# The Probability Conflation: A Reply 

Nassim Nicholas Taleb*, Ronald Richman ${ }^{\dagger}$, Marcos Carreira ${ }^{\ddagger}$, James Sharpe ${ }^{\S}$<br>*Tandon School of Engineering, New York University and Universa Investments ${ }^{\dagger}$ Old Mutual Insure and University of Witwatersrand ${ }^{\ddagger}$ Ecole Polytechnique ${ }^{\S}$ Sharpe Actuarial Limited Forthcoming, International Journal of Forecasting, 2023


#### Abstract

We respond to Tetlock et al. (2022) showing 1 ) how expert judgment fails to reflect tail risk, 2) the lack of compatibility between forecasting tournaments and tail risk assessment methods (such as extreme value theory).More importantly, we communicate a new result showing a greater gap between the properties of tail expectation and those of the corresponding probability.


## I. The Fat Tails Problem

Tetlock et al. (2022) [1], in their criticism of claims by a paper titled "On single point forecasts for fat-tailed variables" in this journal (Taleb et al., 2022 [2]) insist that discriminating between a binary probability and a continuous distribution is a false dichotomy, that binary probabilities derived from expert forecasting tournaments can provide information on tail risk, in addition to some claims about a collaboration with the first author and a "challenge".

We apologize for not answering most of their points as these are already amply covered in two papers in this very journal, which includes the one they are criticizing, [2] and the more formal [3]. Alas "probability" is a mathematical concept that does not easily accommodate verbal discussions and requires a formal treatment, which necessitates precise definitions.

At the gist of what we referred to as "the so-called masquerade problem" is the following conflation, we simplify from [3] using a continuous distribution for ease of exposition:

Let $K \in \mathbb{R}^{+}$be a threshold, $f($.$) a density function for a$ random variable $X \in \mathbb{R}^{+}, P_{K}=\mathbb{P}(X>K) \in[0,1]$ the probability of exceeding it, and $g(x): \mathbb{R}^{+} \rightarrow \mathbb{R}$, an impact function. Let $G_{K}$ be a partial expectation of $g($.$) for the$ function of realizations of $X$ above $K$ :

$$
G_{K}=\int_{K}^{\infty} g(x) f(x) \mathrm{d} x
$$

and for clarity lets write the survival function (that is, the complementary cumulative distribution function, CDF) at $K$ :

$$
P_{K}=\int_{K}^{\infty} f(x) d x
$$

The error comes from conflating the properties of $G_{K}$ which those of $P_{K}$, often associating $P_{K}$ with some constant representing the presumed impact associated with the threshold $K$.

[^0]
## Complementary CDF



Fig. 1. The inverse survival function (complementary quantile function) corresponding to a complementary $\mathrm{CDF}\left(P_{K}\right)$ is unbounded while the complementary CDF is bounded; it is extremely concave for tail probabilities, and compounds the estimation errors on $p$. Simply the transformation reverses the signs of the second derivative and compounds it. We show how errors on $P_{K}$ translate in larger and larger values for $K$, possibly infinite.

The intuition of the difference can be shown as follows: assuming $g(x)=x$, for $X$ a random variable with finite first moment, we have, focusing on the positive domain, generalizing the Tail Probability Expectation Formula,

$$
\begin{equation*}
G_{K}=\underbrace{K P_{K}}_{\text {Prob times impact at threshold }}+\underbrace{\int_{K}^{\infty} P_{x} d x}_{\text {additional term }} \tag{1}
\end{equation*}
$$

with a second term that can dominate the first, particularly under heavy tailed distributions.

As explained in the two referenced papers, $P_{K}$ as a random variable, being bounded between 0 and 1 , necessarily has thin tails, with finite mean, variance, and all moments. By the probability integral transform, its unconditional probability distribution is the Uniform $\mathcal{U}(0,1)$-and its conditional is usually treated, particularly in the Bayesian literature, as a Beta (a special case of which is the Uniform) -the Beta distribution can accommodate practically all shapes of double bounded unimodal distributions. A sum of such bets becomes rapidly Gaussian.

