63	Effectiveness rating	Leader rated ability to meet effectiveness targets
IT development teams	1	83	Team performance rating	Leader rated response regarding team outcomes
Customer service teams	1	122	Machine	Average number of copies made between service calls
	1	122	Parts	Percentage of budget associated with replacing parts
	1	122	Response	Average length of time between call and arrival of team
	1	122	Customer satisfaction	Ratings of customer satisfaction
	1	71	Parts	Percentage of budget associated with replacing parts
	1	71	Response time	Average length of time between call and arrival of team
	1	71	Performance composite	Composite score for team response times, parts, and calls
Mining communities of practice	1	33	Leader rated performance	Composite rating by leaders on team effectiveness
Business simulation teams	1	516	Balanced scorecard ratings	Composite of finance, processes, growth, and service indicators
Total	274	200,825		

Notes. n = number of teams per distribution—for cases for more than one distribution, n is the average. NCAA, National Collegiate Athletic Association; NCAAF, National Collegiate Athletic Association Football; NHL, National Hockey League; MLB, Major League Baseball; ATP, Association of Tennis Professionals; TD, touchdowns; RBI, runs batted in. Data set sources: soccer teams: www.soccerstats.com; NCAAF football teams, NHL teams, MLB teams: www.sports-reference.com; ATP tennis teams: www.atpworldtour.com; Ragnar relay teams: www.webscorer.com/ragnar; ProCycling teams: www.procyclingstats.com; NCAA bowling teams: www.collegebowling.com; Journal editorial teams: us.sagepub.com/en-us/nam/impact-factor-ranking-results, www.scimagojr.com; State campaign teams: www.opensecrets.org; U.S. Congress: www.house.gov, www.senate.gov; Pub trivia teams: www.triviakings.com/results; Auto engineering teams: students.sae.org/cds/formulaseries/results/; bridge engineering teams: nssbc.com; firefighter teams: firefighterchallenge.com, lafd.org/fsla/stations-map; video game teams: gosugamers.com; movie production teams: boxofficemojo.com; Indiegogo teams: indiegogo.com; general organizational teams: Rego et al. (2019), Van Bunderen et al. (2018); sales teams: Ahearne et al. (2010); virtual supply chain teams: Maynard et al. (2012); IT virtual teams: Maynard et al. (2019); IT development teams: Rapp and Mathieu (2019); customer service teams: Mathieu et al. (2006), Rapp et al. (2016); mining communities of practice: Kirkman et al. (2011); business simulation teams: Dierdorff et al. (2019).

coding protocol that included definitions and examples for each of the three characteristics in Hypotheses 1–3 and Research Question 5. Three graduate student coders with training on team theory were presented with the following task: "Within each of the occupations listed below there is considerable variance regarding each of the three dimensions. What we ask you to do is to capture what a *typical* team in each occupation experiences in terms of authority differentiation, temporal stability, and skill differentiation. Rate each sample on these three dimensions on a Likert scale ranging from '1' (very low on this dimension) to '5' (very high on this dimension). Again, as you go through the coding, try to picture what a *typical* team looks like before assigning a code to each sample." The three coders had significant personal experience working with a wide range of teams (e.g., military teams, sports teams, student teams).

Coders were then instructed to rate each of the 274 distributions on each of the three characteristics. They first independently coded a subsample of the teams to check for agreement. Intraclass correlation (ICC(2)) levels were acceptable for authority differentiation (0.73) and skill differentiation (0.75), with a lower value for temporal stability (0.65) (LeBreton and Senter 2008). After receiving additional training, coders then continued with the full sample of teams. Results of this second round of coding showed acceptable ICC(2) for authority differentiation (0.84), temporal stability (0.74), and skill differentiation (0.81). These ICC levels are adequate for the purpose of this project as shown by the statistically significant results we observed (i.e., results are unlikely to be statistically significant in the presence of substantial measurement error). We used averages of the three raters for each characteristic in the analyses. Table S1 in the online appendix shows the scores for authority differentiation, temporal stability, and skill differentiation for each of the sample types.

Criterion Used in Hypothesis Testing: Performance Variability and Proportion of Star Teams in a Distribution. We used parameters from the power law with exponential cutoff (PLC) distribution to assess performance variability and the proportion of star teams in a distribution for testing Hypotheses 1–3 and answering

Research Question 5. Specifically, a set of values follows a PLC distribution if it fits the following probability distribution (Joo et al. 2017):

$$p(x) \propto x^{-\alpha} e^{-\lambda x}, \tag{1}$$

where Euler's number e ≈ 2.718, and alpha (α) and lambda (λ) are parameters indicating the rate of decay that dictate the proportion of star teams. Although both α and λ affect the distribution shape, λ is the stronger parameter and has a more dominant impact on the overall proportion of star teams (Joo et al. 2017). Moreover, using α instead of λ would not change substantive conclusions because $\alpha = 1 + 1/\lambda$ (Hanel et al. 2017). Therefore, we focused on λ as the indicator of variability and the proportion of star teams (i.e., greater variability and proportions are associated with smaller λ values). Finally, the λ parameter was the most appropriate choice given that, as we describe later, PLC was overwhelmingly dominant (i.e., 73% of nonnormal distributions). Moreover, λ also is a parameter used in the equation describing exponential distributions (see equation in Figure 1). Combining PLC and exponential distributions shows that 84% of the nonnormal distributions are accurately described by this particular parameter. Moreover, λ also captures the rate of decay in other types of distributions because exponents of power laws can also be estimated from frequency distributions (Hanel et al. 2017).

Data Analysis

Distribution Pitting Methodology. We used a novel methodological approach in team research called distribution pitting to answer Research Questions 1–4 (Joo et al. 2017). Distribution pitting compares observed distributions to each of the seven theoretical distributions shown in Figure 1 and computes fit indices for each. Then, using a falsification approach, it compares each distribution's fit index to the other distributions' fit indices (e.g., normal versus pure power law, normal versus exponential) to identify the best fitting distribution. Therefore, if in any of these comparisons a distribution is found to be a worse fitting distribution, it is ruled out as the dominant one. This methodology has proven to be accurate in the past (Joo et al. 2017) and allows for distribution-level analyses. Additionally, distribution pitting allows us to identify the proportion of star teams in each distribution. For each sample we made 21 pairwise comparisons of distribution fit: 7!/(2![7–2]!). Following the same procedures as Joo et al. (2017), we implemented three decision rules for identifying the best fitting distribution for each of the 274 distributions. First, for each of the 274 distributions, we compared loglikelihood ratios with their associated p values for each of the 21 comparisons using a cutoff of 0.10 as a conservative cutoff score (Clauset et al. 2009). Second, we used the principle of parsimony to further eliminate distributions with the greater number of parameters when there were two nested distributions remaining (e.g., power law and PLC). Third, many of the distributions are "flexible" in that they approximate other distributions when using certain parameter values. Therefore, we again used the principle of parsimony and opted for the distribution with fewer possible distribution shapes (e.g., the inflexible distributions). In the case of nested distributions, the one with the largest number of parameters always fits at least as well as the distribution with fewer parameters (Virkar and Clauset 2014). However, this increased precision of fit comes at the cost of lower external generalizability (Joo et al. 2017). Accordingly, the second decision rule errs on the side generalizability and relies on the more parsimonious distribution. We conducted all analyses with the Dpit package in R.

Hypothesis Testing. We conducted analyses involving relations between the three predictors (i.e., authority differentiation, temporal stability, and skill differentiation) and the λ values (indicating team performance variability and the proportion of star teams). Although Pearson's r is the most frequently used correlational measure, results can be biased when variables are not normally distributed (de Winter et al. 2016). We performed distribution pitting analysis on the distribution of λ parameters and found that the PLC distribution was the best fitting one. Accordingly, given the nonnormal nature of the λ distribution, we used Spearman correlations (r_S) instead of Pearson's rs to test our hypotheses.

