How much does performance differ between people?

Max Daniel & Benjamin Todd

Some people seem to achieve orders of magnitudes more than others in the same job. For instance, among companies funded by Y Combinator the top 0.5% account for more than ⅔ of the total market value; and among successful bestseller authors, the top 1% stay on the New York Times bestseller list more than 25 times longer than the median author in that group.

This is a striking and often unappreciated fact, but raises many questions. How many jobs have these huge differences in achievements? More importantly, *why* can achievements differ so much, and can we identify future top performers in advance? Are some people much more talented? Have they spent more time practicing key skills? Did they have more supportive environments, or start with more resources? Or did the top performers just get lucky?

More precisely, when recruiting, for instance, we'd want to know the following: when predicting the future performance of different people in a given job, what does the distribution of *predicted* ('ex-ante') performance look like?

This is an important question for EA community building and hiring. For instance, if it's possible to identify people who will be able to have a particularly large positive impact on the world ahead of time, we'd likely want to take a more targeted approach to outreach.

More concretely, we may be interested in two different ways in which we could encounter large performance differences :

1. If we look at a random person, by how much should we expect their performance to differ from the average?
2. What share of total output should we expect to come from the small fraction of people we're most optimistic about (say, the top 1% or top 0.1%) – that is, how *heavy-tailed* is the distribution of ex-ante performance?

(See this appendix for how these two notions differ from each other.)

Depending on the decision we're facing we might be more interested in one or the other. Here we mostly focused on the second question, i.e., on how heavy the tails are.

This document contains our findings from a shallow literature review and theoretical arguments. Max was the lead author, building on some initial work by Ben, who also provided several rounds of comments.

You can see a short summary of our findings below.

We expect this post to be useful for:

- (Primarily:) Junior EA researchers who want to do further research in this area. See in particular the section on *Further research*.
- (Secondarily:) EA decision-makers who want to get a rough sense of what we do and don't know about predicting performance. See in particular this summary and the bolded parts in our section on *Findings*.
    - We weren't maximally diligent with double-checking our spreadsheets etc.; if you wanted to rely heavily on a specific number we give, you might want to do additional vetting.

To determine the distribution of *predicted* performance, we proceed in two steps:

1. We start with how ***ex-post*** performance is distributed. That is, **how much did the performance of different people vary when we *look back* at completed tasks?** On these questions, we'll review empirical evidence on both typical jobs and expert performance (e.g. research).
2. Then we ask how ***ex-ante*** performance is distributed. That is, when we employ our best methods to ***predict future performance*** **by different people**, how will these predictions vary? On these questions, we review empirical evidence on measurable factors correlating with performance as well as the implications of theoretical considerations on which kinds of processes will generate different types of distributions.

Here we adopt a **very loose conception of *performance* that includes both short-term (e.g. sales made on one day) and long-term achievements (e.g. citations over a whole career)**. **We also allow for performance metrics to be influenced by things beyond the performer's control.**

Our overall bottom lines are:

- ***Ex-post*** **performance appears 'heavy-tailed' in many relevant domains, but with very large differences in *how* heavy-tailed: the top 1% account for between 4% to over 80% of the total.** For instance, we find 'heavy-tailed' distributions (e.g. log-normal, power law) of scientific citations, startup valuations, income, and media sales. By contrast, a large meta-analysis reports **'thin-tailed' (Gaussian) distributions for ex-post performance in less complex jobs such as cook or mail carrier[1]: the top 1% account for 3-3.7% of the total**. These figures illustrate that **the difference between 'thin-tailed' and 'heavy-tailed' distributions can be modest in the range that matters in practice, while differences *between* 'heavy-tailed' distributions can be massive**. (More.)
- ***Ex-ante*** **performance is heavy-tailed in at least one relevant domain: science.** More precisely, future citations as well as awards (e.g. Nobel Prize) are predicted by past citations in a range of disciplines, and in mathematics by scores at the International Maths Olympiad. (More.)

[1] For performance in "high-complexity" jobs such as attorney or physician, that meta-analysis (Hunter et al. 1990) reports a coefficient of variation that's about 1.5x as large as for 'medium-complexity' jobs. Unfortunately, we can't calculate how heavy-tailed the performance distribution for high-complexity jobs is: for this we would need to stipulate a particular type of distribution (e.g. normal, log-normal), but Hunter et al. only report that the distribution does *not* appear to be normal (unlike for the low- and medium-complexity cases).

- More broadly, there are **known, measurable correlates of performance in many domains** (e.g. general mental ability). Several of them appear to **remain valid in the tails**. (More.)
- However, **these correlations *by itself* don't tell us much about the shape of the ex-ante performance distribution:** in particular, they would be consistent with either thin-tailed or heavy-tailed ex-ante performance. (More.)
- **Uncertainty should move us toward acting *as if* ex-ante performance was heavy-tailed –** because if you have some credence in it being heavy-tailed, it's heavy-tailed in expectation **– but not all the way, and less so the smaller our credence in heavy-tails.** (More.)
- To infer the shape of the ex-ante performance distribution, it would be more useful to have a mechanistic understanding of the process generating performance, but such **fine-grained causal theories of performance are rarely available**. (More.)
- Nevertheless, **our best guess is that moderately to extremely heavy-tailed ex-ante performance is widespread at least for 'complex' and 'scaleable' tasks**. (I.e. ones where the performance metric can in practice range over many orders of magnitude and isn't artificially truncated.) This is based on our best guess at the causal processes that generate performance combined with the empirical data we've seen. However, we think this is debatable rather than conclusively established by the literature we reviewed. (More.)
- **There are several opportunities for valuable further research.** (More.)

Overall, doing this investigation probably made us a little less confident that highly heavy-tailed distributions of ex-ante performance are widespread, and think that common arguments for it are often too quick. That said, we still think there are often large differences in performance (e.g. some software engineers have 10-times the output of others[2]), these are somewhat predictable, and it's often reasonable to act on the assumption that the ex-ante distribution is heavy-tailed in many relevant domains (broadly, when dealing with something like 'expert' performance as opposed to 'typical' jobs).

Some advice for how to work with these concepts in practice:

- In practice, **don't treat 'heavy-tailed' as a binary property**. Instead, ask *how* heavy the tails of some quantity of interest are, for instance by identifying the frequency of outliers you're interested in (e.g. top 1%, top 0.1%, …) and comparing them to the median or looking at their share of the total.[3]

---

[2] Claims about a 10x output gap between the best and average programmers are very common, as evident from a Google search for '10x developer'. In terms of value rather than quantity of output, the WSJ has reported a Google executive claiming a 300x difference. For a discussion of such claims see, for instance, this blog post by Georgia Institute of Technology professor Mark Guzdial. Similarly, slide 37 of this version of Netflix's influential 'culture deck' claims (without source) that "In creative/inventive work, the best are 10x better than the average".

[3] Similarly, don't treat 'heavy-tailed' as an asymptotic property – i.e. one that by definition need only hold for values above some arbitrarily large value. Instead, consider the range of values that matter in practice. For instance, a distribution that exhibits heavy tails only for values greater than 10^100 would be heavy-tailed in the asymptotic sense. But for e.g. income in USD values like 10^100 would never show up in practice – if your distribution is supposed to correspond to income in USD you'd only be interested in a much smaller range, say up to 10^10. Note that this advice is in contrast to the standard definition of 'heavy-tailed' in mathematical contexts, where it usually *is* defined as an asymptotic property. Relatedly, a distribution that only takes values in some finite range – e.g. between 0 and 10 billion – is *never* heavy-tailed in the mathematical-asymptotic sense, but it may well

- **Carefully choose the underlying population and the metric for performance**, in a way that's tailored to the purpose of your analysis. In particular, be mindful of whether you're looking at the full distribution or some tail (e.g. wealth of all citizens vs. wealth of billionaires).

In an appendix, we provide more detail on some background considerations:

- The **conceptual difference between 'high variance' and 'heavy tails'**: Neither property implies the other. Both mean that unusually good opportunities are much better than typical ones. However, only heavy tails imply that outliers account for a large *share of the total*, and that naive extrapolation underestimates the size of future outliers. (More.)
- We can often distinguish heavy-tailed from light-tailed data by eyeballing (e.g. in a log-log plot), but **it's hard to empirically distinguish different heavy-tailed distributions from one another** (e.g. log-normal vs. power laws). When extrapolating beyond the range of observed data, we advise to proceed with caution and to not take the *specific* distributions reported in papers at face value. (More.)
- There is a small number of papers in industrial-organizational psychology on the specific question whether performance in typical jobs is normally distributed or heavy-tailed. However, we don't give much weight to these papers because their broad high-level conclusion ("it depends") is obvious but we have doubts about the statistical methods behind their more specific claims. (More.)
- We also quote (in more detail than in the main text) the **results from a meta-analysis of predictors of salary, promotions, and career satisfaction**. (More.)
- We provide a technical discussion of how our metrics for heavy-tailedness are affected by the 'cutoff' value at which the tail starts. (More.)

Finally, we provide a glossary of the key terms we use, such as *performance* or *heavy-tailed*.

---

be in the "practical" sense (where you anyway cannot empirically distinguish between a distribution that can take arbitrarily large values and one that is "cut off" beyond some very large maximum).

Findings

Ex-post performance can be heavy-tailed depending on domain and metric, with large differences in *how* heavy-tailed

There is abundant evidence that **the *ex-post* distribution of some measures of performance in some relevant domains, e.g. scientific citations or startup valuations, is heavy-tailed (to varying degrees) across people (see *Table 1*).** This roughly means that when we *look back* at completed tasks, outliers account for a disproportionately large share of total output.

However, **heavy-tailed performance distributions are not universal**. Depending on how performance is measured, we may find a light-tailed (e.g. normal) distribution instead, especially in 'typical' rather than unusually complex jobs. For examples and a systematic discussion, see Aguinis et al. (2016), Beck et al. (2014), and Hunter, Schmidt, & Judiesch (1990). **For selected light-tailed examples, see *Table 2.***

Note that performance measures can also be light-tailed 'by design'. For instance, the popular website IMDb.com rates movies on a scale from 1 to 10. The highest-rated movies could only account for a significant share of the sum of ratings across all movies if people rated the majority of movies with values that are orders of magnitude smaller than 1, which in practice is not how raters interpret and use this scale. It is therefore no surprise that Liu et al. (2018) uses a light-tailed distribution to model IMDb ratings.

Of course, this doesn't tell us anything about whether the 'performance' of movies might be heavy-tailed when measured in a different way, for example by their box office revenue.

As one example of heavy-tailed ex-post performance, among companies funded by Y Combinator the top 1% account for more than ⅕ of the total market value. Other cases are less extreme, e.g.:

- In 2005 the global top 1% accounted for 21% of world income (adjusted for purchasing power).[4]
- Among scientists with long careers, the top 1% most-cited ones get around 7% of all citations, and the top 1% most prolific ones author 4.0% of all papers.[5]

For comparison, in a normal distribution fitted to output data from various "medium-complexity" jobs (e.g. cook) the top 1% account for 3.7% of total output. This is not that different from the 4.0% top-1%-share figure for papers-by-author despite the former being from a 'light-tailed' and the latter being from a 'heavy-tailed' distribution.[6] On the other hand, we have seen that *among* different heavy-tailed distributions the top-1%-share can vary by a factor of more than 10.

This illustrates that it's useful to **ask *how* heavy the tail of performance is in a specific case**, rather than just asking the binary question whether the tails are heavier than for an exponential distribution (a common maths definition of 'heavy-tailed' as a binary property). **For practical purposes there are large differences among heavy-tailed distributions. They range from "winner takes most" situations to ones where the difference to a normal distribution remains modest across the full range of values we'll ever encounter in practice.**[7]

To compare the heavy-tailedness of different distributions, we suggest the share of expected value in the top X% (top 20%, 10%, 1%, etc.) – **see *Table 3***. This measure has several advantages: it focuses on the key difference between heavy-tailed and light-tailed distributions; it highlights that the property of interest isn't binary but varies continuously; it has a straightforward interpretation; and we can apply it to different families of distributions as well as unparameterized data.

One disadvantage of this measure is that it depends on the 'underlying population' – for instance, whether we look at *all* authors of scientific papers or only those who have published consistently over many years. (Consider that a large share of paper authors only

---

[4] Anand & Segal (2014, Table 11.5), the first estimate of the world income distribution that takes into account estimates of top earners within countries.

[5] Based on Sinatra et al. (2016); note that the figures are based on their fitted distribution, not the actual data. The distribution of total citations to authors has no closed-form expression but can be simulated based on Sinatra and colleagues' model. Specifically, I generated 100 independent samples of 1 million scientists each. The mean top-1%-share across the 100 samples was 7.17%, with a standard deviation of 0.0267%. The script I used for the simulation is here, and a screenshot of the output from running the script on March 6th 2021 is here.

[6] Note that the difference between any 'heavy-tailed' and 'thin-tailed' distribution must become arbitrarily large in the limit of increasingly extreme top-shares. That is, if we look at the top 1%, top 0.1%, top 0.01%, and so on, the top-share of the light-tailed distribution will eventually fall much quicker than that for the heavy-tailed distribution (and so the ratio of the top-shares becomes arbitrarily large). However, in practice it matters *when* the difference becomes large: e.g. we often deal with sufficiently large groups of people that it's useful to know about the top 1% but we will rarely if ever be interested in some property of the top 10^{-100}.

[7] Among the distributions we found, the share of the top 0.01% (1 in 10,000) differed by a factor of less than 2 between the heaviest 'thin-tailed' and the least heavy 'heavy-tailed' distribution – but by a factor of more than 500 among different 'heavy-tailed' distributions!

have very few publications, e.g. people who leave academia after their PhDs.) More broadly, certain distributions such as power laws usually only apply in the tails of performance, and then we need to be careful to distinguish between the share of the total *in the tail* and the share of the total *in the full distribution.* For instance, the third row in *Table 3* says that the top 20% of US billionaires account for about 88% of the wealth *of all US billionaires.* If the underlying population instead were all Americans, then the top 20% of billionaires would correspond to a much higher quantile of *that* population (perhaps the top 0.0001%) but since we're now comparing to a larger amount of total wealth their share would also be lower than 88% (in fact at most 30%).[8]

This means that a direct comparison of different entries in our tables may be misleading if they report the same quantity (e.g. citations) for different population subsets (e.g. all scientists vs. tail of highly cited scientists).

(Another disadvantage is that it has a straightforward interpretation only for distributions that range only over positive values. If negative values 'cancel out' positive ones in expectation, this will push the mean toward zero, and thus increase the share of the top X% irrespective of how fast the tails diminish. In the extreme case of a distribution with mean zero, the "share of the total" for any top X% would involve a division by zero and thus be undefined.)

The data in *Table 3* suggests that the same quantity (e.g. wealth) tends to be *more* heavy-tailed for more "elite" populations, i.e. smaller populations that have been more heavily selected for performance (e.g. US billionaires vs. all Americans, scientists published in *Nature* vs. less prestigious journals). This is also suggested by some theoretical considerations[9], but we don't know how generally it holds. If it holds more widely, it would for instance be relevant to assessing replaceability in competitive jobs.

In *Tables 3* and *4* we provide more detail on some heavy-tailed performance distributions. However, as explained in the appendix, our confidence in this more specific information – including the type of distribution, say log-normal vs. Pareto – is low; we think the main robust findings simply are that (i) many performance metrics across many domains are, in a broad sense, heavy-tailed, and that (ii) different performance metrics or samples can vary considerably in *how* heavy the tails are, even for data in the same broad domain (e.g. wealth).

If you are more interested in variance, for some of these distributions we provide a scale-free measure of variance (the 'coefficient of variation', i.e. standard deviation over mean) in *Table 5.*

---

[8] The 0.0001% and 30% figures given here are very crude ballpark estimates based on assuming that there are 1,000 US billionaires (as of March 2020 probably an overestimate by a factor of ~2) in a population of 300 million (in fact ~330 million in 2019), and that the share of total wealth held by billionaires is 35% (which in fact is about the share held by the top 1%, whereas billionaires are on the order of the top 0.0001%).

[9] Roughly: the more heavily the population has been selected, the more room there was for 'success begets success' dynamics to amplify differences, and the more performance tends to depend on a larger number of factors – both of which push toward more heavy tails. For a more detailed explanation, see the later subsection on why we'd guess that *ex-ante* performance is often heavy-tailed.