## Remark 1: Moments

$P_{K}$ corresponds to the zeroeth moment and is always thin-tailed. $G_{K}$ maps to higher moments, and up to the infinite moment (i.e. the extremum) dealt with in extreme value theory ${ }^{a}$
${ }^{a}$ Consider the expectation of the $p^{t h}$ moment $\mathbb{E}\left(x^{p}\right)$. $P_{K}$ corre-
sponds to $p=0$ and the expected maximum to $\lim _{p \rightarrow \infty} \mathbb{E}\left(x^{p}\right)$.
In general, risk management concerns extrema, another point of
divergence with Tetlock et al.

As discussed extensively in Taleb (1997) reflecting the author's experiences as a derivatives market-maker [4], one fails to "hedge" the other in practice (and, of course, in theory). For instance, a rise in skewness of the distribution will tend to increase one side $\left(G_{K}\right)$ while decreasing the other $\left(P_{K}\right)$ : simply, the number of realizations above $K$ drops, but their impact becomes larger ${ }^{1}$

And if one does not hedge the other, then being "good at predicting $P_{K}$ " provides no information on $G_{K}$.

## Remark 2: Probability Classes

If $P_{K}$ and $G_{K}$ are not in the same probability class, that is, while $P_{K}$ is always thin tailed, if $G_{K}$ is not thin-tailed, then one cannot be a practical proxy for the other.

The field of extreme value theory was designed to deal with the issue. Basically, "probability" is not a tangible object like a tomato; it is, mathematically, a kernel inside an integral (or a summation), inseparable from other integrands, and one should avoid drowning it with verbalism. This point applies no matter the probability interpretation (Bayesian, frequentist, propensional, or subjectivist). Furthermore, science is not about precise measurements of exceedance probability, but understanding properties in a comprehensive and useful ways. As explained in the paper criticized by Tetlock et al (2022), one does not handle pandemics via forecasting tournaments and reliance on champions for single point forecasts, but by getting full distributional properties, particularly the shape in

[^1]the tails. Decisions must be informed by the shape of the total distribution, and some understanding of the dynamics involved in generating such a distribution (multiplicative effects cause thick tails while additive ones tend to be benign stochastic outcomes).

Just as warning is not predicting, understanding distribution classes and tail properties is not forecasting. Furthermore, the language of "false positive", while useful in medicine and similar applications based on signal, in not useful in risk and insurance based on distributional considerations.

> Remark 3: Gambling
> It is worth noting that binary options on financial instruments (that is, the trading of $P_{K}$ or $1-P_{K}$ ) proved of little economic value and are not considered an investment in the U.S. and the European Union; they are banned by most corresponding regulators, as they are considered gambling devices. The European Securities and Markets Authority (ESMA) have disallowed retail dealing with binary options. These binaries were also traded at Lloyds, until banned by U.K. legislation with the 1909 Marine Insurance Act.

Note on "dichotomies": Tetlock et al (2022) seem to mix the "dichotomy" between binary and full payoffs with another distinction, that probability estimates help to flag tail risk. Our representation can accommodate both with the function $g($. which as mentioned earlier can reflect the infinite moment, i.e. the extremum.

## II. A New Result

At the Global Uncertainty Reading Group discussion around Tetlock et al (2022), on Dec 1, 2022, a new useful result emerged, which we find worth communicating ${ }^{2}$.

> Remark 4: Events are not defined
> A well known problem with heavy tails is that, at the core, in that class of distributions, "events" are not defined verbally: a "war" can have 200 or a million casualties, so it does not have a quantitative meaning. But setting a precise threshold no longer maps to a precise probability under an error rate, and, vice versa.

As illustrated in Fig. 1, The error in the evaluation of the probability $P_{K}$ can translate into an explosive value for the corresponding $K$, and the more fat tailed the distribution, the more explosive such corresponding value.

There is no space for a general proof, so we shall provide one for any distribution that ends (for large values) with Paretan tails, which is the standard case. Let us assume the probability $P_{K}$ follows a Beta Distribution $\beta(a, b)$ (both

[^2]parameters $>0$ and as we mentioned this fits the unconditional uniform with $a=b=1$ ).
$$
f_{P_{K}}(p)=\frac{p^{a-1}(1-p)^{b-1}}{B(a, b)}
$$
$0<p<1$, where ( $B .,$.$) is the standard Beta function.$
The mean and variance will be $M_{P_{K}}=\frac{a}{a+b}, V_{P_{K}}=$ $\frac{a b}{(a+b)^{2}(a+b+1)}$.