Results

Table 2 provides a summary of results regarding the number of times each distribution was identified as the dominant one after implementing each of the three decision rules. In the interest of full transparency and replicability, the online appendix includes the following additional and more detailed tables: Table S2 shows the dominant distribution after implementing each of the three decision rules sequentially for each of the distributions, Table S3 offers detailed distribution pitting results for four illustrative samples (i.e., the first four listed in Table S2), and Table S4 shows a detailed summary of results based on the broad categories of teams: sports, academic journals, politics, and miscellaneous.

Research Question 1: Overall Nature of the Team Performance Distribution. As shown in Table 2, after implementing all three decision rules, only 11% (30 of 274) were best described by a normal distribution. In contrast, 63% (172 of 274) were best described by one

Table 2. Summary of Results from Distribution Pitting Methodology Identifying the Most Dominant Distribution

	After rule 1		After rules 1 and 2		After all three rules	
Distribution	Count	Percentage	Count	Percentage	Count	Percentage
Heavy-tailed distributions (total)	**64**	**23**	**69**	**25**	**172**	**63**
Power law with exponential cutoff	36	56	36	52	126	73
Lognormal	17	27	17	25	17	10
Exponential	0	0	5	7	18	11
Weibull	11	17	11	16	11	6
Power law	0	0	0	0	0	0
Poisson	0	0	0	0	0	0
Normal (total)	**0**	**0**	**0**	**0**	**30**	**11**
Undetermined (total)	**210**	**77**	**205**	**75**	**72**	**26**
Lognormal	88	42	88	43	51	71
Normal	78	37	78	38	48	67
Weibull	111	53	94	46	27	38
Power law with exponential cutoff	122	58	103	50	10	14
Poisson	114	54	113	55	8	11
Exponential	21	10	16	8	3	4

of the heavy-tailed distributions, and the remaining 26% (72 of 274) were classified as undetermined because more than one plausible distribution remained after implementing the distribution pitting procedures. In other words, we were able to empirically identify a dominant distribution for 74% of the distributions. In addition, we conducted a chi-square analysis to formally test whether the number of distributions is equally represented by normal compared with nonnormal shapes (i.e., 50% of the 274 distributions expected to follow normality and 50% expected to not follow nonnormality). This is an extremely conservative test given that most team research assumes normality, and hence, a more realistic test would be close to 100% of the distributions are expected to follow normality. Even based on a very conservative 50-50 test, results showed a clear dominance of nonnormality: χ^2 (1, n = 274) = 50.82, $p = 1.014 \times 10^{-12}$.

The PLC accounted for 73% (126 of 172) of the heavy-tailed distributions. Table 2 further breaks down the heavy-tailed distributions to give a more detailed count of each of the seven distributions summarized in Figure 1. Overall, results provided evidence about the relative rarity of normal distributions and the dominance of nonnormality, particularly the PLC distribution. To offer a visual representation of our results, Figure 2 includes examples of four observed distributions (i.e., PLC, lognormal, and exponential) overlaid with a normal distribution.

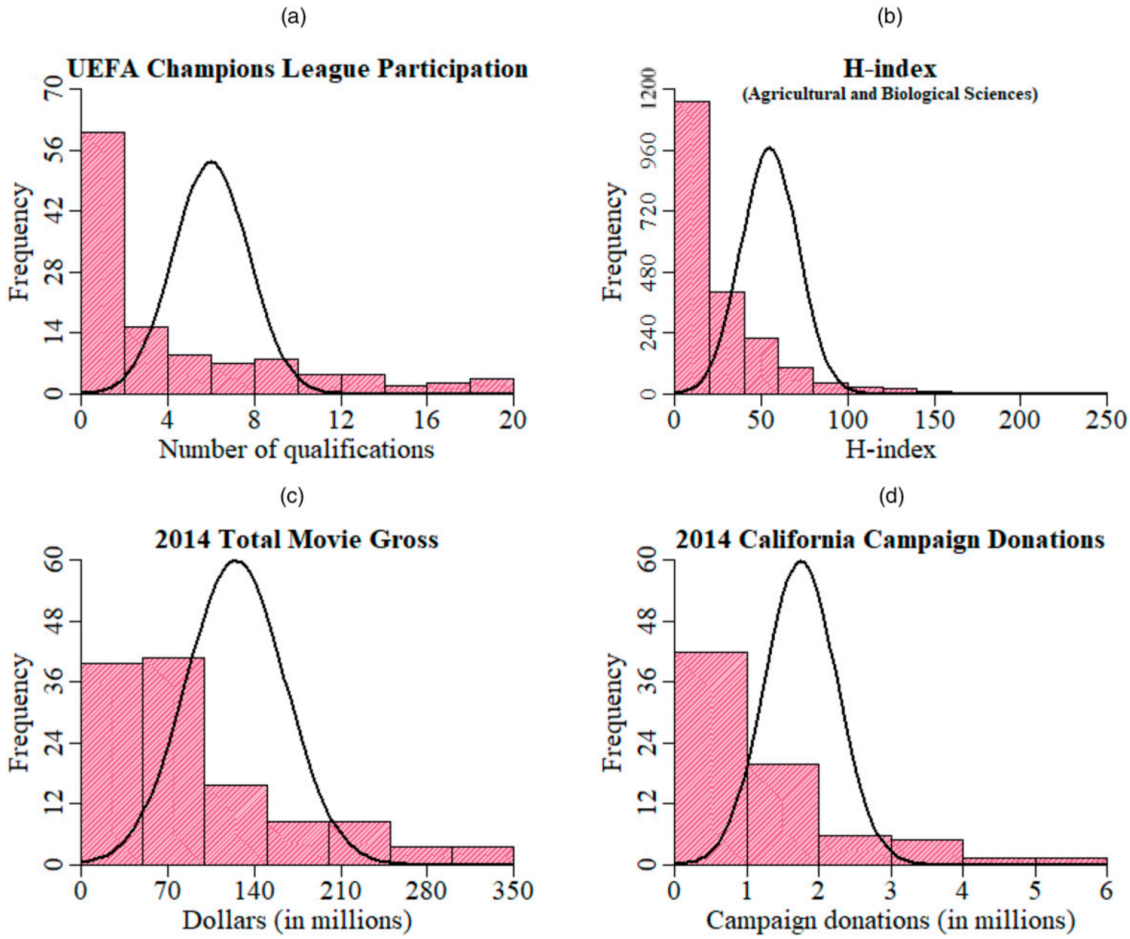
In addition, because journal editorial teams contributed 30% of the distributions, we conducted a subgrouping analysis comparing them to the other samples. Results shown in Table S4 in the online appendix demonstrate that nonnormal distributions are dominant regardless of the context in which teams operate. Specifically, for journal editorial teams, 95% (79 of 83 distributions) followed a heavy-tailed distribution. Similarly, 98% of political teams followed a heavy-tailed distribution (52 of 53 teams).

Despite the dominance of nonnormality, we were intrigued by the few distributions for which normality was the dominant one. As shown in Table S4 in the online appendix, of the 30 distributions that classified as normal, 28 are from the miscellaneous team category. Moreover, of these 28 distributions, 25 came from the same pub trivia team category indicating that factors unique to this specific and particular context are driving the dominance of the normal distribution.