Table 1. Examples of heavy-tailed distributions of ex-post performance.

Performance-relevant quantity found to be heavy-tailed	**Sources**
Citations by scientist (whole career)	Liu et al. (2018), Sinatra et al. (2016), Petersen, Wang, & Stanley (2010)
Number of publications by scientists (whole career)	Sinatra et al. (2016), Petersen, Wang, & Stanley (2010), Clauset et al. (2009)
Profits by startup founders	80,000 Hours (2014a, 2014b)
Various metrics of success in arts & entertainment by artist, e.g. weeks on the NYT bestseller list by fiction author or movie box office gross by director	Tauberg (2018)
Wealth by individual (worldwide and within various countries)	Atkinson & Bourguignon (eds., 2014), Clauset et al. (2009)
Income by individual (worldwide and within various countries)	Atkinson & Bourguignon (eds., 2014), Our World in Data
Citations by paper	Brzezinski (2015), Golosovsky & Solomon (2012), Wallace, Larivière, & Gingras (2009), Clauset et al. (2009), Radicchi, Fortunato, & Castellano (2008), Redner (1998), Price (1965)
Programmer output	Bryan (1994)
Returns to stock indices by time period (e.g. 1-min returns of the S&P 500)	Malevergne, Pisarenko, & Sornette (2005)
Auction prices by artwork	Liu et al. (2018)

Table 2. Examples of light-tailed distributions of ex-post performance.

Performance-relevant quantity found to be light-tailed (i.e. *not* heavy-tailed)	**Source**
Average call handle-time by call center employee	Beck et al. (2014, Fig. 2, p. 541f.)
Points scored *per minute on court* in the NBA (by basketball player)	Beck et al. (2014, Fig. 12b, p. 554f.)
Output count of various ‘low-’ and ‘medium-complexity’ jobs such as machine	Hunter et al. (1990, Tables 4-6, pp. 33ff.)

operators, mail handlers, file clerks, proofreaders	

Table 3. Share of the right tail in the total for various metrics of ex-post performance (see notes below)

Quantity	**Share of the total held by the top ...**				
	20%	**10%**	**1%**	**0.1%**	**0.01%**
Startup founder equity by company, among Y Combinator companies [80,000 Hours 2014]			>80 %		
Wealth (England & Wales, 1910), whole economy [Roine & Waldenström 2014, Fig. 7.17]		93%	69%		
Wealth (US, 2003), among individuals with net worth > 600 million $ [Newman 2005]	88%	83%	68%	57%	47%
Box Office Gross among directors of major US movies (1970-2018) [Tauberg 2018]	80%				
Donations among EA survey 2019 respondents			57%		
Weeks on NYT Fiction Bestseller list by author with at least 6 weeks on that list [Tauberg 2018]	76%	68%	46%	32%	22%
Wealth (US, 2010), whole economy, by household [Roine & Waldenström 2014, Fig. 7.18]		74%	34%		
Weeks in Billboard Hot-100 by musician, top 5500 artists [Tauberg 2018]	70%				
'Citation shares' (split between coauthors) of papers published in *Nature* 1958-2008, among scientists with roughly above-average citations of that type [Petersen et al. 2010, Table II]	62%	51%	26%	13%	6.6%
Income (worldwide, 2005) [Anand & Segal 2014]		60%	21%		
Income (South Africa, 2011) [Our World in Data]		51%			
'Paper shares' (split between coauthors) published in *Nature* 1958-2008, among scientists with at least the equivalent of one single-authored paper (~8% of all data) [Petersen et al. 2010, Table III]	50%	38%	14%	5.3%	2.0%
Income (US; 2013 for top 10% [Our World in Data], 2005 for top 0.1% [Bakija, Cole, & Heim 2012])		30%		7.3%	
Citations to scientists (whole career) [Sinatra et al.	51%	34%	7.2%	1.3%	.21%

2016][10]					
Box Office Gross by US top-200 movie director [Tauberg 2018]	40%	27%	7.1%	1.9%	.5%
Exponential distribution [see: when this is a sensible comparison?]	*52%*	*33%*	*5.6%*	*.79%*	*.10%*
Weeks in Billboard Hot-100 (1970-2018) by musician, among artists with at least 282 weeks in these charts [Tauberg 2018]	35%	22%	5.0%	1.1%	.25%
Income (Sweden, 2014) [Our World in Data]		22%			
Papers coauthored by mathematicians with at least 133 publications [Clauset et al. 2009]	33%	20%	4.0%	.81%	.16%
Papers written by scientist (whole career) [Sinatra et al. 2016]	39%	24%	4.0%	.59%	.083%
Right half of a standard normal distribution [see: when this is a sensible comparison?]	*44%*	*26%*	*3.6%*	*.45%*	*.052%*
Output in typical jobs ("medium" complexity, e.g. cook) among applicants for such jobs [Hunter, Schmidt, & Judiesch 1990]	58%	31%	3.7%	.41%	.045%
Output in typical jobs ("low" complexity, e.g. mail carrier) among applicants for such jobs [Hunter, Schmidt, & Judiesch 1990]	51%	27%	3.0%	.33%	.035%

Table 3. Share of the right tail in the total for various metrics of ex-post performance, as calculated in this spreadsheet. Italicized are non-performance-related 'benchmarks' we report for comparison. Ordered by descending share of the top 1%. Color scheme: Green = descriptive share in observed or estimated data; Yellow = predicted share by log-normal model; Orange = predicted share by power-law model[11]; Blue = shares in non-heavy-tailed distributions for comparison.[12]

Table 4. Quantiles as multiple of the median for various metrics of ex-post performance (see notes below).

Quantity	**Quantiles as multiple of median**				
	.8	**.9**	**.99**	**.999**	**.9999**

[10] See footnote 4.
[11] We used a continuous power law, i.e. a Pareto distribution, even for discrete data. Our best guess is that this doesn't make much of a difference for this purpose, but haven't checked.
[12] Figures based on models are less affected by noise and allow us to extrapolate beyond the range of observed data (e.g. there aren't actually 10,000 US citizens with net worth > 600 million). On the other hand, such extrapolated numbers may be misleading because the models may be invalid beyond the range of observed data (cf. the appendix).

Equity held by startup founders after startup death or acquisition [80,000 Hours 2014, and paper linked there]	Infinite – about 75% of founders (so incl. the median) end up with nothing.				
Wealth (US, 2003), among individuals with net worth > 600 million $ [Newman 2005]	2.3	4.4	36	300	2500
Donations among EA survey 2019 respondents	7.3	17	160		
Weeks on NYT Fiction Bestseller list by author with at least 6 weeks on that list [Tauberg 2018]	2.1	3.8	26	180	1200
'Citation shares' (split between coauthors) to papers published in *Nature* 1958-2008, among scientists with roughly above-average citations of that type [Petersen et al. 2010, Table II]	1.9	3.1	16	80	400
Income (worldwide, 2005) [Anand & Segal 2014, 11.5.1, median stipulated from Table 11.5]			19		
Income (worldwide, 2005) [Anand & Segal 2014, stipulated from Table 11.5]	2.7	4.5	15	37	78
Income (US, 2005, pre-tax) within 10 highest-paying professions (e.g. medicine, law) [80,000 Hours]			6.6–28		
'Paper shares' (split between coauthors) published in *Nature* 1958-2008, among scientists with at least the equivalent of one single-authored paper (~8% of all data) [Petersen et al. 2010, Table III]	1.7	2.5	9.5	36	130
Citations to scientists (whole career) [Sinatra et al. 2016][13]	2.1	3.1	7.5	14	25
Income (US, 2013) [LIS Database]		2.2			
Box Office Gross by US top-200 movie director [Tauberg 2018]	1.5	2.0	5.3	14	38
Exponential distribution *[see: when this is a sensible comparison?]*	*2.3*	*3.3*	*6.6*	*10*	*13*
Weeks in Billboard Hot-100 (1970-2018) by musician, among artists with at least 282 weeks in these charts [Tauberg 2018]	1.4	1.8	3.9	8.8	20
Income (Sweden, 2005) [LIS Database]		1.6			
Papers coauthored by mathematicians with at least 133 publications [Clauset et al. 2009]	1.3	1.6	3.3	6.6	13
Papers written by scientist (whole career) [Sinatra et al. 2016]	1.6	2.1	3.8	5.9	8.5

[13] See footnote 2.

Right half of a standard normal distribution [see: when this is a sensible comparison?]	*1.9*	*2.4*	*3.8*	*4.9*	*5.8*
Output in typical jobs ("medium" complexity, e.g. cook) among all job applicants [Hunter, Schmidt, & Judiesch 1990]	1.3	1.4	1.7	2.0	2.2
Output in typical jobs ("low" complexity, e.g. mail carrier) among all job applicants [Hunter, Schmidt, & Judiesch 1990]	1.2	1.2	1.4	1.6	1.7

Table 4. Quantiles as multiple of the median for various metrics of ex-post performance, as calculated in this spreadsheet. Italicized are non-performance-related 'benchmarks' we report for comparison. Color scheme: Green = descriptive values in observed or estimated data; Yellow = predicted by log-normal model; Orange = predicted by lower-law model[14]; Blue = predicted by non-heavy-tailed model

Table 5. Coefficient of variation of various metrics of ex-post performance (see notes below).

Quantity	**Coefficient of variation, i.e. stdev/mean**
Wealth (US, 2003), among individuals with net worth > 600 million $ [Newman 2005]	infinity (i.e. infinite standard deviation but finite mean)
Weeks on NYT Fiction Bestseller list by author with at least 6 weeks on that list [Tauberg 2018]	infinity
'Citation shares' (split between coauthors) to papers published in *Nature* 1958-2008, among scientists with roughly above-average citations of that type [Petersen et al. 2010, Table II]	infinity
Income (worldwide, 2005) [Anand & Segal 2014, stipulated from Table 11.5]	1.71
'Paper shares' (split between coauthors) published in *Nature* 1958-2008, among scientists with at least the equivalent of one single-authored paper (~8% of all data) [Petersen et al. 2010, Table III]	infinity
Citations to scientists (whole career) [Sinatra et al. 2016][15]	1.06
Box Office Gross by US top-200 movie director [Tauberg 2018]	1.10
Exponential distribution [see: when this is a sensible comparison?]	*1*

[14] We used a continuous power law, i.e. a Pareto distribution, even for discrete data. Our best guess is that this doesn't make much of a difference for this purpose, but we're not sure.
[15] See footnote 4.

Pareto distribution with cdf shape parameter alpha = sqrt(2)+1 (around 2.41)[16]	*1*
Weeks in Billboard Hot-100 (1970-2018) by musician, among artists with at least 282 weeks in these charts [Tauberg 2018]	.638
Papers coauthored by mathematicians with at least 133 publications [Clauset et al. 2009]	.483
Papers written by scientist (whole career) [Sinatra et al. 2016]	.625
Output in typical jobs ("high" complexity, e.g. physician) among applicants for such jobs [Hunter, Schmidt, & Judiesch 1990]	.475
Output in typical jobs ("medium" complexity, e.g. cook) among applicants for such jobs [Hunter, Schmidt, & Judiesch 1990]	.318
Output in typical jobs ("low" complexity, e.g. mail carrier) among applicants for such jobs [Hunter, Schmidt, & Judiesch 1990]	.193
Height of contemporary adult US men [Wikipedia]	*.0429*

Table 5. Coefficient of variation – i.e. the standard deviation as fraction of the mean – of various metrics of ex-post performance, as calculated in this spreadsheet. Italicized are non-performance-related 'benchmarks' we report for comparison. Color scheme: Green = descriptive values in observed or estimated data; Yellow = predicted by log-normal model; Orange = predicted by lower-law model[17]*; Blue = predicted by non-heavy-tailed model*

Scientific achievement is heavy-tailed *ex-ante*

On **academic performance measured by citations**, there is evidence suggesting that **performance can be well predicted by a product of a person-internal factor and luck, *both* of which are heavy-tailed** (see *Table 6*). In addition, for scientists at least about 15 years into their career, **we can estimate the person-internal factor** based on their citation record. Thus, **in at least one highly relevant case, there is *direct* empirical evidence in favor of a heavy-tailed ex-ante performance distribution**.

Quantity	**Share of the total held by the top ...**				
	20%	**10%**	**1%**	**0.1%**	**0.01%**

[16] The coefficient of variation of a Pareto distribution is independent of its scale parameter.
[17] We used a continuous power law, i.e. a Pareto distribution, even for discrete data. Our best guess is that this doesn't make much of a difference for this purpose, but we're not sure.

'Luck' factor *p* proportional to expected number of citations per paper for a fixed scientist, by paper	55%	38%	8.7%	1.7%	.29%
Citations to scientists (whole career)[18]	51%	34%	7.2%	1.3%	.21%
Exponential distribution	*52%*	*33%*	*5.6%*	*.79%*	*.10%*
Papers written *N*, by scientist (whole career)	39%	24%	4.0%	.59%	.083%
Right half of a standard normal distribution	*44%*	*26%*	*3.6%*	*.45%*	*.052%*
Scientist's 'ability factor' *Q* proportional to the expected number of citations per paper	35%	21%	3.1%	.42%	.056%

Table 6. Top shares of distributions relevant to scientific citations from Sinatra et al. (2016), as calculated in this spreadsheet. Color scheme: Yellow = log-normal distributions from Sinatra et al. (2016); Blue = non-heavy-tailed distributions for comparison

Quantity	**Quantiles as multiple of median**				
	.8	**.9**	**.99**	**.999**	**.9999**
'Luck' factor *p* proportional to expected number of citations per paper for a fixed scientist, by paper	2.3	3.4	9.4	20	36
Citations to scientists (whole career)[19]	2.1	3.1	7.5	14	25
Exponential distribution	*2.3*	*3.3*	*6.6*	*10*	*13*
Papers written by scientist (whole career)	1.6	2.1	3.8	5.9	8.5
Right half of normal distribution	*1.9*	*2.4*	*3.8*	*4.9*	*5.8*
Scientist's 'ability factor' proportional to the expected number of citations per paper	1.5	1.8	2.9	4.1	5.5

Table 7. Top quantiles as multiple of the median for distributions relevant to scientific citations from Sinatra et al. (2016), as calculated in this spreadsheet. Color scheme: Yellow = log-normal distributions from Sinatra et al. (2016); Blue = non-heavy-tailed distributions for comparison

We briefly remark that there also is **some evidence suggesting heavy-tailed ex-ante citations and productivity specifically in the tails**, albeit based on just one discipline (mathematics) and a different predictor: each additional point scored on the International Mathematics Olympiad "is associated with a 2.6 percent increase in mathematics publications and a 4.5 percent increase in mathematics citations" (Agarwal & Gaulé 2018, p. 3). In other words, **the ex-ante distribution of citations (or productivity) conditional on IMO score is log-normal**.

[18] See footnote 2.
[19] See footnote 2.

Apart from that, our conclusions are based on a *Science* paper by Sinatra et al. (2016). Most of their analysis and the quantitative results reported below are based on a large sample of physics publications. Specifically, they use the dataset of all publications (n > 450,000) in the *Physical Review* family of journals[20] between 1893 and 2010. However, they've checked that their qualitative conclusions are also valid for the cognitive sciences, chemistry, ecology, economics, biology and neuroscience by using data from Web of Science and Google Scholar.

Before performing statistical analysis, Sinatra et al. excluded from their data sets scientists with short careers.[21] In the *Physical Review* dataset, 2,887 scientists remain. All **results should thus be interpreted as being about the population of scientists who regularly publish papers throughout a long career**. 'Citations' generally refers to the number of citations to papers ten years after publication. Sinatra et al. (2016, S1.4, S1.6) perform various robustness checks to ensure their conclusions don't depend on the details of dataset selection or the citation measure.

They assume that citations to papers are independent draws from a product of two factors, a scientist's "ability" and (paper-specific) luck. Specifically, citations for one paper are $Q_i * p$, where Q_i is an 'ability factor' specific to each scientist i (and constant throughout their career) and p is a random factor representing 'luck', with the same distribution for all scientists. For each of the N_i papers that scientist i writes over their career, we take an independent draw of the 'luck' component p.

In this model, the expected total number of citations to a scientist over their whole career thus depends on three things:

- The distribution of the 'luck' component p. (The same for all scientists.)
- The value of the 'ability factor' Q_i. (Different for each scientist.)
- The number N_i of published papers. (Different for each scientist.)