Now assume the underlying distribution for $X$ where $X>$ $K$, lies in the strong Pareto basin, meaning $P(X>K)=$ $L^{\alpha} K^{-\alpha}$, where $\alpha$ is the tail index an $L$ a scaling constant -this is general for large values of $X$ under all fat-tailed distributions.

The inverse complementary CDF (quantile function) can be expressed as: $K=L\left(P_{K}\right)^{-1 / \alpha}$ if $0 \leq P_{K} \leq 1$.

If probability taken as a random variable $P_{K}$, that is, $1-$ $C D F$ (the cumulative distribution function), follows a Beta distribution $\beta(a, b)$, then $K$, the inverse complementary CDF has for density:

$$
\begin{equation*}
f_{K}(k)=\frac{\alpha K^{-a \alpha-1} L^{a \alpha}\left(1-K^{-\alpha} L^{\alpha}\right)^{b-1}}{B(a, b)} \tag{2}
\end{equation*}
$$

with mean

$$
M_{K}=\frac{L \Gamma\left(a-\frac{1}{\alpha}\right) \Gamma(a+b)}{\Gamma(a) \Gamma\left(a+b-\frac{1}{\alpha}\right)}
$$

and variance

$$
V_{K}=\frac{L^{2} \Gamma(a+b)\left(\frac{\Gamma(a) \Gamma\left(a-\frac{2}{\alpha}\right)}{\Gamma\left(a+b-\frac{2}{\alpha}\right)}-\frac{\Gamma\left(a-\frac{1}{\alpha}\right)^{2} \Gamma(a+b)}{\Gamma\left(a+b-\frac{1}{\alpha}\right)^{2}}\right)}{\Gamma(a)^{2}}
$$

The proof is done via the standard Jacobian Method for the transformation of probability distributions.

As we can see the first moment exists only if $\alpha>\frac{1}{a}$ and $\alpha>\frac{1}{a+b}$; the second moment exists only if $\alpha>\frac{2}{a}$ and $\alpha>$ $\frac{2}{a+b}$; more generally the $n^{t h}$ moment exists only if $\alpha>\frac{n}{a}$ and $\alpha>\frac{n}{a+b}$.

## Remark 5: Error Propagation

While the error on the probability can be small and controlled, the error on the corresponding quantity under consideration can be infinite.

We note that the tail index $\alpha$ for pandemics (as addressed at length in the paper criticized by Tetlock et al. (2022)) is well below 1. The same applies to wars, which means that when it comes to conflicts, forecasts for tail events are not compatible with probability theory.

## III. A Rather Unscientific "Challenge"

"We challenge Taleb et al. (2022) to be equally transparent about the performance of their tail-risk hedging strategies -a controversial topic (Brown, 2020)."

We are surprised to see such a remark coming from professional evidence-based researchers: requesting the single track
record of a tail hedging program as a back-up for a claim about the mathematical inadequacy of using a binary forecast for, say, Covid 19 or similar events under fat tailed distributions. Said tail-hedging strategy consists in capturing the difference between idiosyncratically selected option prices in the market and subsequent market jumps. (Incidentally Professor Tetlock has appeared to conflate tail events -which take place in the tails of any distribution - and the fat-tailedness attributes of statistical distributions). And, on top, such request is made to the author of an entire book, Fooled by Randomness about the futility of such claims. To repeat the famous disparagement by the economist Jagdish Bhagwati of the claim by the speculator George Soros that he "falsified the random walk" [5], we find it highly unscientific to use a single track record to make any general claim - this bizarre demand on the part of professional researchers is no different from anecdotal claims ( $n=1$ ) used by medical charlatans. In addition, trading records are not like points in soccer games, particularly when they can hide tail risks 3

So we prefer the more robust challenge which, under these circumstances, becomes fair. We believe that academia is about search for knowledge and understanding the world, not a commercial enterprise. The same with societal risk management, which is about the public good and does not issue precise point forecasts (recall that Taleb et al.(2022), as its title indicates, is against single point prediction in some domains). So the burden is on forecasters to forecast. And we fail to get how a practical project with remarkable forecasting skills could work for government and not the private sector. The superforecasting project is just about such betting. So, as much as we would have preferred to voice the unspoken question, here we are obligated to put the old adage as "if they claim to be so good at forecasting, and their forecasts are actionable and related to reality, why aren't they so rich?" -in other words, why do they depend on taxpayer funds and, possibly, tax deductible (that is, charitable) contributions to finance such forecasts?