Research Question 2: Tournament vs. Nontournament Contexts as a Boundary Condition. Table S5 in the online appendix shows that teams in a tournament setting included all of the sports samples, most political samples, and several samples from the miscellaneous category (e.g., pub trivia teams). Teams included in the nontournament subgroup included the journal editorial board samples and several of the miscellaneous samples (e.g., IT professionals). For the tournament subgroup, the normal distribution accounted for only 19% of the distributions (30 of 158). Of the 75 nonnormal distributions, the PLC was dominant in 83% (i.e., 62 of 75). For the nontournament subgroup, the normal distribution accounted for 0% of the distributions (0 of 116). Of the 97 nonnormal distributions, the PLC distribution accounted for the majority of distributions with 66% (64 of 97). Therefore, regardless of tournament context, the normal distribution is not the best descriptor of team performance distributions.

Research Question 3: Team Performance Conceptualization and Operationalization as a Boundary Condition. Of the 274 distributions, 208 included team level performance measures based on truly collective constructs (e.g., number of wins for a National Hockey

Figure 2. Visual Representation of Four Illustrative Empirically Observed Distributions Overlaid with a Normal Distribution



Note. (a) PLC distribution, (b) lognormal distribution, (c) exponential distribution, and (d) PLC distribution.

League team in a season), whereas 66 were based on aggregation of individual level data (e.g., number of home runs hit by a Major League Baseball (MLB) team in a season). Of the 208 distributions where performance was a team level construct, 162 (78%) were best described by a nonnormal distribution. For these 162 nonnormal distributions, the PLC was the most dominant in 120 (i.e., 74%). Therefore, results regarding the dominance of nonnormal distributions, and specifically the PLC, are not due to aggregating individual level data to the team level of analysis.

Research Question 4: Hard Left-Tail of Zero in Measuring Team Performance as a Boundary Condition. We calculated the number of distributions that included zero and found that 169 of the 274 distributions (i.e., 61.68%) did not. For example, this hard left-tail of zero did not exist for National Collegiate Athletic Association Football football teams (i.e., none of the teams had an average of zero touchdowns scored on runs over a season or touchdowns scored on passes over a season), or sales teams (i.e., none of the teams met zero percent of their sales quota). Of the 105 distributions that did contain a hard-left tail of zero, 97 (i.e., 92%) were best described by a heavy-tailed distribution and 79 (i.e., 81%) of those fit a PLC distribution. Similarly, of the 170 distributions that did not contain a hard-left of zero, 75 (i.e., 44%) were best described by a heavy-tailed distribution versus only 27 (i.e. 16%) by a normal distribution. Of the heavy-tailed distributions, 47 (i.e., 63%) were best described by a PLC distribution. Therefore, results regarding the prevalence of nonnormality are not explained by the presence of a hard left-tail of zero.

Additional Post Hoc Internal Validity Evidence: Effect of Sample Homogeneity. Although we made an effort regarding sample and performance measure diversity to enhance generalizability (i.e., external validity), we were also concerned about establishing evidence regarding internal validity. Accordingly, to examine yet another possible boundary condition for our results, we created highly homogenous subgroups by randomly selecting five years of data from MLB (total of 20 distributions). We focused on the following performance indicators: total wins, total runs, total runs batted in, and total bases. Table S6 in the online appendix shows that 0% of the distributions are best described by normality (0 of 20). Non-Gaussian distributions were either dominant or codominant for all 20 distributions.

Hypotheses 1–3: Structural Characteristics of Teams as Predictors of Heaviness of Distributions' Tails. Smaller values for λ, which is the parameter describing the rate of decay that dictates the proportion of star teams, indicate more distribution variability (i.e., greater proportion of stars). Therefore, a negative correlation suggests that as the structural team characteristic increases, there is more variability. Hypothesis 1 predicted that greater authority differentiation would be negatively related to variability. Results provided support for this hypothesis: r_S (274) = −0.15, $p < 0.01$. That is, as authority differentiation increases, performance variability and the proportion of star teams in the distribution increases. As an example, heavier tails emerge in distributions of firefighter team performance where there is a clear commander in charge of decision making. Hypothesis 2 was also supported because results demonstrated a positive correlation between temporal stability and λ: r_S (274) = 0.66, $p <$ 0.01. In other words, lower levels of temporal stability were associated with distributions with greater variability (i.e., greater proportion star teams). As an example, distributions with lighter tails emerge in National Collegiate Athletic Association (NCAA) football teams because there is generally a multiyear commitment from players to a school and stringent rules governing college athlete transfers that generate long-term commitment. Finally, the correlation between skill differentiation and the scaling parameter was not significantly different from zero: r_S (274) = −0.09, p = 0.18. Thus, Hypothesis 3 was not supported.

Research Question 5: Relative Importance of the Three Structural Characteristics. A comparison of Spearman correlations showed that the effect of temporal stability was more than four times as large as the effect of authority differentiation (i.e., $|0.66|$ versus $|0.15|$). Therefore, temporal stability was the strongest predictor of distribution variability and the proportion of star teams.

Discussion

We examined 274 team performance distributions from a wide range of industries, occupations, and contexts. Results showed that only 11% of the samples were best described by a normal (i.e., Gaussian) distribution. In contrast, distribution pitting methodology results uncovered that 63% of the distributions were clearly nonnormally distributed, and 73% of these were best described by a power law with exponential cutoff distribution. For 26% of the distributions, results were undetermined in that there was not a single dominant distribution and some of the distributions shapes cannot be completely ruled out with certainty. An examination of possible boundary conditions showed that the dominance of nonnormality was replicated regardless of whether teams are in tournament versus nontournament contexts, whether performance was conceptualized and measured as an aggregation of individual-level performance or as a team-level construct, whether performance was measured including a hard-left of zero, and whether samples were more or less homogeneous. Regarding predictors of heaviness of distributions' tails, results showed that authority differentiation and temporal stability are associated with distributions with greater variability (i.e., a greater proportion of star teams). A comparison of these two showed that temporal stability had the largest effect.

Implications for Existing Organization Science Theory

First, from a descriptive perspective, the empirical discovery regarding the overall nonnormal nature of the team performance distribution provides a more accurate description of reality. To use a metaphor that an anonymous reviewer shared with us, our results show that "the world is round, not flat." Specifically, there is a much greater difference in performance levels across teams than assumed based on the normal distribution. Consider predictions of extreme scores based on how many teams would be found three standard deviations (SDs) to the right of the mean using our team performance distributions. For the distribution of engineering journal editorial teams, whereas a normal distribution would only predict 7 of the 5,339 teams to achieve a count of two citations per published article, our results revealed a total of 90. Therefore, our results show that, based on the type of the team performance distribution, it is not appropriate to discount star teams as being "anomalies." Rather, there is a need to recognize and expect some teams to vastly outperform others. Some theories of team performance mostly focus on the variability of team performance, but this variability is believed to be limited because the majority of teams are assumed to cluster around the average (Mathieu et al. 2000, DeShon et al. 2004). Our results based on team performance distributions suggest the need to focus not only on the average and on the (limited) variability of team performance, but more specifically on how team theories can adequately account for the teams that exist in the upper end of the performance distribution. Certainly, star teams may exist even in the absence of thick tails and there are situations where team capabilities may not be the major driver in the emergence of these stars (e.g., luck, unintentional performance; Vancouver et al. 2016). However, the presence of heavier tails than previously believed provides a strong empirical basis for investigating teams that appear to have found a "winning formula." In fact, research on teams typically does not investigate stars (i.e., "outliers"). Instead, it asks questions such as (Mathieu et al. 2019): What are the factors that explain variance in team performance (i.e., why do some teams perform better than others, distinguishing between low and high performing teams)? However, given the finding that team performance is not normally distributed, team theories should also address questions such as: What are the antecedents leading to the emergence of star teams and how are these outlying teams, which are more common than previously assumed, qualitatively different from others?