Sinatra et al. assume that each of these three factors is log-normal (where for Q_i and N_i the distribution is across scientists), and that luck is independent of ability and productivity. Using maximum-likelihood estimation, they find the following parameters (ibid., p. aaf5239-3). They refer to the means and covariance matrix of the trivariate normal distribution of log p, log Q, log N.

- mu = (mu_p, mu_Q, mu_N) = (0.92, 0.93, 3.34)

[20] Physical Review A, B, C, D, E, I, L, ST, and Review of Modern Physics

[21] In more detail, they include only "scientists that (i) have authored at least one paper every 5 years, (ii) have published at least 10 papers, (iii) their publication career spans at least 20 years in the APS dataset and at least 10 years in the WoS dataset". (Sinatra et al. 2016., S1.3)

$$\sum \equiv \begin{pmatrix} \sigma_p^2 & \sigma_{p,Q} & \sigma_{p,N} \\ \sigma_{p,Q} & \sigma_Q^2 & \sigma_{Q,N} \\ \sigma_{p,N} & \sigma_{Q,N} & \sigma_N^2 \end{pmatrix}$$

$$= \begin{pmatrix} 0.93 & 0.00 & 0.00 \\ 0.00 & 0.21 & 0.09 \\ 0.00 & 0.09 & 0.33 \end{pmatrix}$$

Sinatra et al. performed various statistical checks to support the validity of their model. In particular, they statistically rejected a simpler model that assumed no ability differences between scientists (ibid., Figs. 3CDE), and they showed that the data is consistent with randomness and constant ability *within* fixed careers (ibid., Figs. 2 and 5).[22] However, they did not compare their log-normal model to other heavy-tailed distributions (cf. our appendix); therefore, we think that for the purpose of extrapolating beyond the range of observed data their results should at most be considered weak evidence in favor of a log-normal distribution *in particular* (as opposed to, e.g., a power law).

A few qualitative conclusions from this model are:

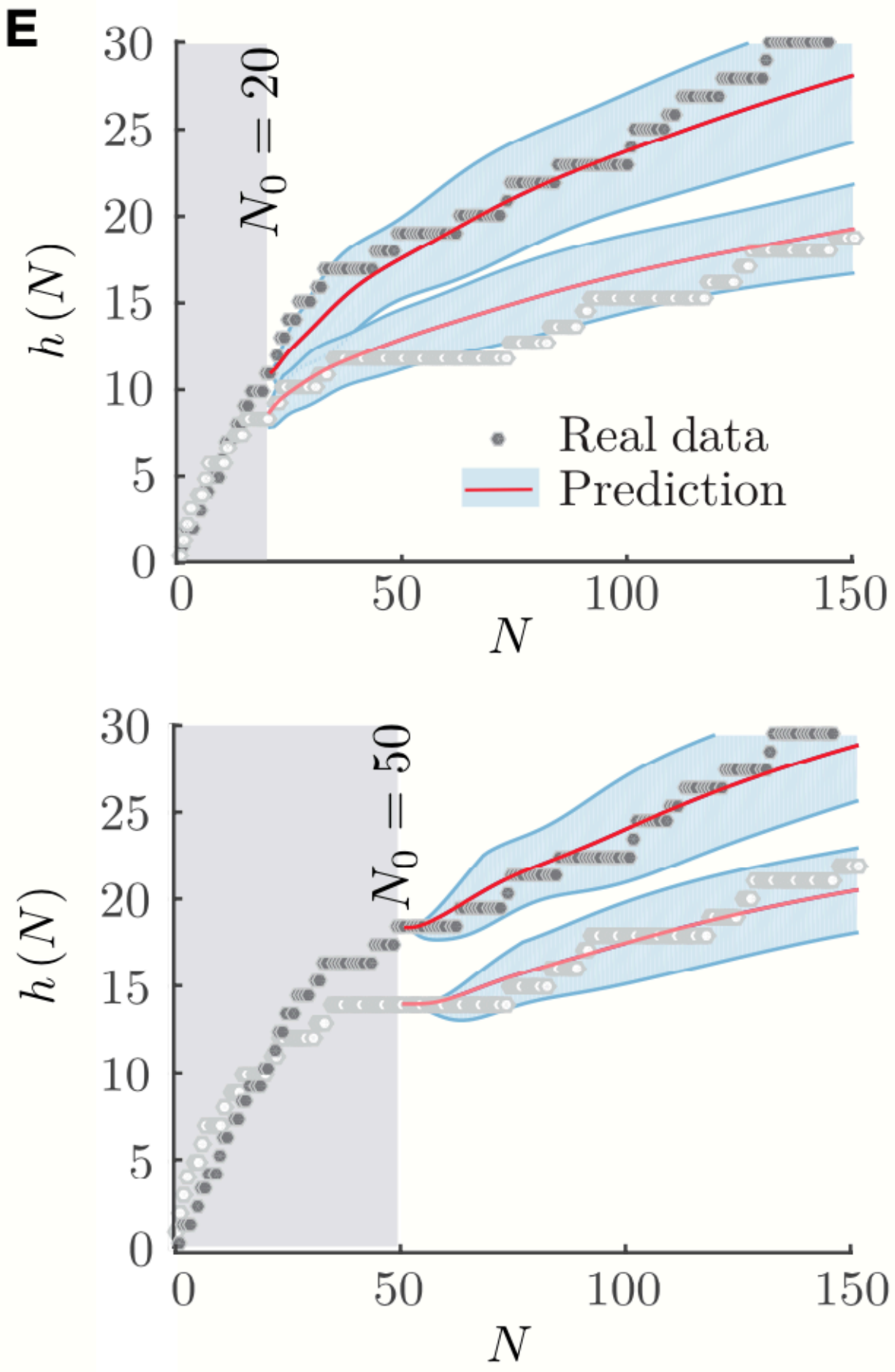
- Average citations per paper *Q* and the number of publications *N* are positively correlated, but only very weakly.
- For *a single paper*, the variance in citations is dominated by luck. However, since career scientists publish many papers, when comparing whole careers the effect of luck 'averages out'. That is, the variance of total citations *over full careers* is mostly *not* due to differences in luck.
- Productivity varies a bit more between scientists than the 'ability factor' *Q*; however, the latter improves citations for *each* paper, thus having a large effect over a whole career.
- As a consequence of the previous two points, ***total* citations vary dramatically between scientists primarily because of differences in the ability factor *Q_i*** (e.g. ibid., Fig. 3E).
- Total citations are more heavy-tailed than each factor individually: we'll see disproportionately many citations to scientists who have high ability *and* high productivity *and* got lucky.

[22] Liu et al. (2018) add an interesting wrinkle: they find evidence of "hot streaks" within scientific – as well as artistic and cultural – careers, i.e. short periods of increased performance. However, they still find that the *timing* of such hot streaks within each career is random (i.e. each piece of work has the same probability of starting a hot streak, no matter whether it's early or late in a career). This contradicts Sinatra and colleagues' assumption that, for a given scientist, the expected number of citations to any single paper is constant throughout a career, and determined for each paper independently. For example, on Sinatra and colleagues' model, the locations of the most-cited and second-most-cited paper within each career should be independent, but Liu et al. find a higher chance of them being close to each other. However, at a more coarse-grained level the results from Liu et al. (2018) and Sinatra et al. (2016) are consistent, and in particular they both find that any paper is as likely as any other to be a scientist's most-cited one.

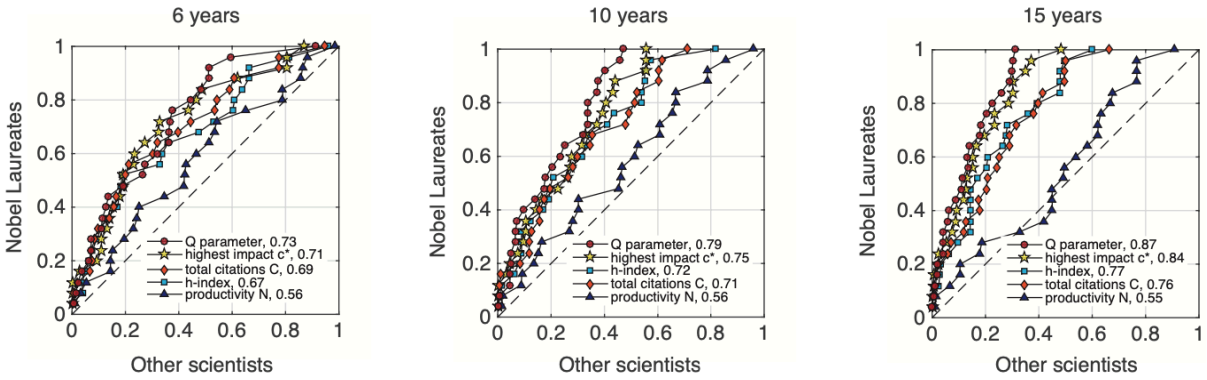
Of course, these are conclusions from a model fitted to *ex-post* data. However, Sinatra et al. also show that we can do reasonably well at *predicting* future citations based on estimating the 'ability factor' *Q_i* from just the early-career citation record. For example, this figure (ibid., Fig. 6E) illustrates how we can predict the Hirsch index *h* – a popular citation measure – based on the first 20 or 50 papers:[23][24]

[23] Note that the evolution of the Hirsch index depends on two things: (i) citations to future papers and (ii) the evolution of citations to *past* papers. It seems easier to predict (ii) than (i), but we care more about (i). This raises the worry that predictions of the Hirsch index are a poor proxy of what we care about – predicting citations to future work – because successful predictions of the Hirsch index may work largely by predicting (ii) but not (i). This *does* make Sinatra and colleagues' ability to predict the Hirsch index less impressive and useful, but the worry is attenuated by two observations: first, the internal validity of their model for predicting successful scientific careers is independently supported by its ability to predict Nobel prizes and other awards; second, they can predict the Hirsch index over a very long period, when it is increasingly dominated by future work rather than accumulating citations to past work.

[24] Acuna, Allesina, & Kording (2012) had previously proposed a simple linear model for predicting scientists' Hirsch index. However, the validity of their model for the purpose of predicting the quality of future work is undermined more strongly by the worry explained in the previous footnote; in addition, the reported validity of their model is inflated by their heterogeneous sample that, unlike the sample analyzed by Sinatra et al. (2016), contains both early- and late-career scientists. (Both points were observed by Penner et al. 2013.)


E
$N_0 = 20$
$h(N)$
N
Real data
Prediction
$N_0 = 50$
$h(N)$
N

Similarly, estimates of *Q_i* based on the first 6, 10, or 15 years of publication activity do better at predicting Nobel prize winners than other metrics including the total number of citations or the Hirsch index, as shown in the following ROC plots (ibid., Fig. S48):



In these plots, the dashed diagonal would represent a predictor that's no better than chance, and predictors are more valid the further their curves are shifted to the left. Two interesting observations are that predictions based on productivity are barely better than chance, and that predictions based on *Q_i* get much better if based on more years of publication activity, especially in the upper tail.

(Note that, to perform e.g. the estimation 'based on the first 6 years' of publication activity, we'd in fact need to wait 16 years since everything is based on citations *10 years after* publication.)

These predictions are much better than chance, but their reliability is far from perfect: if we estimate *Q_i* based on the first 20 papers, then for about 40% of scientists with productive careers of > 70 papers the predicted Hirsch index will be off by more than two standard deviations (ibid., Fig. 3G).

More broadly, there is a large field quantitatively studying citations and other scientific metrics – called bibliometrics, scientometrics, or the science of science. For reviews see Clauset, Larremore, & Sinatra (2017) and Fortunato et al. (2018).

We know of measurable predictors of performance in many domains, including for the tails of performance

Our best predictions generally model **performance as depending on three types of factors**:

- Measurable **'person-internal' characteristics** such as intelligence or conscientiousness;
- Measurable **'personal-external' characteristics** such as the size of a market or the difficulty of a task;
- **'Luck'**, i.e. unmeasured additional factors that appear as random variation of performance.

By *person-internal* we roughly mean characteristics that would be *unaffected by changing the environment of an individual*. For example, if a worker changes companies, their person-internal characteristics should stay the same.

The boundary between person-internal and person-external is admittedly fuzzy. It depends on the performance measure, and in particular the time scale over which we observe performance. For example, consider a skill that improves with practice. The level of skill *at a particular point in time* may be 'person-internal'; but if we measure skill-dependent performance over an extended period of time, then the amount of improvement in the skill will be partly 'person-external' (e.g. workers who get more feedback from their managers might improve faster). Conversely, an individual's environment also depends on 'person-internal' characteristics, e.g. more hard-working people may be able to access better universities.

However, in a fixed context we believe it's often reasonably clear which property would count as person-internal, person-external, or luck, and that this will have major practical relevance.

Suppose you know that performance in some domain is heavy-tailed. Should you try to become a high performer in that domain? If performance was purely determined by luck (as e.g. in a lottery), then everyone has the same chance to become a high performer, and it could be worth trying. If instead the key driver of performance was a heavy-tailed personal-external contribution (e.g. amount of available capital), this would mean you should focus your efforts on modifying your environment accordingly (e.g. seek large amounts of funding). And if performance was heavy-tailed primarily because of measurable person-internal characteristics (e.g. the level in certain skills), then the crucial question would be how you measure up on these characteristics – some people will (predictably) perform far better than others.

We **mostly searched for person-internal predictors of performance**, which is reflected in the findings reported below. However, the literature also discusses several person-external predictors of performance. For example, the book *Chasing Stars* (Groysberg 2012) argues that context – or 'firm-specific capital' – is more important to performance than commonly assumed. Regarding academic performance, there are a number of papers investigating the effects that 'superstar' scientists may have on others in their department (e.g. Azoulay, Zivin, & Wang 2010; Waldinger 2012), and similarly for CEOs (e.g. Brown 2011; Ammann, Hoersch, & Oesch 2016). More broadly, there is a small industry trying to identify "peer effects" in academia or companies.

Performance in typical jobs is predicted by general mental ability, but unclear by how much

There is a large literature in industrial-organizational psychology on performance in typical jobs. In fact, Judiesch & Schmidt (2000, p. 529) state that "job performance is perhaps the most important dependent variable in industrial-organizational psychology."

There are dozens of individual studies across multiple decades that report measures of ex-post job performance as well as potential predictors. On one hand, we can thus draw on a lot of primary data. For example, a meta-analysis by Aguinis et al. (2016, p. 3) is based on

“229 datasets including 633,876 productivity observations collected from approximately 625,000 individuals in occupations including research, entertainment, politics, sports, sales, and manufacturing” – and they mostly limited their analysis to studies published since 2006.

On the other hand, this data comes with a lot of challenges such as convenience samples, small sample sizes, unreliable measurements, and data heterogeneity (observations are from different populations in different jobs with different measures of performance). Psychologists disagree on the extent to which it’s possible to statistically ‘correct for’ these problems in order to reach robust conclusions based on pooled data.

On the optimistic end, psychologists Frank Schmidt, John Hunter, and collaborators (e.g. Hunter & Hunter 1984; Hunter et al. 1990; Schmidt & Hunter 1992, 1998, 2004; Schmidt et al. 2016) have in particular stressed the validity of general mental ability (GMA, similar to general intelligence *g* or IQ) as a reasonably strong predictor of job performance across domains. For instance, **Schmidt et al. (2016, Table 1) report a positive correlation of *r* = 0.65 between GMA tests and job performance, the largest correlation among the 31 predictors reported.** (The second-largest correlation is *r* = 0.58 for employment interviews.) They also find that combining GMA with a second predictor doesn’t add much – the highest gain in the correlation coefficient, for adding an integrity test, is 0.13 or 20%. Another theme in that literature is that GMA is a better predictor of performance, and that performance is higher-variance, in more complex jobs (see in particular Hunter & Hunter 1984 and Hunter et al. 1990).

On the pessimistic end, Richardson & Norgate (2015) based on broadly the same data urge for caution, for instance citing the fact that **different ways of correcting for range restriction and measurement unreliability have led two different meta-analyses – Schmidt & Hunter (1998) and Hartigan & Wigdor (1989) – to wildly different reported GMA-job performance correlations of 0.51 and 0.22, respectively.** They also question the findings on the role of job complexity.