## IV. Conclusion: Some More Evidence Required

Finally, can Tetlock et al. prove that better estimates of $P_{K}$ can provide real benefits to decision makers?

In addition to the problems in the financial domain mentioned above, we completely fail to see the link in insurance -and in event risk in general. In our experience as risk and insurance practitioners, decision makers are usually not well equipped to deal with probabilistic information (compounding the difficulty in translating probabilistic information into practical effects $\sqrt{4}^{4}$. For instance, in a military context, if we refine

[^3]the estimate of a South China sea conflict from e.g. 15\% to $17 \%$, can Tetlock et al prove that this makes a difference in any practical situation?

Also, one is allowed to wonder why the superforecasting project is not applied to sports and election forecasts where 1) $P_{K}$ applies, 2) compatible with probability theory, 3) provides repeatable tests with overly abundant data and, centrally, 4) is "bankable" (that is, translatable into dollars and cents)?

We conclude with the following recommendation: in future work, it would be helpful if Tetlock et al. provided more rigorous backing of their claims about the link between $P_{K}$ and $G_{K}$. For, as it stands, we see neither theoretical nor practical benefits to that "superforecasting" enterprise.

## References

[1] P. E. Tetlock, Y. Lu, and B. A. Mellers, "False dichotomy alert: Improving subjective-probability estimates vs. raising awareness of systemic risk," International Journal of Forecasting, 2022.
[2] N. N. Taleb, Y. Bar-Yam, and P. Cirillo, "On single point forecasts for fat-tailed variables," International Journal of Forecasting, 2022.
[3] N. N. Taleb, "On the statistical differences between binary forecasts and real world payoffs," International Journal of Forecasting, 2021.
[4] ——, Dynamic Hedging: Managing Vanilla and Exotic Options. John Wiley \& Sons (Wiley Series in Financial Engineering), 1997.
[5] J. Bhagwati, "Wheel of fortune," The New Republic, pp. 40-41, October 5, 1987.
[6] N. N. Taleb, The Statistical Consequences of Fat Tails. STEM Academic Press https://arxiv.org/abs/2001.10488, 2020.

## Supplementary Material

The recording of the discussion session (webinar) is available at https://www.techincertoreadingclub.com


[^0]:    Corresponding author: N N Taleb, nnt1@nyu.edu. We thank Fergal McGovern and participants in the webinar Global Uncertainty Reading Group (formerly Global Technical Incerto Reading Group) special meeting on Dec 2, 2022. We also thank Pasquale Cirillo, Sara Lapsley, Brandon Yarckin and Mark Spitznagel.

[^1]:    ${ }^{1}$ Tetlock et al states that Taleb and Tetlock (2013) claims that the two methods are "complementary". Our understanding of the latter paper is that says the exact opposite: there is no such complementarity fat tailed variables, which is what this discussion about "masquerade" is about. This is the reason why the first author (Taleb) refused to be involved in the IARPA forecasting exercise.

[^2]:    ${ }^{2}$ This result can be useful in financial risk management, particulaly the mapping between "VaR" (Value at Risk, maps to $K$ for a set probability $P_{K}$ of losses above that threshold) and expected shortfall, "CVaR" (maps to $G_{K}$, that is, includes the impact of losses).

[^3]:    ${ }^{3}$ Since Tetlock et al (2022) is uncritically citing a web opinion article by Aaron Brown, we would like to debunk the claim in it that the performance record is not available: not being a retail product, it has been continuously available to what the Security and Exchange Commission (S.E.C.) defines as qualified investors, not Twitter activists, and Mr. Brown (who by his own disclosure had a severe conflict of interest) violated, willingly or unwillingly journalistic standards. For it turned out that Brown never did the fact checking, and never asked to see the audited returns. We also note that only dimensionless returns are to be compared.
    ${ }^{4}$ There is the other problem that payoffs are in in expectation space not in frequency space. For instance hedge funds with the best track records turned out to be the most vulnerable to tail events in 2008, see [6].