Second, it is ironic that, although many contemporary team theories actually predict the heavy-tailed nature of team performance (e.g., performance spirals, team composition, team learning, punctuated equilibrium), team researchers do not seem to acknowledge these theories. Otherwise, contemporary team research would not use procedures and methods that assume normality or transform nonnormality away. This is a substantive rather than a trivial methodological detail because simulation studies show that by relying on the assumption of normality when data are actually heavy-tailed introduces a large amount of bias—and the amount of bias increases with greater deviations from normality (de Winter et al. 2016). As an illustration of the meaning of our results for team theory, consider a recently published meta-analysis that investigated the relationship between transactive memory systems and team performance[2] and reported $r^2 = 0.152$. Our results showed that 63% of distributions are heavy-tailed. Based on large-scale Monte Carlo simulation results, consider a conservative amount of bias of 24% due to nonnormality (de Winter et al. 2016). If 63% of the samples in this published meta-analysis are nonnormal, the resulting meta-analytic r^2 would be 0.108 instead of 0.152. We derive two notable implications from re-examining this published meta-analysis in light of our results. First, the coefficient of determination now likely falls outside of the 90% confidence interval, meaning that we can no longer conclude with confidence that there is a nonzero relation between transactive memory systems and team performance. Second, the corrected coefficient of determination means that only 11% of variance in performance is explained by transactive memory systems, in contrast with the original conclusion that 15% of variance is explained. This represents a decrease in 26.66% in the size of the effect. We emphasize again that this is a conservative corrected estimate and the difference in effect sizes would increase with increased nonnormality (i.e., increase in the thickness of the tail). Therefore, by scrutinizing past empirical research based on whether the normality assumption may have been tenable, it is likely that many past substantive conclusions may need to be revised and updated.

Additional Implications for Future Research Directions

Although our study focused on the distribution level of analysis, our results have implications for future research addressing the distribution as well as the team level of analysis. An especially salient implication for theory and opportunity for future research lies in the variability that exists in distribution shapes even when teams are ostensibly created for similar purposes. For instance, consider journal editorial teams. Although editorial teams may all have a similar goal (i.e., publish high-quality articles), different generative mechanisms are present depending on constraints that exist due to different disciplines and fields of study. For example, open-access publishing has recently become more common across all domains of scholarly research; however, it is more common among the sciences, and science, technology, engineering, and math (STEM) in particular, compared with the humanities and social sciences (Gross and Ryan 2015). With access to a larger number of journals for publishing, incremental differentiation may be a driving mechanism as journals that are able to maintain a higher performance trajectory will generate heavy-tailed performance distributions. Similarly, fields that generally publish papers that require resource-intensive studies as the norm (e.g., biomedical clinical trials; de la Torre Hernández and Edelman 2017) especially emphasize the initial level of performance of a journal (e.g., higher impact factor, higher h-index) due to the massive amounts of resources required to conduct a study. Because of this, we may see proportionate differentiation as the driving mechanism. Having a clearer understanding of the generative mechanism present in these situations can lead to better theory aimed at improving our understanding of the conditions that drive the emergence of different distribution shapes.

Additionally, given the prevalence of the generative mechanism of incremental differentiation in the observed performance distributions, team theory can begin to incorporate these findings into future conceptualizations. Because team performance trajectories appear to impact the team performance distribution shape, there is a need to incorporate these trajectories in understanding team performance. For instance, consider research on group pride (Beal et al. 2003). Our results showed that star teams are more abundant than previously thought and belonging to these teams likely has an impact on feelings of pride toward the team. Currently, theory on status differences of groups is not clearly defined in the team literature (Driskell et al. 2018), and group pride has received significantly less attention than other aspects of cohesion (Beal et al. 2003). However, our findings can provide a direct link from the experience of group pride to the ability a team has to maintain high performance trajectories. Also, theory on multiteam systems (MTSs) (De Vries et al. 2016) could be updated to explicitly consider these generative mechanisms as well. Specifically, MTSs can be characterized along a number of dimensions such as competency separation (Luciano et al. 2018). In the presence of greater performance variability, systems of multiple star teams could be constructed where competency separation would be low. On the other hand, the impact of team processes and output of teams may also be negatively impacted when competency separation is high, owing to the presence of one star team surrounded by several lower-performing teams. This is especially important given recent research that has investigated competency-based trust compared with relationship-based trust and their impact on creating and maintaining partnerships (Connelly et al. 2018). Thus, the shape of the distribution and the generative mechanisms involved can help address how competency separation impacts performance of the whole system as well as the internal relationships between teams.

Second, although our results provided evidence that the most prevalent type of nonnormal team performance distribution was PLC, other distributions also emerged. Considering the context in which a team operates may help explain why different types of heavy-tailed distributions exist and provide fruitful avenues for future research directions. For example, although appearing in other domains, the PLC distribution was especially dominant in the "Journal" and "Politics" domains. In each of these domains, it appears that the key factor in driving nonnormality is differentiation in the linear growth of team performance trajectories. Therefore, for instance, journal editorial boards that can continually increase the citation count of published articles over time and rise to stardom, while those that stagnate do not achieve these higher levels of performance. Similarly, there might be a differential ability of political campaign teams to increase their campaign donations linearly over time.

Third, the context in which a team operates has relevance for the emergence of heavy-tailed distributions. For example, in situations that are characterized by the lognormal distribution, there is likely a greater importance placed on the starting point of performance compared with exponential tail distributions. In these contexts, the initial starting value of a team will combine with the performance trajectory to dictate which teams rise to the level of stars. Future research is necessary to investigate which additional characteristics of team contexts drive the emergence of different distributions in addition to the three we examined in our study. Consider, for instance, the competitive nature of the team environment that may serve as a potential contextual factor that could lead to a greater proportion of stars. Our results showed that when we subgrouped the distributions into 116 tournament and 158 nontournament environments, results still showed a prevalence of heavy-tailed distributions and dominance of the PLC among nonnormal distributions. In less competitive environments, star teams emerged at much higher rates than would be expected by a normal distribution, providing further evidence for a generalized theory of heavy-tailed emergence. Nevertheless, results of distribution pitting showed that the PLC was not the unanimous winner among all nonnormal distributions (Table 2).

Fourth, another major contextual factor that could affect the emergence of different types of distributions and the proportion of star teams is the presence of a ceiling. In the presence of limiting factors to overall performance, generative mechanisms such as homogenization may cause a normal distribution to emerge. In an especially powerful case that provides initial empirical support for this possibility, almost all our pub trivia performance samples were best described by a normal distribution. In this context, a hard ceiling (i.e., maximum possible point total) limits the range of performance scores, likely resulting in distributions approaching normality and fewer star teams.

Fifth, another implication for future research is the need to address the impact of star teams within their organizations. Specifically, although star teams are beneficial in terms of their output, other factors may make star teams detrimental to organizations due to issues such as within-organization competition. Because of this, it will be important to further address intraorganizational team dynamics (e.g., interorganizational team conflict; Rose and Shoham 2004) given the prevalence of star teams as well as those teams on the lower end of the performance distribution.

Sixth, although a focus on the heavy tail is warranted given our results, another important consideration is what is occurring at the left side of the performance distribution where teams are extremely underperforming compared with the star teams. With heavy-tailed distributions, there is often a large cluster of these teams that would fall below the average level of performance, as shown by the classic study by Schachter et al. (1951), in which increased team cohesiveness surprisingly led to both higher and lower performing teams. Thus, there is a need to account for those underperforming teams.

Seventh, our focus was on the distributions of teams (i.e., distribution level of analysis) and not on individual teams (i.e., team level of analysis). Future research focusing on the team level of analysis could investigate possible dynamics leading to the exceptional performance of specific teams. For example, there is a current debate on the ideal proportion of individual stars leading to team stardom (Swaab et al. 2014, Gula et al. 2021).