We did not try to adjudicate this debate, though one of us (Max) got the tentative impression that the optimistic perspective is closer to the received wisdom in the field. In any case there seems to be no doubt that there is *some* positive correlation between GMA and job performance for most jobs. However, whether the GMA–job performance correlation is closer to 0.2 or 0.7 would make the difference between GMA being one predictor among many (see *Other predictors of performance* below) and the by far single best one (at least for performance in a wide range of typical jobs).

General mental ability predicts a number of other performance-related quantities

GMA also correlates with a number of other quantities that are in some loose sense related to performance. Examples include:

- *Patentable inventions.* For instance, based on data from Finland, Aghion et al. (2017, p. 3) find that “IQ has [...] a direct effect on the probability of inventing which is almost five times as large as that of having a high-income father”.
- *Academic achievement,* e.g. grades in school: see Wikipedia.

- *Income.* For instance, based on a representative sample of the working-age population of 19 high-income countries (total *n* = 69,901), Ganzach & Patel (2018) claim that "there is not much more than *g* [general mental ability]" for predicting wages, at least after controlling for age and sex.
- *Occupational attainment,* i.e. roughly how prestigious one's job is as opposed to how well one performs in a given job. E.g., in a large US data set Schmidt & Hunter (2004, p. 163) found an uncorrected correlation with GMA of 0.65.

Other predictors of performance

The literature has identified other correlates of job performance or performance-related life outcomes. However, our impression is that these are less well supported, are less strongly correlated with performance, or only apply to more specific tasks (e.g. height predicts success in basketball much better than for most other jobs).

With a very quick search, we found only one meta-analysis (Ng et al. 2005) that examines a wide range of different predictors at the same time (human capital, organizational sponsorship, socio-demographic variables, and stable individual differences including personality, GMA, proactivity, and locus of control). It did not, however, include all of the predictors we've encountered in the literature and list below.

Ng et al. (2005) find that:

- The predictors surveyed by them tend to work better for salary level (corrected correlations up to *r_c* = .29, and many above .2) than for promotions (most *r_c* below .1, only one predictor with *r_c* barely above .2).
- These two measures of 'objective career success' tend to be predicted by different variables than 'subjective career success', i.e. career satisfaction. Support from employers, personality, and non-cognitive skills tend to correlate more strongly with subjective career success, while human-capital and socio-demographic variables tend to correlate more strongly with objective career success.
- The 5 best predictors of salary level among the 27 variables surveyed are (*r_c* between .29 and .26, in descending order): Education level, political knowledge & skills, cognitive ability, work experience, age.

One caveat is that Ng et al. include cognitive ability only in their analysis of predictors of salary levels – but not for promotions or career satisfaction. Another limitation is that they don't provide information about correlations *between* predictors (e.g. age and work experience are clearly related); taking their results at face value would thus understate the role of variables that are causally prior to many others (e.g. we would guess that cognitive ability causally contributes to education level; similarly, some personality traits and non-cognitive skills may, via influencing motivation, causally contribute to hours worked etc.). For their full results, see the appendix.

Beyond that, predictors of job performance or career success that appear in the literature include:

- *Personality,* especially conscientiousness and related constructs such as "integrity", "self-discipline", or Duckworth's "grit".

    - Borghans et al. (2016), in a paper published in *PNAS*, analyze data from 4 cohorts from high-income countries with between 347 and 8,874 individuals. They find correlations of 0.29 to 0.45 between personality measures and scores on achievement tests at school, and correlations of 0.25 to 0.43 between personality and grades. (Their reported correlation with IQ is stronger than that for achievement tests, but weaker for grades.) Regarding longer-term outcomes, they state that "Personality is generally more predictive than IQ on a variety of important life outcomes." (p. 13354).[25]
    - Barrick & Mount (1991) performed a meta-analysis of correlations between the Big Five personality traits with 3 measures of job performance (job proficiency, training proficiency, personnel data) in 5 occupational groups (professionals, police, managers, sales, skilled/semi-skilled). They found that conscientiousness correlates at about 0.2 with all measures of performance in all studied occupations. (Though 'uncorrected' correlations were smaller, at most 0.13.) Results for other personality traits were more mixed or inconclusive. Another meta-analysis by Tett, Jackson, and Rothstein (1991) features similar conclusions.
    - However, in the meta-analysis by Ng et al. (2005), conscientiousness does *not* appear to be more predictive of objective career success (salary and promotions) than other Big Five personality traits. Instead, correlations with all personality traits are similarly small, between -0.12 (neuroticism and salary) and 0.18 (extroversion and promotions).
    - Kaufman et al. (2016), in four samples of in total n = 1,035 individuals, find that two different facets of the Big Five trait openness to experience – namely 'openness' and 'intellect' – correlate with achievement in the arts and sciences, respectively.
        - More broadly, there is a recent literature trying to identify correlations with subfacets of the Big Five traits.
    - Using nonstandard personality dimensions and questionnaires administered to 196 biologists, 201 chemists, and 171 physicists, Busse & Mansfield (1984) found that their measure of "commitment to work" correlates with the number of publications, while their measure of "originality" correlates with citations.
- *Non-cognitive skills* (i.e. not primarily cognitive abilities that can be changed through practice or developments rather than stable traits), e.g. "character skills" (Kautz et al. 2017[2014]).
- *Educational attainment* (e.g. highest degree obtained, academic discipline), see e.g. Wai (2014) and this UK government study.
- *Academic performance,* e.g. grades or test scores. For instance, the *PNAS* paper by Borghans et al. (2016, p. 13354) mentioned above states that "both grades and achievement tests are substantially better predictors of important life outcomes than IQ."
- *Organizational sponsorship,* i.e. the extent to which individuals receive career support by their employers. For instance, a meta-analysis by Ng et al. (2005) reports 'corrected' correlation of 0.05 to 0.24 between different measures of organizational

[25] At a glance, Max perceives some claims from Borghans et al. (2016) to be at odds with what he read elsewhere. This might indicate either that his understanding of other views is mistaken or that there is some problem with this study. Max didn't try to resolve this issue.

sponsorship and objective career success (salary or promotions), and of 0.38 to 0.44 between different measures of organizational sponsorship and career satisfaction.
- *Attractiveness.* See e.g. Hamermesh and Biddle (1994) and Hamermesh, Meng, & Zhang (2002)
- *Demographic characteristics,* e.g. age or marital status. For instance, Azoulay et al. (2018, p. 1) found that, in the US, the "mean founder age for the 1 in 1,000 fastest growing new ventures is 45.0".
- *Socio-economic status (SES) of parents*.
    - Strenze (2007), in a meta-analysis of longitudinal studies, finds that "intelligence is a powerful predictor of success [as measured by education, occupation, and income] but, on the whole, not an overwhelmingly better predictor than parental SES or grades".
    - Aghion et al. (2017) find both a direct correlation and an interaction effect with IQ of father income with the probability of becoming an inventor.
- *Career success of parents,* see e.g. 80,000 Hours (2015).
- *Specific skills or abilities* (rather than a *general* ability factor), see e.g. 80,000 Hours (2017) and Grobelny (2018).

We didn't try to be comprehensive and didn't examine any of these studies in more detail.

In several cases, predictors remain valid in the tails

Several studies have specifically examined the tails of performance or the tails of measured predictors. Examples include:

- *Wealth.* Wai & Lincoln (2016) analyze a data set of *n* = 18,245 ultra-high net worth individuals (wealth > $30 million). They find that "smarter (more educated) people were wealthier, gave more, and had more powerful social networks (but when controlling for multiple confounds the association between education/ability and wealth was found to be quite small)" (p. 1).
- *Executive management.* Adams et al. (2018, p. 392) find that in a large sample of Swedish men "the median large-company CEO belongs to the top-17% of the population in cognitive ability, and to the top-5% in the combination of cognitive, non-cognitive ability, and height".
- *GMA and educational attainment as predictors of wealth, income, and influence*. Wai (2014) analyzes *N* = 1,426 billionaires, *N* = 231 'powerful' people (by *Forbes* ranking), and *N* = 2,624 World Economic Forum attendees. In this highly 'elite' sample, he finds that, in the US, top-1%-ability individuals were overrepresented by a factor of 45 to 85, and that "[e]ven within the top 0.0000001% of wealth, higher education and ability were associated with higher net worth, even within self-made and non-self-made billionaires, but not within China and Russia."
- *GMA as a predictor of income.* Gensowski et al. (2011) report a correlation of IQ and lifetime earnings in a prospective cohort study of *n* = 617 high-IQ individuals (IQ > 135). They also find correlations of income with personality and education, and confirm that significant correlations with IQ and personality remain after controlling for education.
- *GMA as a predictor of academic, creative, and scientific achievement.*
    - Park, Lubinski, & Benbow (2008), in a cohort of *n* = 1,586 individuals with exceptional maths abilities assessed during adolescence (top 1% of

performance in the maths portion of the SAT at age 13), find that adolescent SAT scores correlate with the probability of having at least one patent or scientific publication, even after controlling for the highest academic degree obtained. See also Park, Lubinski, & Benbow (2008) and Robertson et al. (2010).
    - Makel et al. (2016), in two cohorts of *n* = 320 and *n* = 259 individuals assessed to be in the top 0.01% of verbal or maths ability before age 13, found markedly higher levels of achievement than in samples of top-1%-ability individuals. In fact, on average their top-0.01% individuals by age 40 had achieved at least as much as top-1% individuals by age 50 (p. 9).
- *IMO scores as a predictor of success in academic mathematics.* Agarwal & Gaulé (2018) find that performance in the International Maths Olympiad (IMO) correlates with various measures of success in academic mathematics[26], e.g. completing a PhD, citation counts, and getting a Fields medal (the most prestigious award in mathematics, comparable to a Nobel Prize). These correlations hold across the whole range of a sample that is in the extreme right tail of maths ability, and seem strong e.g. "the conditional probability that an IMO gold medalist will become a Fields medalist is two order of magnitudes larger than the corresponding probability for a PhD graduate from a top 10 mathematics program." (p. 4) The authors perform two additional analyses to control for confounders. First, they look at the subsample of IMO participants who later got a maths PhD; second, they compare individuals who participated in the IMO in the same year *and* got their PhD from the same university. In both analyses, the correlations remain positive and are almost as large as in the full sample.
- *Height as a predictor of success in basketball.* This *Forbes* article suggests that the proportion of 20-40 year-old men who play in the NBA (the US's top basketball league), as well as the average earnings of basketball players, increases with height up to heights greater than 7 feet, the top 0.000038% of height.

Again, we didn't try to be comprehensive.

Note that even if a predictor remains valid in the tails, ex post the highest performers will usually exhibit very high but not the absolute highest values of the predictor.

Measurable predictors of heavy-tailed ex-post performance don't imply that predicted performance is heavy-tailed

Recall that our best predictions generally model **performance as depending on three types of factors**:

- Measurable **'person-internal' characteristics** such as intelligence or conscientiousness;
- Measurable **'personal-external' characteristics** such as the size of a market or the difficulty of a task;

[26] "Each additional point scored on the IMO (out of a total possible score of 42) is associated with a 2.6 percent increase in mathematics publications and a 4.5 percent increase in mathematics citations." (p. 3) (Correlation with log cites is still around 4% among subsample who got maths PhDs, Table 4.)

- **'Luck'**, i.e. unmeasured additional factors that appear as random variation of performance.

Observing heavy-tailed ex-post performance doesn't *by itself* tell us whether or not any contributing factor of any type (person-internal, person-external, or luck) is heavy-tailed. This is because this observation is consistent with any of the following possibilities:

a) All three factors (internal, external, and luck) being heavy-tailed.
b) A single factor being heavy-tailed, and all other factors being light-tailed or even constant. For instance, a lottery can have heavy-tailed ex-post results that are purely due to luck. For another example, suppose that the amount of investor optimism (measured in the amount of seed funding they're providing for a given startup) was heavy-tailed, and that startup success is the sum (or product) of investor optimism, market size, founders' intelligence, and how much founders work per week. Startup success would then be heavy-tailed even if all these other factors were constant across startups.
c) *No* individual factor being heavy-tailed, but performance depending on the *product* of many factors. This is possible because of the mathematical fact that, under certain conditions that often hold in practice, the product of an increasing number of light-tailed factors will converge toward a heavy-tailed distribution.[27]

Furthermore, **knowing a measurable correlate of heavy-tailed ex-post performance doesn't *by itself* imply heavy-tailed ex-ante performance.**

Here is why. Suppose we know that:

- Some metric of performance *Y* is heavy-tailed ex-post (say, scientific citations); and
- Some characteristic (person-internal or -external, say a scientist's IQ or the ranking of their university) *X* that is measurable ex-ante is positively correlated with *Y.*

The short version is that the predictor might only tell us about a factor that doesn't drive the heavy tail.

Formally, ex-ante performance then is the conditional expected value **E**[*Y* | *X*] (note that unlike the unconditional expected value **E**[*Y*], the conditional expected value is a *random variable,* i.e. something that has a probability *distribution*).

What is **E**[*Y* | *X*]? In practical terms, imagine that you measure *X* for a large number of randomly selected people, thus obtaining a sample of measured values *X* = *x_1, X* = *x_2,* … , *X* = *x_N* (e.g., *x_1* could be the first person's IQ, *x_2* the second person's IQ, etc.). You can then calculate these people's *expected performance y_1* = **E**[*Y* | *X* = x_1], *y_2* = **E**[*Y* | *X* = *x_2*], … , *y_N* = **E**[*Y* | *X* = *x_N*]. Each *y_i* is a single number representing the predicted level

[27] For example, if after taking the logarithm the conditions of the Central Limit Theorem are fulfilled, then the product will converge to a log-normal distribution. We've sometimes encountered the misconception that products of light-tailed factors *always* converge to a log-normal distribution. However, in fact, depending on the details the limit can also be another type of heavy-tailed distribution, such as a power law (see, e.g., Mitzenmacher 2004, sc. 5-7 for an accessible discussion and examples). Relevant details include whether there is a strictly positive minimum value beyond which products can't fall (ibid., sc. 5.1), random variation in the number of factors (ibid., sc. 7), and correlations between factors.

of performance based on *x_i*; for example, *y_i* could be the predicted number of publications by a scientist with IQ *x_i*. The conditional expected value **E**[*Y* | *X*] simply is the *distribution* of the numbers *y_i* that will emerge for large samples sizes *N*.

As a sanity check, if we could predict performance perfectly, then ex-post and ex-ante performance should coincide. And indeed, **E**[*Y* | *Y*] = *Y*. Conversely, if all ex-ante information *X* is irrelevant to ex-post performance *Y* then you can do no better than to predict the unconditional expected value for everyone: and indeed, if *X* and *Y* are statistically independent then **E**[*Y* | *X*] = **E**[*Y*].

The key point is that **ex-ante performance E[*Y* | *X*] can be light-tailed even if ex-post performance *Y* is heavy-tailed and *X* correlates with *Y***. For example, suppose that ex-post performance is the *product* of two *independent* factors *X* and *X'*:

$$Y = X * X'$$

Then, by basic properties of conditional expected values, ex-ante performance is

$$\mathbf{E}[Y \mid X] = \mathbf{E}[X * X' \mid X] = X * \mathbf{E}[X' \mid X] = X * \mathbf{E}[X'].$$

Thus, ex-ante performance is captured wholly by the measurable correlate *X*, up to a constant factor that depends only on the unmeasured part *X'*. In particular, ex-ante performance **E**[*Y* | *X*] is heavy-tailed *if and only if* the measurable predictor *X* itself is heavy-tailed.

This also makes sense intuitively. For example, consider a lottery in which every ticket has the same small chance of winning a fixed price. Suppose we can measure *how many lottery tickets X* each participant has bought. We then know that ex-post lottery winnings are heavy-tailed and can measure a correlate *X* of these heavy-tailed winnings – but the shape of our distribution of *predicted* lottery winnings will look exactly like the distribution of observed ticket sales. Ex-ante lottery winnings will be normally distributed if and only if ticket sales were normally distributed; ex-ante lottery winnings will follow a power law if and only if ticket sales followed a power law; and so on.