Eighth, some nonintuitive findings also provide an opportunity to pursue further research. Specifically, the greater proportion of nonnormal distributions in the nontournament environments than in the competitive environments demonstrates a need to investigate further the importance of competition in determining the shape of the performance distribution. Something that seems to be consistent across all competitive contexts is the resource constraints that do not necessarily exist in noncompetitive environments. In competitive environments it can be very difficult to "expand the pie" in terms of outputs (e.g., competitions do not allow for all teams to be winners as there will always be losers). In contrast, noncompetitive environments are not generally constrained in the same way as there are ways for teams to not only take a bigger piece of the pie but also to expand the size of the pie for all those involved as well, leading to a larger proportion of star teams. In addition, our results may be something of an anomaly given that most tournament contexts that were normally distributed were specifically from the pub trivia teams. It is possible that the much higher proportion of nonnormal distributions in the nontournament distributions exists because of something that is unique only to those trivia competition environments that other team competitions do not have.

Finally, our findings are consistent with results at the individual level of analysis that showed that individual performance follows a PLC distribution (Joo et al. 2017). An important implication of this result is that there seems to be a generalized and isomorphic phenomenon of nonnormality emergence at multiple levels of analysis that provides opportunities for future multilevel theory development (Morgeson and Hofmann 1999). Therefore, although team-level outcomes are often the result of more than just an aggregation of their individual members (Woolley et al. 2010), our findings replicated across these two types of performance operationalization. An important goal for the field of organization science, as in all scientific fields, is to produce strong, generalizable theories (Pfeffer et al. 1977, Boyd et al. 2005). Our results contribute toward this goal in that they provide evidence for the ubiquity of heavy-tails with their associated generative mechanisms in performance distributions at the team level of analysis. Future research could examine potential isomorphism at additional lower (e.g., within-individual performance) and higher levels (e.g., organizational units larger than teams such as industries).

Methodological Implications

The dominant and widely used organization science data analytic procedures rely on GLM and do not adequately capture the true nature of relations in the presence of heavy-tailed distributions. For example, a regression coefficient with a value of 2.5 means that there is a 2.5 increase in team performance given a one-point increase in the antecedent; however, this is on *average*—a crucial clarification that is usually left out when results are reported (Cohen et al. 2003). In the presence of heavy-tailed distributions, as a measure of central tendency, the average cannot be interpreted in isolation because it is disproportionally influenced by outliers. In fact, models that assume normality treat the first and second moments—the mean and the variance—as key statistics in testing theory (O'Boyle and Aguinis 2012). However, because the mean is a measure of central tendency, it is only informative when it provides a description of a typical data point. In the presence of heavy-tailed distributions and many star teams, the average is moved to the right and no longer captures what could be considered a "typical" team (Buzsáki and Mizuseki 2014). Likewise, due to their extremely high level of heterogeneity, heavy-tailed distributions can have pseudo-infinite variance, making the variance (and SD) estimate unstable and therefore not useful as a descriptive or inferential statistic (Li and Zhao 2012). Accordingly, using the mean and SD in computing parameter estimates (e.g., correlations, regression coefficients), test statistics (e.g., F, t), and associated p values, as is done in all data analytic procedures that assume normality of residuals (e.g., multiple regression, analysis of variance, structural equation modeling, multilevel modeling), can lead to biased results (Jones et al. 2016). Overall, assuming normality implicitly or explicitly means that team research assumes little variability across teams, which may not be a good representation of the actual (nonnormal) distribution. Although the central limit theorem takes care of producing normally distributed sampling distributions even if the raw score distributions are not normal, effect-size estimates computed using the mean and variance (e.g., regression coefficients, correlation coefficients, ds) are biased (Cohen et al. 2003), further highlighting the importance of understanding the underlying distribution shape.

Accordingly, based on our results, authors and journal reviewers should not leave the normality assumption unchecked. First, distribution pitting, which can be implemented using the Dpit package for R available on CRAN, provides a procedure for testing whether a distribution is actually normal or better described by a heavy-tail. Just like we used distribution pitting results to decide to use Spearman correlations instead of Pearson's correlations to test our hypotheses, distribution pitting can be implemented to inform which statistical tools to use (i.e., methods that assume normality should only be used with normality is observed empirically). Additionally, researchers should also consider generating visualizations to provide a clearer understanding of the nature and degree of nonnormality as well as the opportunity to communicate that information to readers. Second, if the distribution is not normal, there will be a need to use alternative data-analytic procedures that do not rely on the normality assumption. For example, these include Spearman correlations instead of Pearson's correlations (de Winter et al. 2016) and additive unrestricted nonparametric multiple regression instead of GLM-based regression (Aguinis et al. 2019). Our results are relevant for all future research on team performance, but particularly so when the purpose is to predict the size and importance of a hypothesized performance antecedent. Instead, if the purpose is to simply show that a hypothesized effect exists or not (i.e., dichotomous decision), then deviations from normal distributions may not change the result. Bayesian analysis provides opportunities for researchers to investigate important theoretical relationships even in the absence of normality (Kruschke et al. 2012). Our results could inform team researchers in the future regarding more accurate prior distributions (Winkler 1967) to be used in Bayesian analysis. A related methodological implication of our results is that by drawing attention to the shapes of distributions, future research will hopefully be more transparent regarding distributions shapes, which will in turn be useful for future meta-analyses focusing on quantitatively aggregating results from prior empirical studies.

Another methodological implication is the importance of understanding and managing data points that are located far from the others (i.e., outliers). This is especially true when dealing with influential teams that fall in the thick tails of the performance distribution (Aguinis et al. 2013). Although these star teams (i.e., outliers) are often viewed as problems, in reality, they can indicate the presence of substantively interesting data points that warrant further investigation and theory development (Gibbert et al. 2021). However, reviews have suggested that previous research has not adequately and transparently managed outliers (Aguinis et al. 2013). Outliers are likely to be ignored if the mean is reported in isolation. At a minimum, it should be reported together with the mode and median—as well as measures of dispersion.

Implications for Practice

Because of the heavy-tailed nature of the team performance distribution, there is an important distinction between the performance of star teams and that of others. This finding suggests a need to implement proper compensation practices that reflect the very large variability in performance across teams. Creating compensation packages focused on spurring equitable, team-based pay that helps distinguish teams can help managers reward top performing teams and motivate other teams to reach higher performance levels (Garbers and Konradt 2014). This is especially true if the rewards are based on equity (Garbers and Konradt 2014), and there is transparency in the compensation plan (Aguinis and Bradley 2015). In addition, in the presence of heavy-tailed distributions, the resource allocation-performance and performance-value functions are not linear (Trevor et al. 2012, Hill et al. 2017). Accordingly, if the goal is to increase an organization's overall performance, return on investment (ROI) will be greater when resources are allocated to star teams. Thus, the specific nature of the performance distribution also informs practices about how to allocate resources among top-performing teams because some distributions show greater levels of differentiation between teams (e.g., pure power law) than others (e.g., exponential tail).

Additionally, although our results demonstrate that star teams are more common than previously thought, it is important to also pay particular attention to teams on the other end of the distribution—those that occupy the lowest levels of performance. Managers of teams should be cognizant of teams that fall at these extreme low levels and improve their performance by providing additional training (Salas et al. 2008) or improving team motivation (Park et al. 2013), which are ways to increase the performance of these lower-performing teams. In fact, based on an organization's strategic priorities and values, it may be beneficial to minimize the heterogeneity between teams in an attempt to enhance the performance of all teams.