How would this look like in more relevant toy models? Suppose that 'performance = intelligence * luck', with intelligence being normally distributed and measurable, luck being log-normally distributed and unmeasurable, and the two factors being independent. Then performance would be heavy-tailed, intelligence would be a measurable predictor of performance, but the ex-ante distribution of predicted performance based on intelligence would be *normally* distributed (i.e. thin-tailed):

$$\mathbf{E}[\text{performance} \mid \text{intelligence}] = \text{intelligence} * \mathbf{E}[\text{luck}]$$

Our best guess is that heavy-tailed ex-ante distributions *are* widespread, at least for expert performance on complex tasks, such as scientific research or organizational leadership. **However, this guess relies more on priors and broad gestalt impressions of the world rather than the specific evidence we investigated here.**

Uncertainty should move us toward acting as if ex-ante performance was heavy-tailed – but not all the way

At first glance there is an argument to act *as if* ex-ante performance was heavy-tailed even in cases where we're uncertain: Suppose, for instance, we're uncertain whether the ex-ante distribution has one of two forms, *X_heavy* or *X_thin*, the former being more heavy-tailed. If we have credence *p* in *X_heavy* and we use expected value to account for our uncertainty, then we should act as if ex-ante performance was distributed like *p * X_heavy + (1-p) * X_thin*. And this sum becomes *as heavy-tailed as* X_heavy if we look sufficiently far down the tail.

However, in practice we usually aren't interested in the limit of infinitesimally unlikely tail events but in a fixed quantile, say the top 1%, compared to the median. At any fixed quantile, the sum *p * X_heavy + (1-p) * X_thin* will be *more* heavy-tailed than *X_thin* but not all the way as heavy-tailed as *X_heavy.* Therefore, **in practice, uncertainty about the tails of ex-ante performance should move us *some but not all the way toward the hypothesis of heavy-tailed ex-ante performance***, by an amount that depends on our credence in the heavy-tailed hypothesis.

Causal models of performance would be useful, but we haven't found one that would be 'shovel-ready' for making predictions in EA contexts

Due to the limitations of observational research – e.g. distinguishing correlation from causation, or distinguishing different heavy-tailed distributions from one another – it would be very helpful to have a causal *theory* of performance: understanding how various resources, traits, and behaviors interact to actually produce performance.

Of particular interest in our context is whether different factors used to explain performance – whatever they are – combine *additively* or *multiplicatively* to yield performance. This is because **the *sum* of many light-tailed factors will usually again be light-tailed, while their *product* will be heavy-tailed**. (Of course, in general a sum can be heavy-tailed as well, e.g. if one of the summands was heavy-tailed itself.)

There is a **debate in industrial-organizational psychology on whether job performance is better modeled as the sum or product of employee traits such as intelligence and personality** (e.g. Sackett, Gruys, & Ellingson 1998 analyze four data sets that support an additive model, thus questioning three earlier papers that found support for a multiplicative model; a recent meta-analysis by Van Iddekinge et al., 2017, also favors an additive model). This debate is thus highly relevant: if traits multiply, then we should job performance to be more heavy-tailed than if the same traits add. On the other hand, like most psychology work on job performance, this debate seems largely based on data from 'typical' jobs rather than high-complexity areas such as science or upper management. Since we were more interested in the latter, **we didn't review this debate in more detail and didn't try to form our own view**.

Beyond that, we found a number of claims about specific causal mechanisms – for example, the 'Matthew effect' (e.g. Merton 1968) according to which the 'rich get richer', i.e. success begets further success. Unfortunately, we found it hard to vet these claims or to synthesize them into a comprehensive theory. So we just list them here:

- Schmidt & Hunter (2004, p. 170) review evidence for the causal hypothesis that general mental ability predicts job performance because it helps with the acquisition of job knowledge, which in turn causes better performance.
- Kremer's (1993) famous "O-ring theory of economic development" posits that many economic production processes consist of many steps, at each of which the whole process can fail, resulting in a product of zero value. (Similar to how the whole Space Shuttle Challenger exploded because a single part – an 'O-ring' – failed.) Kremer proposes a multiplicative model to capture this property. He describes several implications, for example that maximization of total output leads to 'assortative matching', i.e., a division into consistently high-quality and consistently low-quality production processes (e.g. the most able employees will flock to the same few 'elite' firms).
- Rosen (1981) presents potential causes for an increase in highly concentrated markets (which have a heavy-tailed distribution of e.g. revenue across sellers) such as imperfect substitution or zero marginal cost.
- Shockley (1957, pp. 284ff., sc. VI) presents two hypotheses that could explain the heavy-tailed distributions of scientific citations: one is that there are small differences (normally distributed) in *how many ideas people can consider simultaneously*, which results in heavy-tailed performance differences because the total number of idea combinations one can consider increases rapidly with this parameter; the second is that publishing papers depends on the *multiplicative* interactions of many traits such as "1) ability to think of a good problem, 2) ability to work on it, 3) ability to recognize a worthwhile result, 4) ability to make a decision as to when to stop and write up the results, 5) ability to write adequately, 6) ability to profit constructively from criticism, 7) determination to submit the paper to a journal, 8) persistence in making changes" (ibid., p. 286).
    - Both of these explanations are speculative. Indeed, the relevant section is titled "Speculations on the origin of the log-normal distribution" (ibid., p. 284).
- To explain career success, some psychology work (e.g. Turner 1960, Spilerman 1977, Rosenbaum 1984, Dreher & Ash 1990) distinguishes between a "contest-mobility model" (an increasingly small number of positions is allocated to the best applicants) and a "sponsored-mobility model" (career progression depends on how much organizations 'invest into' their employees).
- Psychologist Angela Duckworth (who pioneered the study of 'grit') has suggested that "Performance = Skill * Effort" and "Skill = Talent * Effort", thus resulting in the model that "Performance = Talent * Effort^2".
- Gensowski (2018, p. 177) hypothesizes that, in a sample of high-IQ men, conscientiousness and extraversion predict lifetime earnings because personalities high on these two traits accumulate human capital in school at a higher rate, which in turn allows people to perform higher-paid work.

More theoretically, for common types of distributions (normal, exponential, log-normal, Pareto, etc.) we can ask which sort of mathematical processes will generate them. The

earlier observation on additive vs. multiplicative processes is a special case of this. Understanding the possible origins of different distributions could be useful because, in addition to looking at ex-post performance data, we could then infer the type of the performance distribution from empirical information on the origins of performance.

We didn't pursue this line of investigation, but here are some examples:

- Two high-level insights are the principle of maximum entropy (e.g. Frank 2009) and the fact that for certain kinds of processes the stable distributions are the only possible attractors.
- On generating mechanisms for power laws, see Newman (2005).

Why we'd guess that ex-ante performance at complex tasks is often heavy-tailed

If we look beyond the literature we've reviewed for this post and consider all relevant evidence (including things like gut feelings), our best guess is that, for complex tasks, it will often be possible to identify predictors relative to which ex-ante performance is heavy-tailed.

We only gesture at why we (tentatively) believe this, in a way that we expect won't necessarily be convincing to people who have different impressions.

First, recall that we've found many examples of *ex-post* performance being heavy-tailed. We think there are theoretical reasons to expect this property to be widespread for many tasks, at least if performance is cashed out in terms of 'impact on the world' in some sense. Specifically, it seems that for the effects of many 'complex' tasks there is a metric that can range over many orders of magnitude and depends on a complicated combination of largely independent factors. *If* such a model is correct, then there are mathematical reasons to expect a heavy-tailed distribution.

As an example, consider the contribution a CEO makes to the profit of their company. This seems to depend on many factors such as their cognitive skills, their personality, their health, how well their personal life is going, the actions of various people in their company, the actions of competitors, 'exogenous' events such as natural disasters, political developments, etc. – many of these seem to be independent from others, e.g. whether San Francisco is hit by an earthquake does not depend on the CEO's skills or personality. At the same time, many of these factors seem to *interact* – e.g. the impact an earthquake would have on company performance *does* seem to depend on the CEO's skills (have they put safeguards in place? how quickly would they be able to resume production? etc.). This suggests that the CEO's contribution to profit depends on a complicated combination of largely independent factors.

Similarly, many complex tasks can be broken down into successive steps of simpler tasks, such that the task can fail at each step. This suggests a multiplicative model similar to Kremer's O-ring theory. E.g. it is often argued that heavy-tailed paper citations arise because there are many steps involved in a scientific publication: having a good idea, finding collaborators, running an experiment, analyzing the data, writing the paper, responding to reviewer comments, etc.

Second, why expect heavy-tailed *ex-ante* performance? Basically we would guess that in many cases where performance depends on the combination of many factors we will be able to measure (correlates of) a significant fraction of these factors – or variables that themselves depend on a combination of many of the same factors. This means that we can measure a predictor of performance which *itself* depends on a combination of many factors, and thus is heavy-tailed.

Or, alternatively, we might be able to measure *one* predictor that correlates with *many* of the performance-determining factors. This is particularly plausible when performance depends on a combination of cognitive tasks since it is well established that performance on such tasks is positively correlated (the "positive manifold" of cognitive abilities), and it's possible to psychometrically measure a 'general mental ability' factor that is positively correlated with performance on all these specific tasks.

As a toy example, suppose that performance Z depends on the product of 50 measurable factors X_i and 50 unmeasurable factors Y_j, all of which are mutually independent. Denote the product of all X_i with X, and the product of all Y_j with Y. Then by the same calculation as in an earlier subsection, $\mathbf{E}[Z \mid X\_1, \ldots, X\_50] = X * \mathbf{E}[Y]$, and X is heavy-tailed because it is the product of many factors. This is not *literally* what we'd encounter in practice, e.g. because the things we can measure are rarely mutually independent. But the analog argument still goes through for more complicated models, and so we think the toy model is a good illustration for why we think heavy-tailed ex-ante performance is widespread.

(Similarly, if we can't directly determine the value of any individual X_i but can only measure some variable X' that correlates with all X_i, we believe it follows that $\mathbf{E}[Z \mid X']$ is heavy-tailed – though we haven't checked this.)

This still seems true to at least some extent if we restrict ourselves to 'person-internal' predictors. For example, intelligence and motivation seem both relevant for performance at many tasks, and do seem to interact: more intelligent people can make more use of their motivation and vice versa – it's not like performance in the first half of the day depends only on motivation and performance in the second half only on intelligence. In a more fine-grained model of cognition, intelligence may in turn depend on several interacting factors such as 'processing speed', memory, ability to focus, etc.; task-specific motivation may depend on factors such as sleep, nutrition, genetic contributions to personality traits, and which books one read as a child.

Separately, 'success begets success' dynamics suggest that predictions of longer-term outcomes may be heavy-tailed even if they're based on only one thin-tailed predictor. If we can measure some predictor X such that these dynamics tend to much more strongly[28] 'amplify' success for people with higher values in X, then we should expect *ex-ante* that over

[28] More precisely, it needs to be the case that the *marginal* amount of 'amplification' increases with X. That is, a small increase in X 'helps' you more with success *the larger your value of* X *already was.* This condition does not hold in the lottery counterexample from a previous section: each additional lottery ticket 'amplifies' your expected winning by the *same* amount, no matter how many lottery tickets you already had. But if the chance of winning the lottery depended on e.g. the square of tickets purchased, then each additional ticket would be more valuable the more tickets you already have.

time they might turn a thin-tailed *X* into a heavy-tailed distribution of success. For example, someone with strong cognitive abilities from a privileged background is more likely to do well in school, therefore is more likely to get into a good university, which in turn means they're more likely to land a first job in which they'll get good mentorship and learn a lot, etc.

If we're looking at a notion of performance that requires high-performers to secure unusually influential and competitive positions, there is some direct evidence that educational attainment is a heavy-tailed ex-ante predictor. That is, some of these positions are dominated by graduates from the very few top universities. For example, a UK government study found that more than ½ of UK Cabinet members and more than ⅔ of Senior Judges have a degree from Oxford or Cambridge. Similarly, Wai (2014, p. 54) found that:

> *"[R]oughly 34% of billionaires, 31% of self-made billionaires, 71% of powerful males [by* Forbes *ranking], 58% of powerful females, and 55% of Davos participants attended elite schools worldwide. [...] In the U.S., top 1% ability individuals were highly overrepresented: 45 times (base rate expectations) among billionaires, 56 times among powerful females, 85 times among powerful males, and 64 times among Davos participants. [...] Even within the top 0.0000001% of wealth, higher education and ability were associated with higher net worth, even within self-made and non-self-made billionaires, but not within China and Russia. [...] These global elites were largely drawn from the academically gifted, with many likely in the top 1% of ability."*

Another data point is from the Canadian Inventors Assistance Program[29] (IAP). Inventors can pay the IAP to predict the success of their invention. Many then try to develop and market their invention even if the IAP was pessimistic about commercial viability. This means we have data on the accuracy of the IAP's predictions, and we know that 55% of highest-rated inventions achieve commercial success, compared to 0% for the lowest rating.

Finally, another argument is based on the evidence showing that predicting future citations based on past citations results in a heavy-tailed distribution. We think this is at least weak evidence that the phenomenon of "predicted performance conditional on past performance is heavy-tailed" is more widespread: put differently, we can't think of a plausible reason why this relationship would be highly specific to science.

Beyond these explicit arguments we've tried to gesture at, we also feel our take is supported by our broad impression of recruiting practices in highly competitive fields, anecdotes from our own experience, and other broad gestalt impressions of the world.

Further research

Here are some avenues for further research which we think might be promising, especially for people whose background is a good fit for answering some of these questions.

[29] H/T Ben West for making us aware of this data.

They are in no particular order, and lightly held suggestions rather than carefully vetted and strongly recommended research projects. Their value and tractability likely differs substantially.

- **What can we say about the *negative* tails of performance or impact?** In particular, when is negative impact heavy-tailed? We know that harmful things can have heavy-tailed distributions, e.g. earthquake intensity, forest fire size, or war casualties (see e.g. Clauset et al. 2009). But data on negative impact by people is scarce as most performance metrics are by definition restricted to positive values. Can we learn anything from existing metrics that can take both positive and negative values?
    - See also Kokotajlo & Oprea (2020) for an argument for why this question is important for EA. They also provide an argument for why we should expect heavy tails of negative impact to be common.
    - We do have data on this from some domains, e.g.:
        - The negative tail of financial returns looks similar to the positive one (e.g. Jondeau and Rockinger 2001).
        - Ben found a negative tail in a cost-benefit analysis for about 370 US social policies (archive), compiled by the Washington State Institute for Public Policy benefit-costs results database.
        - A negative dimension of job performance that has been extensively studied in industrial-organizational psychology are "counterproductive work behaviors" such as bullying, lateness, or theft (e.g. Dallal 2005). At first glance, these seem less relevant in many EA contexts, but is there anything useful we can infer from this literature?
- **Suppose that for some task the true ex-post distribution of performance is very heavy-tailed across people but our ability to predict performance is very limited. Which heuristics should we adopt in such a world?** Should we e.g. rely more or less on gut judgments, allocate resources by lottery, or try to learn from analogs such as venture capital and science funding?
- **Do tasks differ in whether we get increasing or decreasing returns (in terms of altruistic impact) to better performance?** As an extreme possibility, is it the case that one needs to exceed some performance threshold to have *any* impact through work in early-stage research fields without established questions or methods (such as perhaps some areas of AI safety)?
    - Put differently, what can we say about ex-ante *altruistic impact*, i.e. the conditional expected value **E**[altruistic impact | performance predictor]? (Rather than just ex-ante *performance*, i.e. **E**[performance metric | performance predictor].)
- At a high level, **we can distinguish different types of interventions aimed at increasing the EA community's total impact: better allocation of existing resources, e.g. improving hiring processes helping people identify which job they're the best fit for; intensive growth, e.g. helping current EAs to improve their skills; and extensive growth, which could be either untargeted or aimed at particular audiences, e.g. promoting EA in mass media versus giving EA-related material to IMO participants. What are the key parameters that determine how cost-effective these different types of intervention are?** For

instance, what's a good way to operationalize how good current hiring and funding processes are, and how costly it would be to improve them?

- **Are there any sources of data that are more directly relevant to EA use cases, thus ameliorating worries about external validity?** For example, what do we know about the distributions of donations to EA organizations, karma on various EA fora, the number and value of behavior changes caused by EA conferences, or the number and value of plan changes caused by 80,000 Hours?
    - Partial answers:
        - Donations: EA Survey [2019, 2018, 2017]
        - Plan changes influenced by 80,000 Hours: Annual reviews [2019, 2018]
- Denrell & Liu (2012) show that, **when using predictors of wildly different reliability, then naive selection by best predicted performance can be *predictably* suboptimal**. (This is roughly because a very high level of predicted performance is disproportionately likely due to a large prediction error for one of the low-reliability predictors.) This is an extension of the familiar Optimizer's Curse (Smith & Winkler 2006). **What are the implications of this finding? Are its conditions ever plausible fulfilled in practice (perhaps when comparing interventions or cause areas using very different types of evidence)?**
- What should we conclude from the debate in industrial-organizational psychology on **whether job performance is better modeled as the sum or product of employee traits such as intelligence and personality** (e.g. Sackett, Gruys, & Ellingson 1998; Van Iddekinge et al., 2017)?
- Can we make the following statement more precise, and what does this imply in practice? "If we can measure some predictor *X* such that 'success-begets-success' dynamics tend to much more strongly 'amplify' success for people with higher values in *X*, then we should expect *ex-ante* that over time they might turn a thin-tailed *X* into a heavy-tailed distribution of success."
- **Can we infer anything useful from theoretical statements on which kinds of stochastic processes will result in which type of distribution?** (See the end of our section on *Causal models of performance* for a brief discussion.)

Appendix

High variance vs. heavy tails

Which properties of the performance distribution are particularly interesting? Both the academic literature and previous discussions in EA have sometimes focused on variance and sometimes on heavy tails.[30]

[30] E.g., CEA's (deprecated) page on their current thinking has a section *Talent is high variance*, while Owen Cotton-Barratt's popular talk *Prospecting for gold* includes a section on *Heavy-tailed distributions*. In the psychology literature on job performance, Hunter, Schmidt, & Judiesch (1990) focus on variance, while Aguinis et al. (2016) focus on heavy tails.

These are distinct concepts – **a heavy-tailed distribution can have arbitrarily small variance, and a light-tailed distribution can have arbitrarily high (finite) variance**.[3132]

Depending on the purpose of your analysis, you might care about variance, heavy tails, or both. Here we won't make claims about which is more important when, but simply try to explain how they differ.

(One caveat is that there are *different* definitions of "heavy-tailed" in the literature. Throughout this post we take *heavy-tailed* to roughly mean *having heavier tails than an exponential distribution*. For instance, we consider any log-normal distribution to be heavy-tailed. There are other definitions that impose a tighter relationship between heavy tails and variance, e.g. ones that require heavy-tailed distributions to have infinite variance. For a more formal discussion, see here.)

Both high variance and heavy tails imply that an unusually good individual opportunity is much better than an individual typical one. However, outliers – data points with much higher values than anything you've seen so far – are more common and more extreme for heavy-tailed distributions. We highlight two ways how this matters.[33]

First, **the *sum* of large samples from a heavy-tailed distribution will depend disproportionately on the contribution of outliers** – they account for a disproportionate share of the total.[34] For some heavy-tailed distributions, you should even expect that sufficiently large sums will be due to just a *single* extremely large summand ('catastrophe principle'). This is not true of light-tailed distributions, no matter their variance. Clearly it could matter for community building whether or not the total impact of the EA community will largely be due to only very few people.[35]

---

[31] However, only heavy-tailed distributions can have *infinite* variance. Conversely, there are different common definitions of 'heavy-tailed', and some of them imply infinite variance. For our purposes, however, it's useful if log-normal distributions count as heavy-tailed, and for any such definition the statement that heavy-tailed distributions can have arbitrarily small variance is true (since it's true for log-normal distributions).

[32] People also sometimes talk about distributions being *skewed.* This is yet another property conceptually distinct from both variance and heavy tails. Skewness is a conspicuous difference between *some* common heavy-tailed distributions – e.g., the log-normal and Pareto distributions – and the normal distribution, a paradigmatic example of a light-tailed distribution. However, heavy-tailed distributions need not be skewed: the Cauchy distribution is heavy-tailed but symmetric, i.e. not skewed (more generally this is true of any Lévy alpha-stable distribution with alpha < 2). Conversely, the exponential distribution is skewed but not heavy-tailed.

[33] A third difference is that 'heavy-tailed' is a property that's *scale-invariant,* while variance isn't. Thus the practical relevance of the heavy-tailed property is *internal* to the distribution, while variance matters only relative to a specified relationship between the distribution and the real world. For example, if I told you that the distribution of skyscraper heights had variance 100 this wouldn't mean anything to you without specifying the units – if the variance was 100 centimeters you'd think it was very low, if it was 100 kilometers you'd think it was very high. By contrast, saying that the distribution of skyscraper heights is heavy-tailed would tell you a lot without specifying units. [We don't know how skyscraper heights are in fact distributed.] There are ways to specify variance that avoid this problem, e.g. the ratio of the standard deviation to the mean ('coefficient of variation').

[34] This also means that the mean of a heavy-tailed distribution is much larger than its median. However, mean and median coming apart is not sufficient for heavy-tailedness, as shown e.g. by the exponential distribution (which has a larger mean than median but is not heavy-tailed).

[35] For instance, if impact across people is heavy-tailed, then 80K's metric for plan changes needs to be designed in such a way that it can be dominated by outliers.

Second (but relatedly), **for heavy-tailed distributions the sample variance and sample mean will severely understate the true mean and true variance** – even for very large samples. For heavy-tailed phenomena, naive extrapolation can thus be disastrous.[36]

To give a prominent and EA-relevant example, albeit one outside our focus: if battle deaths from war were heavy-tailed, we'd need to be very cautious when using historic casualty data to predict how deadly wars this century might be.[37] Another example, this time within our focus: when evaluating their recruitment efforts, local EA groups would like to know what they can and cannot infer from past trends.

This difference in what we can infer from past experience is intuitive for properties we are familiar with. For example, imagine you're in a room with perhaps a few dozen other people. Consider on one hand their *height* (light-tailed), and on the other hand their *wealth* (heavy-tailed). Additional people enter the room, one after the other. What happens to the height of the tallest person in the room over time, and how does this differ from the wealth of the wealthiest person?

At some point, a person that just entered will be taller than everyone else in the room. However, you'd be very surprised if the height difference between the new person and the previously tallest person was much larger than the height difference between the two previously tallest people.

E.g. if previously the two tallest people were 1.75m and 1.80m (which means there probably aren't that many people with you), you'd be quite surprised if the first person taller than that is 2m: it's much more likely that someone, say, 1.83m tall enters the room first because such people are much more common than 2m tall people. If previously the two tallest people were 2.11m and 2.12m (which probably means that the total number in the room is already much larger), then you will expect a new tallest person to be just barely rather than several cm taller, etc.

For wealth, it would be just the other way around: as new 'wealthiest people' enter the room, their net worth will exceed the previously highest wealth by *increasing* margins. E.g., the first millionaire may well enter the room when previously no-one in the room was worth more than half a million, and the first billionaire may well enter the room before the first person worth more than half a billion.

As a consequence, a single new person – e.g. the first billionaire – may well have a massive impact on the average wealth in the room (the sample mean understates the true mean). This will hardly happen for height. Similarly, by the time the first billionaire enters, she may well have more wealth than all other people in the room combined (the sum is dominated by an outlier), while this is basically impossible for height.

---

[36] Of course, a sample from a heavy-tailed distribution *does* contain *some* information, including on the distribution's mean and variance. The point is that we can only exploit this information with more sophisticated statistical techniques, which is beyond the scope of this post.

[37] For discussion of what data on past wars tells us about future wars, see Pinker (2011), Cirillo & Taleb (2016), and Braumoeller (2019).

It's hard to empirically distinguish different heavy-tailed distributions from one another, e.g. log-normal vs. power law

Fundamental difficulties

It is often easy to see whether data is heavy-tailed or light-tailed. For example, if over a large range the data appears as an approximately straight line in a plot with two logarithmic axes (for example, a rank-frequency plot), then – at least over this range – the data is heavy-tailed.

We can also see approximately how heavy-tailed the data is *in the observed range*, e.g. by looking at the slope of the line in such a plot (the steeper the slope, the less heavy-tailed).

However, **it's hard to identify the particular type of heavy-tailed distribution from observations alone**. For example, it can be impossible to tell whether data was generated by a log-normal or a Pareto distribution (a continuous power law). There are also other contenders that are rarely even considered, e.g. the 'double Pareto' or 'double Pareto-lognormal' distributions proposed by Reed (2003) and Reed & Jorgensen (2004), or the stretched exponential/Weibull distribution (e.g. Malevergne, Pisarenko, & Sornette 2005).

The basic reason for this is simply that different types of heavy-tailed distributions can provide almost equally good fits to the observed data. For example, while a power law is the only distribution that in expectation will generate a straight line in a log-log plot, data from a log-normal distribution can also look very much like a straight line over a large range. Since your observations will be noisy anyway, and your sample might not be big enough to cover the range where a log-normal would visibly deviate from a power law, you cannot tell the power law apart from the log-normal simply by seeing an approximately straight line over a finite range in a log-log plot.

This is no problem *if all you want to do is to describe the data you've seen*. After all, by design, if different distributions provide good fits to the data, they all do well at describing that data. (Though there will be systematic differences in *where* the fit is better or worse, and sometimes you might care about this.)

However, **you should be very careful when extrapolating beyond the range of observed data**.[38] This is because different types of heavy-tailed distributions that fit the observed data about equally well will differ dramatically in what they predict *beyond* the range of observed data. For example, suppose you have observed 10,000 earthquakes and based on this ask yourself how severe a "1 in a million" earthquake would be; a prediction

[38] Cf. footnote 10 in Clauset et al. (2009, p. 680): "In cases where we are unable to distinguish between two hypothesized distributions one could claim that there is really no difference between them: if both are good fits to the data then it makes no difference which one we use. This may be true in some cases but it is certainly not true in general. In particular, **if we wish to extrapolate a fitted distribution far into its tail, to predict, for example, the frequencies of large but rare events like major earthquakes or meteor impacts, then conclusions based on different fitted forms can differ enormously** even if the forms are indistinguishable in the domain covered by the actual data. Thus the ability to say whether the data clearly favor one hypothesis over another can have substantial practical consequences." (emphasis ours)

based on a power law would then predict a much more severe earthquake than one based on a log-normal that fits the observed data about equally well.

Similarly, **because there are only about 1,000 to 10,000 EAs, we probably couldn't say very much about the performance or impact of a "1 in a million"-EA based *just* on observing the performance or impact of existing EAs**, even if we could measure those with perfect reliability.

For more detail on this problem, **I recommend the paper *Power-law distributions in empirical data* by Clauset, Shalizi, & Newman (2009)**.[39]

They look at 24 data sets based on which previous papers claimed to have identified a power law. They *rule out* power laws in 7 cases. For the remaining 17, in all but one case there is another heavy-tailed distribution (e.g. stretched exponential or log normal) that fits the data about as well as a power law. (In 3 cases, even the exponential distribution – usually considered to be just on the edge between light-tailed and heavy-tailed distributions – could be a plausible fit.) In other words, **in only 1 out of 24 cases can we be confident that data was generated by a power law and not some other heavy-tailed distribution**.

Practical difficulties

Of course, if you have enough data from sufficiently reliable measurements, you will *sometimes* be able to rule out *some* heavy-tailed distributions.

However, even then **you'll have to use relatively sophisticated statistical techniques**. In particular, it is usually a *bad* idea to just fit a line to a log-log plot. Instead, use maximum likelihood estimation or more complex tools such as a "uniformly most powerful unbiased test".[40]

(In a polemical blog post, statistician Cosma Shalizi claims that if everyone used appropriate methods when working with heavy-tailed data, this would "lead to a real change in the literature" and that, e.g., "half or more each issue of *Physica A* would disappear".)

---

[39] The difficulty of distinguishing different heavy-tailed distributions based on observations has been acknowledged, either in general or for specific cases, in many other papers. For example, in a paper published in *Science* with the telling title *Critical Truths About Power Laws*, Stumpf & Porter (2012) conclude that "although power laws have been reported in areas ranging from finance and molecular biology to geophysics and the Internet, the data are typically insufficient and the mechanistic insights are almost always too limited for the identification of power-law behavior to be scientifically useful"; Mitzenmacher (2004, p. 227) in a paper on computer file sizes remarks that "Very similar basic generative models can lead to either power law or lognormal distributions, depending on seemingly trivial variations. There is, therefore, a reason why this argument as to whether power law or lognormal distributions are more accurate has arisen and repeated itself across a variety of fields." For instance, there are debates on the distribution of financial returns (e.g. Malevergne, Pisarenko, & Sornette 2005), city sizes (e.g. Malevergne, Pisarenko, & Sornette 2011) or citations (e.g. Golosovsky & Solomon 2012, Brzezinski 2015).

[40] Again see Clauset et al. (2009) for some basics on how to do this well. A uniformly most powerful unbiased (UMPU) test is used by Malevergne, Pisarenko, & Sornette (2011) to settle the 'log-normal vs. power law' debate on city sizes in favor of the latter, and they "advocate the UMPU test as a systematic tool to address similar controversies in the literature of many disciplines involving power laws, scaling, 'fat' or 'heavy' tails."

Even the easier problem of identifying the 'right' power law – i.e. ignoring the question whether a log-normal or other type of heavy-tailed distribution would fit the data just as well – can be tricky, in part because the inferred exponent can be very sensitive to the 'cutoff', i.e. the value above which the power law is supposed to apply.

As one cautionary tale, consider Michael Tauberg's 2018 *Medium* post on *Power Law in Popular Media* (from which we report data in our section on *Ex-post performance*). Tauberg fitted power laws to media data, using "existing R libraries that are designed for this sort of analysis". In fact, he analyzed each data set using two *different* R libraries, saying that this yields "similar results".

However, even small differences in the inferred power law exponent can have a significant impact on the tails.

For instance, for "weeks on the NYT bestseller list" R library *igraph* gives an exponent of 2.08, while library *poweRlaw* gives an exponent of 2.20 (perhaps because the former concluded that the power law holds above a cutoff of 5 weeks on the list, while the latter used 6 weeks as a cutoff). (These are the exponents of the probability *density* function, from which you have to subtract 1 to get the exponent of the *cumulative* distribution function.) This difference may look innocent at first glance; but in the distribution inferred by *igraph* the 'top 1-in-a-million' bestseller authors would account for 36% of all time on the bestseller list, while in the distribution inferred by *poweRlaw* their share would be 'only' 10%. Even for the top 1%, a frequency that clearly matters in practice, the difference is sizable: the predicted shares of the total are 71% and 46%, respectively.

Thus, if you wouldn't appreciate the import of power law exponents that differ by about 0.1, or if you wouldn't be able to adjudicate conflicting results spat out by different standard software, you might easily mislead yourself.

Worse, even if you're a maximally sophisticated statistician, your **conclusions will still be quite sensitive to a small number of outliers in your data**. For one, you might simply not be able to get enough data to observe, e.g., a "1-in-10,000" event. In addition, you'll often struggle with measurement error at the far end of the data you *can* get in principle – and this measurement error matters. For instance, to accurately determine the distribution of income you would need reliable information about top earners, which is hard to get (Anand & Segal claim to provide "the first estimates of global inequality that take into account data on the incomes of the top one percent within countries" – in a paper from 2014 [!]; see also 80,000 Hours).

I/O psychology papers on whether job performance is heavy-tailed don't update us much

In the psychology literature, there's a debate specifically on whether performance in typical jobs is normally distributed or heavy-tailed.

For example, in an influential meta-analysis, Hunter, Schmidt, & Judiesch (1990) found that performance in 'high-complexity' jobs (e.g. physician) and sales jobs is *not* normally

distributed. More recently, business scholar Herman Aguinis and collaborators have attacked the "long-held assumption in human resource management, organizational behavior, and industrial and organizational psychology that individual performance follows a Gaussian (normal) distribution" (O'Boyle & Aguinis 2012, p. 79; see also e.g. Aguinis & O'Boyle 2014, Aguinis et al. 2016).

Others have explicitly defended the claim that job performance – at least when measured appropriately – is usually normally distributed. For instance, Beck et al. (2014, p. 531) conclude that "large departures from normality are in many cases an artifact of measurement".

In fact, as we said in our section on *Ex-post performance*, it seems clear that performance data can be heavy-tailed or normal depending on the domain and performance measure used.