Limitations

First, our data set included many sports samples, which is a potential limitation in terms of generalizability. However, sports data can be used to build and test theory in many domains (Day et al. 2012). Additionally, the sports teams used in our studies share many important characteristics with teams in more traditional work contexts (Day et al. 2012). For example, like many organization teams, sports teams perform tasks in high stress situations, are required to communicate extensively with teammates, work collaboratively to achieve a desired outcome, and compete with other teams for limited resources. Second, although much of the sports data we used reflect only part of the entire team performance construct (e.g., NCAA football yards per game reflects only the offensive performance of a sports team while ignoring defensive contributions to the team), our measures are nevertheless consistent with our definition of team performance because they capture various aspects of team output and results. Similarly, while there are antecedents of team performance that may impact the ultimate output a team is able to generate (e.g., luck, organizational prestige) the team performance indicators we chose closely follow our definition of team performance and capture a crucial aspect of performance for each of the samples chosen. Nevertheless, we readily acknowledge that other team performance measures warrant future consideration. For instance, consider team creative output (Somech and Drach-Zahavy 2013). Consistent with our results, Choi and Lee (2020) reported that this type of performance may also follow a heavy-tailed distribution. Third, given our research design, we did not investigate how the generative mechanisms may have effects over time. However, the shape of the team distribution is a necessary and sufficient condition to conclude which generative mechanism is responsible. Although a longitudinal design would add additional evidence, the presence of a specific distribution shape such as PLC is, in itself, enough to indicate which generative mechanism (i.e., incremental differentiation) is present given empirical evidence about the formation of distributions across numerous fields such as physics, zoology, ornithology, and geology (McKelvey and Andriani 2005, Joo et al. 2017). Finally, although incremental differentiation is dominant, this does not preclude the existence of other mechanisms that may also occur simultaneously. For example, although incremental differentiation does not rely on feedback loops to generate a heavy-tail, those feedback loops may still exert some influence over the shape of the heavy-tailed distribution (Joo et al. 2017). However, our findings suggest that the impact from those mechanisms is either short lived or not as influential as the incremental growth that leads to the emergence of PLC distributions.

Conclusions

Although many team theories imply the existence of nonnormal team performance distributions, empirical team research assumes normality implicitly by using statistical procedures that rely on the normality assumption or explicitly by transforming (i.e., squeezing) nonnormal data and giving less or no weight to very high-performing teams (e.g., by eliminating outliers). We sought to first ascertain the overall nature of the team performance distribution and critically examine theory-based boundary conditions (i.e., tournament versus nontournament environments, performance conceptualized and measured as an aggregation of individual-level performance versus team-level performance, performance measured in the presence versus absence of a hard left-tail zero, more versus less homogenous samples). Then, using team learning curve as an overarching conceptual framework, we examined three theory-based predictors of the shape of the team performance distribution and the proportion of star teams: Authority differentiation, temporal stability, and skill differentiation. Results indicated that the normal distribution is not nearly as common as has been assumed in the past. Instead, nonnormal distributions, and the power law with exponential cutoff distribution in particular, emerged as the most prevalent across a wide range of samples and contexts and boundary conditions. Moreover, incremental differentiation provides the best explanation for the observed large degree of performance variability and greater proportion of star teams, and temporal stability was the strongest antecedent. These findings challenge existing assumptions regarding team performance, open up new areas for future research directions, and lead to recommendations on methodological and managerial practices.

Acknowledgments

The authors thank Anita Williams Woolley and three *Organization Science* reviewers for highly constructive feedback that allowed us to improve our article in a substantial manner. Also, the authors thank the following individuals for generously allowing us to use their data in our study: John F. Mathieu, Erich C. Dierdorff, Daan van Knippenberg, and Kai Chi Yam. A previous version of this manuscript was presented at the meetings of the Academy of Management, Chicago, IL, August 2018, and portions of this manuscript are based on Kyle J. Bradley's doctoral dissertation, which was conducted at the Kelley School of Business, Indiana University, under the supervision of Herman Aguinis. The datasets and R scripts used in this study are available from the authors upon request (except for the data that were shared with us by authors of published articles).

Endnotes

[1] Although the majority of our samples contain data from the United States, Table 1 shows that we also drew several samples from non-U.S. contexts (e.g., UEFA Champions League, global virtual supply chain teams), thus providing additional diversity in the teams used in our analyses.

[2] Given that the normality assumption is so pervasive, we do not find it useful to single out the authors of this meta-analysis. However, we make the source available upon request.

References

Aggarwal I, Woolley AW, Chabris CF, Malone TW (2019) The impact of cognitive style diversity on implicit learning in teams. *Frontiers Psych.* 10(112):1–11.

Aguinis H, Bradley KJ (2015) The secret sauce for organizational success. *Organ. Dynamics* 44(3):161–168.

Aguinis H, O'Boyle E Jr (2014) Star performers in twenty-first century organizations. *Personality Psych.* 67(2):313–350.

Aguinis H, Gottfredson RK, Joo H (2013) Best-practice recommendations for defining, identifying, and handling outliers. *Organ. Res. Methods* 16(2):270–301.

Aguinis H, Ji YH, Joo H (2018) Gender productivity gap among star performers in STEM and other scientific fields. *J. Appl. Psych.* 103(12):1283–1306.

Aguinis H, Ramani RS, Alabduljader N, Bailey JR, Lee J (2019) A pluralist conceptualization of scholarly impact in management education: Students as stakeholders. *Acad. Management Learn. Edu.* 18(1):11–42.

Ahearne M, MacKenzie SB, Podsakoff PM, Mathieu JE, Lam SK (2010) The role of consensus in sales team performance. *J. Marketing Res.* 47(3):458–469.

Amitrano D (2012) Variability in the power-law distributions of rupture events. *Eur. Phys. J. Spec. Top.* 205(1):199–215.

Andriani P, McKelvey B (2009) Perspective—From Gaussian to Paretian thinking: Causes and implications of power laws in organizations. *Organ. Sci.* 20(6):1053–1071.

Argote L, Epple D (1990) Learning curves in manufacturing. *Science* 247(4945):920–924.

Argote L, Ingram P (2000) Knowledge transfer: A basis for competitive advantage in firms. *Organ. Behav. Human Decision Processes* 82(1):150–169.

Baiden BK, Price AD, Dainty AR (2006) The extent of team integration within construction projects. *Internat. J. Project Management* 24(1):13–23.

Banerjee A, Yakovenko VM (2010) Universal patterns of inequality. *New J. Phys.* 12(7):075032.

Banker RD, Field JM, Schroeder RG, Sintia KK (1996) Impact of work teams on manufacturing performance: A longitudinal field study. *Acad. Management J.* 39(4):867–890.

Beal DJ, Cohen RR, Burke MJ, McLendon CL (2003) Cohesion and performance in groups: A meta-analytic clarification of construct relations. *J. Appl. Psych.* 88(6):989–1004.

Becker TE, Robertson MM, Vandenberg RJ (2019) Nonlinear transformations in organizational research: Possible problems and potential solutions. *Organ. Res. Methods* 22(4):831–866.

Bell BS, Kozlowski SWJ, Blawath S (2012) Team learning: A review and integration. Kozlowski SWJ, ed. *The Oxford Handbook of Organizational Psychology,* vol. 2 (Oxford University Press, Oxford, UK), 859–909.

Bell ST (2007) Deep-level composition variables as predictors of team performance: A meta-analysis. *J. Appl. Psych.* 92(3):595–615.

Boyd BK, Finkelstein S, Gove S (2005) How advanced is the strategy paradigm? The role of particularism and universalism in shaping research outcomes. *Strategic Management J.* 26(9):841–854.

Bresman H (2010) External learning activities and team performance: A multimethod field study. *Organ. Sci.* 21(1):81–96.

Busse C, Kach AP, Wagner SM (2017) Boundary conditions: What they are, how to explore them, why we need them, and when to consider them. *Organ. Res. Methods* 20(4):574–609.

Buzsáki G, Mizuseki K (2014) The log-dynamic brain: How skewed distributions affect network operations. *Nature Rev. Neurosci.* 15(4):264–278.

Chang A, Bordia P, Duck J (2003) Punctuated equilibrium and linear progression: Toward a new understanding of group development. *Acad. Management J.* 46(1):106–117.

Choi M, Lee CY (2020) Power-law distributions of corporate innovative output: Evidence from US patent data. *Scientometrics* 122: 519–554.

Clauset A, Shalizi CR, Newman ME (2009) Power-law distributions in empirical data. *SIAM Rev.* 51(4):661–703.

Cohen SG, Bailey DE (1997) What makes teams work: Group effectiveness research from the shop floor to the executive suite. *J. Management* 23(3):239–290.

Cohen J, Cohen P, West SG, Aiken LS (2003) *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, 3rd ed. (Lawrence Erlbaum Associates, Mahwah, NJ).

Connelly BL, Crook TR, Combs JG, Ketchen DJ Jr, Aguinis H (2018) Competence-and integrity-based trust in interorganizational relationships: Which matters more? *J. Management* 44(3):919–945.

Day DV, Gordon S, Fink C (2012) The sporting life: Exploring organizations through the lens of sport. *Acad. Management Ann.* 6(1):397–433.

DeShon RP, Kozlowski SW, Schmidt AM, Milner KR, Wiechmann D (2004) A multiple-goal, multilevel model of feedback effects on the regulation of individual and team performance. *J. Appl. Psych.* 89(6):1035–1056.

Devine DJ, Philips JL (2001) Do smarter teams do better: A meta-analysis of cognitive ability and team performance. *Small Group Res.* 32(5):507–532.

de la Torre Hernández JM, Edelman ER (2017) From nonclinical research to clinical trials and patient registries: Challenges and opportunities in biomedical research. *Rev. Esp. Cardiology (English Ed.)* 70(12):1121–1133.

De Vries TA, Hollenbeck JR, Davison RB, Walter F, Van Der Vegt GS (2016) Managing coordination in multiteam systems: Integrating micro and macro perspectives. *Acad. Management J.* 59 (5):1823–1844.

de Winter JC, Gosling SD, Potter J (2016) Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. *Psych. Methods* 21(3):273–290.

Dierdorff EC, Fisher DM, Rubin RS (2019) The power of percipience: Consequences of self-awareness in teams on team-level functioning and performance. *J. Management* 45(7): 2891–2919.

Driskell T, Salas E, Driskell JE (2018) Teams in extreme environments: Alterations in team development and teamwork. *Human Resources Management Rev.* 28(4):434–449.

Edmondson AC, Bohmer RM, Pisano GP (2001) Disrupted routines: Team learning and new technology implementation in hospitals. *Admin. Sci. Quart.* 46(4):685–716.

Edmondson AC, Dillon JR, Roloff KS (2007) Three perspectives on team learning: Outcome improvement, task mastery, and group process. *Acad. Management Ann.* 1(1):269–314.

Foster MK, Abbey A, Callow MA, Zu X, Wilbon AD (2015) Rethinking virtuality and its impact on teams. *Small Group Res.* 46(3): 267–299.

Fries SD (1992) *NASA Engineers and the Age of Apollo* (National Aeronautics and Space Administration Scientific and Technical Information Program, Washington, DC).

Garbers Y, Konradt U (2014) The effect of financial incentives on performance: A quantitative review of individual and team-based financial incentives. *J. Occupational Organ. Psych.* 87(1): 102–137.

Gersick CJ (1988) Time and transition in work teams: Toward a new model of group development. *Acad. Management J.* 31(1):9–41.

Gibbert M, Nair LB, Weiss M, Hoegl M (2021) Using outliers for theory building. *Organ. Res. Methods* 24(1):172–181.

Gibson CB, Gibbs JL (2006) Unpacking the concept of virtuality: The effects of geographic dispersion, electronic dependence, dynamic structure, and national diversity on team innovation. *Admin. Sci. Quart.* 51(3):451–495.

Greer LL, Caruso HM, Jehn KA (2011) The bigger they are, the harder they fall: Linking team power, team conflict, and performance. *Organ. Behav. Human Decision Processes* 116(1):116–128.

Greer LL, de Jong BA, Schouten ME, Dannals JE (2018) Why and when hierarchy impacts team effectiveness: A meta-analytic integration. *J. Appl. Psych.* 103(6):591–613.

Gross J, Ryan JC (2015) Landscapes of research: Perceptions of open access (OA) publishing in the arts and humanities. *Publications* 2015(3):65–88.

Gula B, Vaci N, Alexandrowicz RW, Bilalic M (2021) Never too much: The benefit of talent to team performance in the NBA: Comment on Swaab et al. (2014). *Psych. Sci.* 32(2):301–304.

Hanel R, Corominas-Murtra B, Liu B, Thurner S (2017) Fitting power-laws in empirical data with estimators that work for all exponents. *PLoS One* 12(2):e0170920.

Hill AD, Aime F, Ridge JW (2017) The performance implications of resource and pay dispersion: The case of Major League Baseball. *Strategic Management J.* 38(9):1935–1947.

Hollenbeck JR, Beersma B, Schouten ME (2012) Beyond team types and taxonomies: A dimensional scaling conceptualization for team description. *Acad. Management Rev.* 37(1):82–106.

Hollingshead AB (1998) Group and individual training: The impact of practice on performance. *Small Group Res.* 29(2):254–280.

Ilgen DR, Hollenbeck JR, Johnson M, Jundt D (2005) Teams in organizations. *Annu. Rev. Psych.* 56(1):517–543.

Jones AM, Lomas J, Moore PT, Rice N (2016) A quasi-Monte-Carlo comparison of parametric and semiparametric regression methods for heavy-tailed and non-normal data: An application to healthcare costs. *J. Royal Statist. Soc. Ser. A* 179(4):951–974.

Joo H, Aguinis H, Bradley KJ (2017) Not all non-normal distributions are created equal: Improved theoretical and measurement precision. *J. Appl. Psych.* 102(7):1022–1053.

Kane AA, Argote L, Levine JM (2005) Knowledge transfer between groups via personnel rotation: Effects of social identity and knowledge quality. *Organ. Behav. Human Decision Processes* 96(1): 56–71.

Katz R (1982) The effects of group longevity on project communication and performance. *Admin. Sci. Quart.* 27(1):81–104.

Kim JS, Yum BJ (2008) Selection between Weibull and lognormal distributions: A comparative simulation study. *Comput. Statist. Data Anal.* 53(2):477–485.

Kirkman BL, Mathieu JE, Cordery JL, Rosen B, Kukenberger M (2011) Managing a new collaborative entity in business organizations: Understanding organizational communities of practice effectiveness. *J. Appl. Psych.* 96(6):1234–1245.

Knapp R (2010) Collective (team) learning process models: A conceptual review. *Human Resources Dev. Rev.* 9(3):285–299.

Kruschke JK, Aguinis H, Joo H (2012) The time has come: Bayesian methods for data analysis in the organizational sciences. *Organ. Res. Methods* 15(4):722–752.

LeBreton JM, Senter JL (2008) Answers to 20 questions about interrater reliability and interrater agreement. *Organ. Res. Methods* 11(4):815–852.

Lee SM, Koopman J, Hollenbeck JR, Wang LC, Lanaj K (2015) The team descriptive index (TDI): A multidimensional scaling approach for team description. *Acad. Managment Discovery* 1(1):91–116.

Le Roy F, Fernandez AS (2015) Managing coopetitive tensions at the working-group level: The rise of the coopetitive project team. *British J. Management* 26(4):671–688.