At first glance, we were unsure whether the debate in the literature adds much to this basic observation. We've encountered several qualitative claims on when to expect heavy-tailed vs. normal performance distributions, and while these claims often seemed reasonable to us, we weren't sure about the quantitative analysis that was supposed to support them.

We have neither comprehensively reviewed this debate nor tried to adjudicate it ourselves. A minor reason is that a lot of the debate is about a different question: the 'correct' definition of performance rather than the empirical distribution of agreed-upon quantities (see e.g. Aguinis et al., 2016, pp. 4f. on "behavior-based" vs. "results-based" definitions). We think that the appropriate operationalization of performance depends on the question one asks, and thus that we can simply use whatever data seems most relevant for a given question rather than quarrel about the best general definition.

More seriously, from glancing at the papers, we have tentative doubts about some of the statistical methods, and it would have taken more time to investigate whether these doubts are warranted. For example:

- *Within* the literature some papers (e.g. Micceri, 1989; O'Boyle & Aguinis, 2012) point out potentially severe flaws in others, including on distributions stipulated to be normal without good reason.
- Beck et al. (2014) only test normal against *exponential* distributions, which we find puzzling since the exponential distribution is *not* heavy-tailed, and the paper they respond to (O'Boyle & Aguinis 2012) claims that performance often has a Pareto distribution (rather than an exponential one). Beck et al. (2014, p. 539) explain that this is "because using the exponential distribution the @Risk program was able to converge for nearly all data sets, whereas the Paretian distributions failed to converge in several cases". They add that "in instances where more than one skewed distribution converged (e.g., exponential and Paretian), the results regarding the skewed distributions provided the same interpretation", but we don't find this sufficiently reassuring. If there are good theoretical reasons to use a particular type of distribution, then the mere fact that this causes issues with a particular type of software doesn't seem like a sufficient reason to change one's analytical approach –

at the very least we would want to see how such a "quick hack" may affect the validity of results, or an analysis of *why* the software didn't work.
    - They also make several statements that sound like they are merely checking whether data looks symmetric or skewed, which seems like the wrong question to ask since a symmetric distribution can still have a heavy tail, or conversely a skewed distribution could have a thin tail.
- O'Boyle & Aguinis (2012) and Beck et al. (2014, p. 539) "used the Decision Tools Suite program @Risk which is an add-on to Microsoft Excel". Our impression is that this is an uncommon choice of software among statistically literate communities, and that using Excel (or other spreadsheets) carries a high risk of ending up with unnoticed implementation errors (e.g. typos in which cells are being referenced in a formula).
- Aguinis and colleagues' (2016) essentially operationalize the question "does job characteristics *X* (e.g. complexity, autonomy) predict more heavy-tailed performance?" as "do values on an ordinal scale for *X* correlate with the Kolmogorov-Smirnov (K-S) statistic of the best power law we can fit to performance data". We have several questions about this approach.
    - Are these correlations meaningful at all, i.e. a good measure of whether or not job characteristics predict the extent to which performance follows a power law? We are neither sure whether it makes sense to look at a correlation with a statistical quantity such as the K-S statistic, nor whether the K-S statistic of the best fitted power law is a good measure for how heavy-tailed the data is.
    - Is it justified to simply fit a power law to all data, and ignore other heavy-tailed distributions? If the best fitted power law has a high K-S statistic, this certainly tells us that no power law is a good fit to the data – but does it tell us anything whether the data is instead, say, normally or log-normally distributed?
        - Indeed, some of their own graphs (ibid., Fig. 3AC) look conspicuously like log-normal data.
    - The *p*-values associated with their Kolmogorov-Smirnov tests vary wildly even for at first glance similar data (e.g. $p = 0.75$ for ecology publications and $p = 0.00$ for environmental science publications). However, they seem to ignore this in their further analysis. Is this justified?
    - A power law often only applies to a certain range of data, but their analysis seems to ignore this. Put differently, in their analysis a high K-S statistics could either indicate that the data follows a power law nowhere or that it does over some limited range.
- Hunter, Schmidt, & Judiesch (1990) don't seem to actually test whether the tails of their performance data are thin or heavy. Instead, they seem to simply assume that all distributions are normal by default. The reason why they reject normal distributions for high-complexity and sales jobs is *not* that they observed heavy tails but that their inferred normal distribution would have non-negligible probability mass on negative values. This seems to us to be at best a weak reason to reject a normal distribution (and if so, whether the actual distribution simply is a truncated normal distribution where values cannot fall below a certain minimum, or a different type of distribution altogether), but conversely we feel unsure whether the assumption of normality was well-founded in the first place.

Results from a meta-analysis of predictors of career success

These are Tables 1, 2, and 3 from Ng et al. (2005, pp. 384ff.).

TABLE 1

Meta-Analytic Results of the Predictors of Salary

Predictors	*N*	*k*	r_c	*SDc*	*Q*
Human capital					
Hours worked	15,428	22	.24*	.10	209.61*
Work centrality	9,101	17	.12*	.12	75.74
Job tenure	17,094	20	.07*	.14	361.66*
Organization tenure	39,562	39	.20*	.13	792.75*
Work experience	10,841	27	.27*	.13	260.05*
Willingness to transfer	3,156	6	.11*	.09	21.58*
International experience	4,869	4	.11*	.02	6.97
Education level	45,293	45	.29*	.14	1,126.93*
Career planning	522	2	.11*	.10	4.24
Political knowledge & skills	1,261	5	.29*	.05	4.60
Social capital	3,481	9	.17*	.14	67.56*
Average correlation			*.21*		
Organizational sponsorship					
Career sponsorship	3,406	10	.22*	.21	29.46*
Supervisor support	2,322	5	.05*	.13	24.14*
Training & skill development opportunities	9,670	7	.24*	.15	278.01*
Organizational resources	8,204	18	.07*	.13	159.66*
Average correlation			*.13*		
Socio-demographics					
Gender (*male* = 1, *female* = 0)	33,211	51	.18*	.11	519.21*
Race (*White* = 1, *non-White* = 0)	6,443	13	.11*	.12	115.10*
Marital status (*married* = 1, *unmarried* = 0)	23,303	29	.16*	.09	252.86*
Age	40,197	52	.26*	.16	1,249.90*
Average correlation			*.20*		
Stable individual differences					
Neuroticism	6,433	7	−.12*	.03	12.38
Conscientiousness	6,286	6	.07*	.10	55.95*
Extroversion	6,610	7	.10*	.05	27.00*
Agreeableness	6,286	6	−.10*	.01	2.23
Openness to experience	6,800	7	.04*	.04	9.94*
Proactivity	1,006	4	.11*	.13	11.69*
Locus of control	2,495	7	.06*	.11	21.91*
Cognitive ability	9,560	8	.27*	.07	69.49*
Average correlation			*.11*		

Notes. Average correlation is represented by the absolute value. N = cumulative sample size; k = number of studies cumulated; r_c = sample size weighted corrected correlation; and $Q = Q$ statistics.

*$p < .05$.

TABLE 2

Meta-Analytic Results of the Predictors of Promotion

Predictors	*N*	*k*	r_c	*SDc*	*Q*
Human capital					
Hours worked	12,077	10	.13*	.05	36.22*
Work centrality	5,258	5	.04*	.04	11.84*
Job tenure	11,393	10	−.02*	.07	62.96*
Organization tenure	17,725	17	.03*	.22	993.14*
Work experience	5,400	10	.06*	.26	402.62*
Willingness to transfer	3,982	5	.03*	.14	56.51*
International experience	4,768	3	.12*	.00	1.11
Education level	9,571	26	.05*	.08	95.72*
Political knowledge & skills	432	2	.07	.00	.04
Social capital	2,605	7	.15*	.06	10.67
Average correlation			*.06*		
Organizational sponsorship					
Career sponsorship	4,828	10	.12*	.08	33.53*
Supervisor support	1,235	6	.02	.00	2.68
Training & skill development opportunities	6,503	6	.23*	.21	391.39*
Organizational resources	18,780	14	.06*	.02	23.07*
Average correlation			*.10*		
Socio-demographics					
Gender (*male* = 1, *female* = 0)	19,545	29	.08*	.07	127.65*
Race (*White* = 1, *non-White* = 0)	11,148	11	.01	.03	24.84*
Marital status (*married* = 1, *unmarried* = 0)	26,708	16	.09*	.09	227.18*
Age	28,498	28	.02*	.21	1,334.28*
Average correlation			*.05*		
Stable individual differences					
Neuroticism	4,575	5	−.11*	.05	12.60*
Conscientiousness	4,428	4	.06*	.01	2.61
Extroversion	4,428	4	.18*	.06	8.82*
Agreeableness	4,428	4	−.05*	.00	.60
Openness to experience	4,942	5	.01	.02	7.23
Proactivity	676	2	.16*	.03	1.93
Locus of control	5,911	4	−.03	.03	6.44
Average correlation			*.08*		

Notes. Average correlation is represented by the absolute value. N = cumulative sample size; k = number of studies cumulated; r_c = sample size weighted corrected correlation; and Q = Q statistics.

*$p < .05$.

TABLE 3

Meta-Analytic Results of the Predictors of Career Satisfaction

Predictors	*N*	*k*	r_c	*SDc*	*Q*
Human capital					
Hours worked	9,236	17	.13*	.08	66.46*
Work centrality	14,944	19	.22*	.20	335.92*
Job tenure	6,491	9	−.02	.05	18.02*
Organization tenure	9,246	17	.02	.04	21.72
Work experience	7,318	16	.00	.10	68.93*
Willingness to transfer	1,060	4	−.06	.41	102.02*
International experience	5,068	4	.03	.03	13.62*
Education level	11,890	24	.03*	.07	65.38*
Career planning	2,367	7	.33*	.23	41.24*
Political knowledge & skills	6,112	2	.05*	.04	7.21
Social capital	3,051	8	.28*	.13	36.26*
Average correlation			*.10*		
Organizational sponsorship					
Career sponsorship	6,255	18	.44*	.21	166.75*
Supervisor support	1,653	6	.46*	.26	57.02*
Training & skill development opportunities	5,048	18	.38*	.16	82.84*
Organizational resources	7,096	15	−.02	.12	106.00*
Average correlation			*.31*		
Socio-demographics					
Gender (*male* = 1, *female* = 0)	10,246	22	.01	.08	65.58*
Race (*White* = 1, *non-White* = 0)	2,561	5	.03*	.11	27.92*
Marital status (*married* = 1, *unmarried* = 0)	6,468	14	.06*	.01	9.67
Age	11,913	26	.00	.09	114.62*
Average correlation			*.02*		
Stable individual differences					
Neuroticism	10,566	6	−.36*	.05	67.71*
Conscientiousness	10,566	6	.14*	.06	16.04*
Extroversion	10,566	6	.27*	.07	6.68
Agreeableness	4,634	5	.11*	.05	4.65
Openness to experience	10,962	7	.12*	.03	26.74*
Proactivity	1,072	3	.38*	.02	0.50
Locus of control	668	3	.47*	.29	22.57*
Average correlation			*.24*		

Notes. Average correlation is represented by the absolute value. N = cumulative sample size; k = number of studies cumulated; r_c = sample size weighted corrected correlation; and Q = Q statistics.

*$p < .05$.

How do our metrics of heavy-tailedness depend on the value at which the tail starts?

Suppose we're interested in the distribution of wealth among millionaires. This is the tail of the wealth distribution among *all* people. We might then ask: does that tail look like a Pareto distribution (power law), like the tail of a log-normal distribution, like an exponential distribution, or like the tail of a normal distribution? (And so on for other candidate distributions.) And what difference would this make for the top-shares and top-quantiles among millionaires – the metrics of heavy-tailedness we have reported in our tables?

For a Pareto distribution, these metrics depend on the 'shape' parameter *alpha* – the exponent appearing in the pdf, which controls how fast the density converges to zero. They

do not, however, depend on the ‘cutoff’ – the minimal value above which the Pareto distribution applies.

So if we knew that the tail of wealth is described by a Pareto distribution with *alpha* = 2, then we would know the wealth share of the top 1% (etc.) in that tail, no matter where the tail starts. If the distribution describes the wealth of millionaires, we know how wealthy the top 1% richest *millionaires* are compared to the total wealth owned by all *millionaires.* If the same distribution describes the wealth of *billionaires,* then the same number would describe the wealth of the top 1% richest *billionaires* compared to total *billionaire* wealth.

For an exponential distribution, our metrics of heavy-tailedness do *not* depend on its single ‘rate’ parameter *lambda*. Similarly, if we start with a normal distribution with mean 0, and then consider its right (positive) half as a probability distribution, by our metrics the heaviness of this ‘Gaussian tail’ does not depend on the variance *sigma^2* of the normal distribution we started with. Hence we have included data for exponential and the right half of a mean-0 normal distribution in our tables.

However, this observation is misleading: once we allow positive ‘cutoffs’ for the tail, the parameters *lambda* and *sigma^2 do* matter for heavy-tailedness. The apparent independence of parameters is an ‘artefact’ of the convention that exponential distributions are usually parametrized to ‘start’ at zero. But in this use case we’re actually looking at an exponential distribution starting at, for instance, one million (if we’re looking at the wealth of millionaires).

More precisely, the ‘benchmark’ values we report in our tables for exponential and right-half-of-normal distributions are good approximations if and only if *lambda* is sufficiently small – or *sigma^2* is sufficiently large, respectively. Here, “large” and “small” are in relation to the ‘cutoff point’ at which the tail starts, with the requirement becoming more demanding the larger the cutoff. So e.g. the ‘exponential distribution’ values from our tables (which are for any exponential distribution starting at 0) may be a good approximation for the exponential tail of ‘millionaire wealth’ for some fixed *lambda* (if it is ‘small enough’); but if we were using the *same lambda* to describe ‘billionaire wealth’, the values from our tables might no longer be a good approximation (namely if lambda is ‘small enough’ relative to one million but not ‘small enough’ relative to one billion).

Here is the precise technical result from which this follows.

Let X be a random variable and c be a constant real number; set $Y = X + c$. (Think e.g. of X having an exponential distribution starting at 0, c = 10^6; then Y might be the distribution describing the wealth of millionaires.) Let $0 < p < 1$ and set $q = 1 - p$ (representing probabilities). Denote the top-q-share of X with $t\_X(q)$ – so e.g. if q = 0.1 then $t\_X(q)$ would be the share of the top 10%.

A routine calculation then shows that

$$t\_Y(q) = t\_X(q)/(1 + c') + q/(1 + 1/c'),$$

where *c'* = *c*/**E**[*X*], i.e. the size of the translation 'in relation to' the expected value of the original distribution. We see that if *c'* is very close to zero, then *t_Y*(*q*) approximately equals *t_X*(*q*). As *c'* becomes larger, the first summand becomes smaller and the second one larger, and for *c'* going towards infinity the top-share *t_Y*(*q*) converges toward *q*.

An easier calculation shows that if *r_X* is some quantile of *X* as multiple of the median, then

$$r\_Y = r\_X/(1 + c'') + 1/(1 + 1/c''),$$

this time with *c"* = c / median(X). Thus translations of *X* have a very similar effect on this metric, this time with 1 rather than *q* as the limit for large translations.

(The above claims now follow since the expected value of an exponential distribution is 1/*lambda*, and the expected value of a right-half-of-normal distribution increases with the *sigma*^2 of the original normal distribution. Similar remarks apply for the median.)

Key concepts and terminology

- *Task* = type of deliberate activity or set of activities, described at a level of specificity such that instances of the activity are regularly carried out by different people and by the same person at different times.
    - E.g. driving a car, assembling a chair, writing physics papers.
- *Performance* = how well someone does at a task or set of tasks (e.g. all tasks relevant to a certain job, then called *job performance*). Usually operationalized with a specific *metric* or *proxy.*
    - Example performance metrics could be:
        - For driving a car: frequency of accidents per kilometer; average speed; satisfaction rating on a 1-10 scale by other people in the car.
        - For assembling a chair: required time; how much weight the assembled chair can endure without collapsing; amount of waste produced while assembling.
        - For writing physics papers: number of publications; citations to publications; ratings by academic peers.
    - We deliberately use performance in a very broad and loose sense. On our definition, "performance" can incorporate things one would usually call outcome or impact and that are beyond the performer's control. We also include both performance at a single instance of a task and aggregate performance over potentially long periods of time (e.g. a whole career).
- *Heavy-tailed* = having a heavier tail than an exponential distribution. Loosely this means that the tail of the probability density function approaches zero more slowly than the tail of an exponential distribution. Formally, it means that above some threshold *x* > *x*_0 the *conditional mean exceedance* (also known as *mean residual lifetime*) **E**[*X* - *x* | *X* > *x*] is a strictly increasing function of *x* (where **E** denotes expected value and *X* is a random variable with the distribution we're talking about).[41]
    - E.g. log-normal, Pareto distribution

[41] This definition follows Bryson (1974). There are different definitions of 'heavy tailed' in the literature, see e.g. here.