Li M, Zhao W (2012) Visiting power laws in cyber-physical networking systems. *Math. Problems Engrg.* 2012:302786.

Lindsley DH, Brass DJ, Thomas JB (1995) Efficacy-performing spirals: A multilevel perspective. *Acad. Management Rev.* 20(3):645–678.

Luan K, Rico R, Xie XY, Zhang Q (2016) Collective team identification and external learning. *Small Group Res.* 47(4):384–405.

Luciano MM, DeChurch LA, Mathieu JE (2018) Multiteam systems: A structural framework and meso-theory of system functioning. *J. Management* 44(3):1065–1096.

Mathieu JE, Gilson LL, Ruddy TM (2006) Empowerment and team effectiveness: An empirical test of an integrated model. *J. Appl. Psych.* 91(1):97–108.

Mathieu JE, Gallagher PT, Domingo MA, Klock EA (2019) Embracing complexity: Reviewing the past decade of team effectiveness research. *Annu. Rev. Organ. Psych. Organ. Behav.* 6:17–46.

Mathieu JE, Tannenbaum SI, Donsbach JS, Alliger GM (2014) A review and integration of team composition models: Moving toward a dynamic and temporal framework. *J. Management* 40(1):130–160.

Mathieu JE, Heffner TS, Goodwin GF, Salas E, Cannon-Bowers JA (2000) The influence of shared mental models on team process and performance. *J. Appl. Psych.* 85(2):273–283.

Maynard MT, Mathieu JE, Rapp TL, Gilson LL (2012) Something (s) old and something (s) new: Modeling drivers of global virtual team effectiveness. *J. Organ. Behav.* 33(3):342–365.

Maynard MT, Mathieu JE, Gilson LL, Sanchez D, Dean MD (2019) Do I really know you and does it matter? Unpacking the relationship between familiarity and information elaboration in global virtual teams. *Group Organ. Management* 44(1):3–37.

McKelvey B, Andriani P (2005) Why Gaussian statistics are mostly wrong for strategic organization. *Strategic Organ.* 3(2): 219–228.

Mitzenmacher M (2004) A brief history of generative models for power law and lognormal distributions. *Internet Math.* 1(2):226–251.

Morgeson FP, Hofmann DA (1999) The structure and function of collective constructs: Implications for multilevel research and theory development. *Acad. Management Rev.* 24(2):249–265.

O'Boyle E Jr, Aguinis H (2012) The best and the rest: Revisiting the norm of normality of individual performance. *Personality Psych.* 65(1):79–119.

Pareto V (1897) The new theories of economics. *J. Political Econom.* 5(4):485–502.

Park G, Spitzmuller M, DeShon RP (2013) Advancing our understanding of team motivation: Integrating conceptual approaches and content areas. *J. Management* 39(5):1339–1379.

Pfeffer J, Leong A, Strehl K (1977) Paradigm development and particularism: Journal publication in three scientific disciplines. *Soc. Forces* 55(4):938–951.

Pisano GP, Bohmer RM, Edmondson AC (2001) Organizational differences in rates of learning: Evidence from the adoption of minimally invasive cardiac surgery. *Management Sci.* 47(6):752–768.

Rapp TL, Mathieu JE (2019) Team and individual influences on members' identification and performance per membership in multiple team membership arrangements. *J. Appl. Psych.* 104(3): 303–320.

Rapp TL, Gilson LL, Mathieu JE, Ruddy T (2016) Leading empowered teams: An examination of the role of external team leaders and team coaches. *Leadership Quart.* 27(1):109–123.

Reagans R, Argote L, Brooks D (2005) Individual experience and experience working together: Predicting learning rates from knowing who knows what and knowing how to work together. *Management Sci.* 51(6):869–881.

Rego A, Owens B, Yam KC, Bluhm D, Cunha MPE, Silard A, Goncalves L, et al. (2019) Leader humility and team performance: Exploring the mediating mechanisms of team PsyCap and task allocation effectiveness. *J. Management* 45(3):1009–1033.

Ren Y, Argote L (2011) Transactive memory systems 1985–2010: An integrative framework of key dimensions, antecedents, and consequences. *Acad. Management Ann.* 5(1):189–229.

Riedl C, Kim YJ, Gupta P, Malone TW, Woolley AW (2021) Quantifying collective intelligence in human groups. *Proc. National Acad. Sci. USA* 118(21):1–5.

Romme AGL (2004) Unanimity rule and organizational decision making: A simulation model. *Organ. Sci.* 15(6):704–718.

Rose GM, Shoham A (2004) Interorganizaitojnal task and emotional conflict with international channels of distribution. *J. Bus. Res.* 57(9):942–950.

Salas E, DiazGranados D, Klein C, Burke CS, Stagl KC, Goodwin GF, Halpin SM (2008) Does team training improve team performance? A meta-analysis. *Human Factors* 50(6):903–933.

Schachter S, Ellertson N, McBride D, Gregory D (1951) An experimental study of cohesiveness and productivity. *Human Relations* 4(3):229–238.

Simonton DK (2003) Scientific creativity as constrained stochastic behavior: The integration of product, person, and process perspectives. *Psych. Bull.* 129(4):475–494.

Somech A, Drach-Zahavy A (2013) Translating team creativity to innovation implementation: The role of team composition and climate for innovation. *J. Management* 39(3):684–708.

Sornette D (2006) *Critical Phenomena in Natural Sciences: Chaos, Fractals, Selforganization and Disorder: Concepts and Tools* (Springer Science & Business Media, Berlin).

Steiner ID (1972) *Group Processes and Productivity* (Academic Press, New York).

Sundaresan S, Zhang Z (2012) Parallel teams for knowledge creation: Role of collaboration and incentives. *Decision Support Systems* 54(1):109–121.

Swaab RI, Schaerer M, Anicich EM, Ronay R, Galinsky AD (2014) The too-much-talent effect: Team interdependence determines when more talent is too much or not enough. *Psych. Sci.* 25(8):1581–1591.

Trevor CO, Reilly G, Gerhart B (2012) Reconsidering pay dispersion's effect on the performance of interdependent work: Reconciling sorting and pay inequality. *Acad. Management J.* 55(3):585–610.

Van Bunderen L, Greer LL, Van Knippenberg D (2018) When interteam conflict spirals into intrateam power struggles: The pivotal role of team power structures. *Acad. Management J.* 61(3):1100–1130.

Vancouver JB, Li X, Weinhardt JM, Steel P, Purl JD (2016) Using a computational model to understand possible sources of skews in distributions of job performance. *Personality Psych.* 69(4):931–974.

Vashdi DR, Bamberger PA, Erez M (2013) Can surgical teams ever learn? The role of coordination, complexity, and transitivity in action team learning. *Acad. Management J.* 56(4):945–971.

Virkar Y, Clauset A (2014) Power-law distributions in binned empirical data. *Ann. Appl. Statist.* 8(1):89–119.

Volmer J, Sonnentag S (2011) The role of star performers in software design teams. *J. Management Psych.* 26(3):219–234.

Winkler RL (1967) The assessment of prior distributions in Bayesian analysis. *J. Amer. Statist. Assoc.* 62(319):776–800.

Woolley AW, Chabris CF, Pentland A, Hashmi N, Malone TW (2010) Evidence for a collective intelligence factor in the performance of human groups. *Science* 330(6004):686–688.

Kyle J. Bradley is an assistant professor of management at Kansas State University. His research interests focus on understanding star employees, the impact they have on organizations, and improving research methods in management research. He earned his PhD from Indiana University's Kelley School of Business in 2017.

Herman Aguinis is the Avram distinguished scholar, professor of management, and chair of the Department of Management at The George Washington University School of Business. His research addresses the acquisition and deployment of talent in organizations and organizational research methods, and he is currently serving as president of the Academy of Management.