- *Light-tailed* = having a lighter tail than an exponential distribution. Loosely this means that the tail of the probability density function approaches zero faster than the tail of an exponential distribution. Formally, it means that above some threshold $x > x\_0$ the *conditional mean exceedance* $\mathbf{E}[X - x \mid X > x]$ is a strictly decreasing function of x.
    - E.g. normal distribution
- (Note that any exponential distribution has constant conditional mean exceedance. Thus on this definition, the exponential distribution is neither heavy-tailed nor light-tailed – it is right on the edge between these two properties.)

References

Acuna, D. E., Allesina, S. & Kording, K. P. 2012. Predicting scientific success. *Nature,* 489**,** 201-202.

Adams, R., Keloharju, M. & Knüpfer, S. 2018. Are CEOs born leaders? Lessons from traits of a million individuals. *Journal of Financial Economics,* 130**,** 392-408.

Agarwal, R. & Gaule, P. 2018. *Talent matters: Evidence from Mathematics*, IMF.

Aghion, P., Akcigit, U., Hyytinen, A. & Toivanen, O. 2017. *The Social Origins of Inventors,* Cambridge, MA, National Bureau of Economic Research.

Aguinis, H. & O'Boyle Jr, E. 2014. Star Performers in Twenty-First Century Organizations. *Personnel Psychology,* 67**,** 313-350.

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

Ammann, M., Horsch, P. & Oesch, D. 2016. Competing with Superstars. *Management Science,* 62**,** 2842-2858.

Anand, S. & Segal, P. 2014. The global distribution of income. *In:* Atkinson, A. & Bourguignon, F. (eds.) *Handbook of Income Distribution.* Amsterdam: Elsevier.

Atkinson, A. & Bourguignon, F. (eds.) 2014. *Handbook of Income Distribution, Volume 2A-2B,* Amsterdam: Elsevier.

Azoulay, P., Graff-Zivin, J., Uzzi, B., Wang, D., Williams, H., Evans, J. A., Jin, G. Z., Lu, S. F., Jones, B. F., Börner, K., Lakhani, K. R., Boudreau, K. J. & Guinan, E. C. 2018. Toward a more scientific science. *Science,* 361**,** 1194-1197.

Azoulay, P., Zivin, J. & Wang, J. 2010. Superstar Extinction. *The Quarterly Journal of Economics,* 125**,** 549-589.

Bakija, J., Cole, A. & Heim, B. 2012. *Jobs and Income Growth of Top Earners and the Causes of Changing Income Inequality: Evidence from U.S. Tax Return Data,* Williamstown, MA, Williams College.

Barrick, M. R. & Mount, M. K. 1991. The Big Five Personality Dimensions and Job Performance: A Meta-Analysis. *Personnel Psychology,* 44**,** 1–26.

Beck, J. W., Beatty, A. S. & Sackett, P. R. 2014. On the Distribution of Job Performance: The Role of Measurement Characteristics in Observed Departures from Normality. *Personnel Psychology,* 67**,** 531-566.

Biddle, J. & Hamermesh, D. 1994. Beauty and the Labor Market. *American Economic Review,* 84**,** 1174-1194.

Borghans, L., Golsteyn, B. H. H., Heckman, J. J. & Humphries, J. E. 2016. What grades and achievement tests measure. *Proceedings of the National Academy of Sciences,* 113**,** 13354.

Braumoeller, B. F. 2019. *Only the Dead: The Persistence of War in the Modern Age,* New York, Oxford University Press.

Brown, J. 2011. Quitters Never Win: The (Adverse) Incentive Effects of Competing with Superstars. *Journal of Political Economy,* 119**,** 982-1013.

Bryan, G. E. 1994. Not all programmers are created equal. Proceedings of 1994 IEEE Aerospace Applications Conference Proceedings, 5-12 February 1994 Vail, CO. IEEE, 55-62.

Bryson, M. C. 1974. Heavy-Tailed Distributions: Properties and Tests. *Technometrics,* 16**,** 61-68.

Brzezinski, M. 2015. Power laws in citation distributions: evidence from Scopus. *Scientometrics,* 103**,** 213-228.

Busse, T. V. & Mansfield, R. S. 1984. Selected Personality Traits and Achievement in Male Scientists. *Journal of Psychology,* 116**,** 117-131.

Cirillo, P. & Taleb, N. 2016. The Decline of Violent Conflicts: What Do the Data Really Say? *SSRN Electronic Journal***,** 1-26.

Clauset, A., Larremore, D. B. & Sinatra, R. 2017. Data-driven predictions in the science of science. *Science,* 355**,** 477-480.

Clauset, A., Shalizi, C. R. & Newman, M. E. J. 2009. Power-Law Distributions in Empirical Data. *SIAM Review,* 51**,** 661-703.

Cotton-Barratt, O., Daniel, M. & Sandberg, A. 2020. Defence in Depth Against Human Extinction: Prevention, Response, Resilience, and Why They All Matter. *Global Policy,* 11**,** 271-282.

Dalal, R. S. 2005. A Meta-Analysis of the Relationship Between Organizational Citizenship Behavior and Counterproductive Work Behavior. *Journal of Applied Psychology,* 90**,** 1241-1255.

Denrell, J. & Liu, C. 2012. Top performers are not the most impressive when extreme performance indicates unreliability. *Proceedings of the National Academy of Sciences of the United States of America,* 109**,** 9331-9336.

Dreher, G. F. & Ash, R. A. 1990. A comparative study of mentoring among men and women in managerial, professional, and technical positions. *Journal of Applied Psychology,* 75**,** 539-546.

Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., Petersen, A. M., Radicchi, F., Sinatra, R., Uzzi, B., Vespignani, A., Waltman, L., Wang, D. & Barabási, A.-L. 2018. Science of science. *Science,* 359**,** eaao0185.

Frank, S. A. 2009. The common patterns of nature. *Journal of Evolutionary Biology,* 22**,** 1563-1585.

Ganzach, Y. & Patel, P. 2018. Wages, mental abilities and assessments in large scale international surveys: Still not much more than g. *Intelligence,* 69**,** 1-7.

Gensowski, M., Heckman, J. & Savelyev, P. 2011. *The Effects of Education, Personality, and IQ on Earnings of High-Ability Men,* Bonn, Institute of Labor Economics.

Gensowski, M., 2018. Personality, IQ, and lifetime earnings. *Labour Economics,* 51, pp.170-183.

Golosovsky, M. & Solomon, S. 2012. Runaway events dominate the heavy tail of citation distributions. *The European Physical Journal Special Topics,* 205**,** 303-311.

Grobelny, J. 2018. Predictive Validity toward Job Performance of General and Specific Mental Abilities. A Validity Study across Different Occupational Groups. *Business and Management Studies,* 4.

Groysberg, B. 2011. *Chasing Stars: The Myth of Talent and the Portability of Performance,* Princeton, NJ, Princeton University Press.

Hamermesh, D., Meng, X. & Zhang, J. 2002. Dress for success--does primping pay? *Labour Economics,* 9**,** 361-373.

Hartigan, J. A. & Wigdor, A. K. (eds.) 1989. *Fairness in employment testing: Validity generalization, minority issues, and the General Aptitude Test Battery,* Washington, DC: National Academy Press.

Hunter, J. E. & Hunter, R. F. 1984. Validity and utility of alternative predictors of job performance. *Psychological Bulletin,* 96**,** 72-98.

Hunter, J. E., Schmidt, F. L. & Judiesch, M. K. 1990. Individual differences in output variability as a function of job complexity. *Journal of Applied Psychology,* 75**,** 28-42.

Jondeau, E. and Rockinger, M., 2003. Testing for differences in the tails of stock-market returns. Journal of Empirical Finance, 10(5), pp.559-581.

Judiesch, M. K. & Schmidt, F. L. 2000. Between-Worker Variability in Output Under Piece-Rate Versus Hourly Pay Systems. *Journal of Business and Psychology,* 14**,** 529–552.

Kaufman, S. B., Quilty, L. C., Grazioplene, R. G., Hirsh, J. B., Gray, J. R., Peterson, J. B. & Deyoung, C. G. 2016. Openness to Experience and Intellect Differentially Predict Creative Achievement in the Arts and Sciences. *Journal of Personality,* 84**,** 248-258.

Kautz, T., Heckman, J., Diris, R., Ter Weel, B. & Borghans, L. 2017. *Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success,* Cambridge, MA, National Bureau of Economic Research.

Kokotajlo, D. & Oprea, A. 2020. Counterproductive Altruism: The Other Heavy Tail. *Philosophical Perspectives,* 34**,** 134-163.

Kremer, M. 1993. The O-Ring Theory of Economic Development*. *The Quarterly Journal of Economics,* 108**,** 551-575.

Liu, L., Wang, Y., Sinatra, R., Giles, C. L., Song, C. & Wang, D. 2018. Hot streaks in artistic, cultural, and scientific careers. *Nature,* 559**,** 396–399.

Makel, M. C., Kell, H. J., Lubinski, D., Putallaz, M. & Benbow, C. P. 2016. When Lightning Strikes Twice: Profoundly Gifted, Profoundly Accomplished. *Psychological Science,* 27**,** 1004-1018.

Malevergne, Y., Pisarenko, V. & Sornette, D. 2005. Empirical distributions of stock returns: between the stretched exponential and the power law? *Quantitative Finance,* 5**,** 379-401.

Malevergne, Y., Pisarenko, V. & Sornette, D. 2011. Testing the Pareto against the lognormal distributions with the uniformly most powerful unbiased test applied to the distribution of cities. *Physical Review E,* 83**,** 036111.

Merton, R. K. 1968. The Matthew Effect in Science. *Science,* 159**,** 56-63.

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

Mitzenmacher, M. 2004. A Brief History of Generative Models for Power Law and Lognormal Distributions. *Internet Mathematics,* 1**,** 226-251.

Newman, M. E. J. 2005. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics,* 46**,** 323-351.

Ng, T. W. H., Eby, L. T., Sorensen, K. L. & Feldman, D. C. 2005. Predictors of objective and subjective career success: A Meta Analysis. *Personnel Psychology,* 58**,** 367-408.

O'Boyle Jr, E. & Aguinis, H. 2012. The Best and the Rest: Revisiting the Norm of Normality of Individual Performance. *Personnel Psychology,* 65**,** 79-119.

Park, G., Lubinski, D. and Benbow, C.P., 2007. Contrasting intellectual patterns predict creativity in the arts and sciences: Tracking intellectually precocious youth over 25 years. *Psychological Science*, *18*(11), pp.948-952.

Park, G., Lubinski, D. & Benbow, C. P. 2008. Ability Differences Among People Who Have Commensurate Degrees Matter for Scientific Creativity. *Psychological Science,* 19**,** 957–961.

Penner, O., Pan, R. K., Petersen, A. M. & Fortunato, S. 2013. The case for caution in predicting scientists' future impact. *Physics Today,* 66**,** 8–9.

Petersen, A. M., Wang, F. & Stanley, H. E. 2010. Methods for measuring the citations and productivity of scientists across time and discipline. *Physical Review E,* 81**,** 036114.

Pinker, S. 2011. *The Better Angels of our Nature,* New York, NY, Viking.

Price, D. J. D. S. 1965. Networks of Scientific Papers. *Science,* 149**,** 510-515.

Radicchi, F., Fortunato, S. & Castellano, C. 2008. Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences,* 105**,** 17268.

Redner, S. 1998. How popular is your paper? An empirical study of the citation distribution. *The European Physical Journal B - Condensed Matter and Complex Systems,* 4**,** 131-134.

Reed, W. J. 2003. The Pareto law of incomes—an explanation and an extension. *Physica A: Statistical Mechanics and its Applications,* 319**,** 469-486.

Reed, W. J. & Jorgensen, M. 2004. The Double Pareto-Lognormal Distribution—A New Parametric Model for Size Distributions. *Communications in Statistics - Theory and Methods,* 33**,** 1733-1753.

Richardson, K. & Norgate, S. H. 2015. Does IQ Really Predict Job Performance? *Applied Developmental Science,* 19**,** 153–169.

Robertson, K.F., Smeets, S., Lubinski, D. and Benbow, C.P., 2010. Beyond the threshold hypothesis: Even among the gifted and top math/science graduate students, cognitive abilities, vocational interests, and lifestyle preferences matter for career choice, performance, and persistence. *Current Directions in Psychological Science*, *19*(6), pp.346-351.

Roine, J. & Waldenström, D. 2014. Long-Run Trends in the Distribution of Income and Wealth. *In:* Atkinson, A. B. & Bourguignon, F. (eds.) *Handbook of Income Distribution.* Amsterdam: Elsevier.

Rosen, S. 1981. The Economics of Superstars. *American Economic Review,* 71**,** 845-858.

Rosenbaum, J. E. 1984. *Career Mobility in a Corporate Hierarchy,* New York, NY, Academic Press.

Sackett, P. R., Gruys, M. L. & Ellingson, J. E. 1998. Ability–personality interactions when predicting job performance. *Journal of Applied Psychology,* 83**,** 545-556.

Schmidt, F. L. & Hunter, J. E. 1992. Development of a Causal Model of Processes Determining Job Performance. *Current Directions in Psychological Science,* 1**,** 89–92.

Schmidt, F. L. & Hunter, J. E. 1998. The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin,* 124**,** 262-274.

Schmidt, F. L. & Hunter, J. E. 2004. General Mental Ability in the World of Work: Occupational Attainment and Job Performance. *Journal of Personality and Social Psychology,* 86**,** 162-173.

Shockley, W. 1957. On the Statistics of Individual Variations of Productivity in Research Laboratories. *Proceedings of the IRE,* 45**,** 279-290.

Sinatra, R., Wang, D., Deville, P., Song, C. & Barabási, A.-L. 2016. Quantifying the evolution of individual scientific impact. *Science,* 354**,** aaf5239.

Smith, J. & Winkler, R. 2006. The Optimizer's Curse: Skepticism and Postdecision Surprise in Decision Analysis. *Management Science,* 52**,** 311-322.

Spilerman, S. 1977. Careers, Labor Market Structure, and Socioeconomic Achievement. *American Journal of Sociology,* 83**,** 551-593.

Strenze, T. 2007. Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence,* 35**,** 401-426.

Stumpf, M. P. H. & Porter, M. A. 2012. Critical Truths About Power Laws. *Science,* 335**,** 665-666.

Tauberg, M. 2018. *Power Law in Popular Media* [Online]. Medium. Available: https://michaeltauberg.medium.com/power-law-in-popular-media-7d7efef3fb7c [Accessed 23 March 2021].

Tett, R. P., Jackson, D. N. & Rothstein, M. 1991. Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology,* 44**,** 703-742.

Turner, R. H. 1960. Sponsored and Contest Mobility and the School System. *American Sociological Review,* 25**,** 855-867.

Van Iddekinge, C. H., Aguinis, H., Mackey, J. D. & Deortentiis, P. S. 2017. A Meta-Analysis of the Interactive, Additive, and Relative Effects of Cognitive Ability and Motivation on Performance. *Journal of Management,* 44**,** 249-279.

Wai, J. 2014. Investigating the world's rich and powerful: Education, cognitive ability, and sex differences. *Intelligence,* 46**,** 54-72.

Wai, J. & Lincoln, D. 2016. Investigating the right tail of wealth: Education, cognitive ability, giving, network power, gender, ethnicity, leadership, and other characteristics. *Intelligence,* 54**,** 1-32.

Waldinger, F. 2012. Peer Effects in Science: Evidence from the Dismissal of Scientists in Nazi Germany. *The Review of Economic Studies,* 79**,** 838-861.

Wallace, M. L., Larivière, V. & Gingras, Y. 2009. Modeling a century of citation distributions. *Journal of Informetrics,* 3**,** 296-303.