Separating the Verbalistic from the Mathematical in Risk and Decisions

Nassim Nicholas Taleb
School of Engineering, New York University

***Abstract*—Payoffs (or risks, beliefs, predictions and exposures) are precisely defined as mathematical objects, in a legal-like codification into classes $\mathfrak{P}_1$ through $\mathfrak{P}_4$. Some paradoxes between the formal and the verbalistic are made clear, along with misspecifications and confusions in the decision literature, particularly with regards to fat-tailed processes and "irrationality" and "biases" of decision-makers (psychology tests only cover a subset of payoffs in $\mathfrak{P}_2$ that do not reflect natural tail exposures in $\mathfrak{P}_{>2}$).**

We also show how prediction markets (in $\mathfrak{P}_2$) do not hedge natural risks in $\mathfrak{P}_4$.

We start in dimension 1 and extend to larger dimensions.

TABLE I: Four Classes

Class $\mathfrak{P}_i$	Name	Function notation	Fourier Transform of $\phi(\Psi^+)$: $\widehat{\phi_1}(t)$	$\mathbb{E}(\Psi)^+$
$\mathfrak{P}_1$	Atomic	Ψ_1	1	$p(x)$
$\mathfrak{P}_2$	Binary	Ψ_2^+, Ψ_2^-	$(1-\pi_K)+e^{it}\pi_K$	π_K
$\mathfrak{P}_3$	Vanilla	Ψ_3^+, Ψ_3^-	$(1-\pi_K)$ $+\int_K^\infty e^{it}\,\mathrm{d}\mathrm{P}x$	$-K\,\pi_K+$ $\int_K^\infty x\,\mathrm{d}\mathrm{P}x$
$\mathfrak{P}_{4a}$	Comp.	Ψ_4	$\Sigma\Omega_i\widehat{\phi_i}(t)$,	$\Sigma\Omega_i\mathbb{E}(\Psi_i)$
$\mathfrak{P}_{4b}$	Gen. Sigm.		$\int\widehat{\phi_i}(t)\,\mathrm{d}\Omega$	$\int\mathbb{E}(\Psi_i)\,\mathrm{d}\Omega$

* * *

This is a mathematical-legal attempt at formally mapping payoffs and assessing their memberships in precisely defined classes. By legal we mean as expressed explicitly in a codified term sheet, legal contract, or formal legal code, which naturally converge to the mathematical definitions. The aims is showing the impossibility of verbalistic discussion of risk and exposures and the corresponding biases, and shows how the gap in stochastic properties between the verbalistic and mathematical increases under fat tails. Many biases in the psychology-decision science literature (such as the over-estimation of tail events, or the long shot bias in fat-tailed domains) are shown to simply result from misdefinitions or sloppy verbalism.

The discussion is organized as follows. We first introduce the intuition of the problems through simplified examples. We then maps payoffs and progressively show differences in stochastic properties. We then take a look at the decision theory literature.

I. What is an "event"?

The problem starts with the very definition of an event. Each definition corresponds to different tail properties.

Binary: A binary event is a discrete one that can only take one value. Binary statements, predictions and exposures are functions of ω in probability space (Ω, $\mathcal{F}$,$\mathbb{P}$), with true/false, yes/no types of answers expressed as events in a specific probability space. The outcome random variable $X(\omega)$ is either 0 (the event does not take place or the statement is false) or 1 (the event took place or the statement is true), that is the set {0,1} or (by affine scaling) the set $\{a_L, a_H\}$, with $a_L < a_H$ any two discrete and exhaustive values for the outcomes.

Example of binary: most scientific statements tested by "p-values", or most conversational nonquantitative "events" as whether a person will win the election, a single individual will die, a prince will become king or a team will win a contest.

Vanilla: statements, predictions and exposures, also known as natural random variables, correspond to situations in which the payoff is either 1) continuous or 2) discrete but can take infinite values, i.e. is at-least one tailed in its support.

In our classification below, for the purposes of unification, we will use the state variable approach in place of direct ω, with $x(\omega)$ the elementary security.

Other categories: We add another two categories, for a total of four, one below the binary and one above the vanilla, each one being an integration (or summation) of the previous one in the hierarchy.

The vanillas add a layer of complication: profits for companies or deaths due to terrorism or war can take many, many potential values. You can predict the company will be “profitable”, but the profit could be \$1 or \$10 billion.

The conflation binary-vanilla is a mis-specification often made in probability, seen in as fundamental texts as in J.M. Keynes' approach to probability [1]. Such a conflation is almost always present in discussions of "prediction markets" and similar aberrations; it affects some results in research. It is even made in places by De Finetti in the assessment of what makes a good "probability appraiser"[2].[1]

> The central point here is that decision-making is not about being a good *probability appraiser* –life is not about probability as a standalone concept but something more complex in which probability only enters as a

[1]The misuse comes from using the scoring rule of the following type:if a person gives a probability p for an event A, he is scored $(p-1)^2$ or p^2, according to whether A is subsequently found to be true or false. Consequences of A or the fact that there can be various versions of such event are, at best, an afterthought.

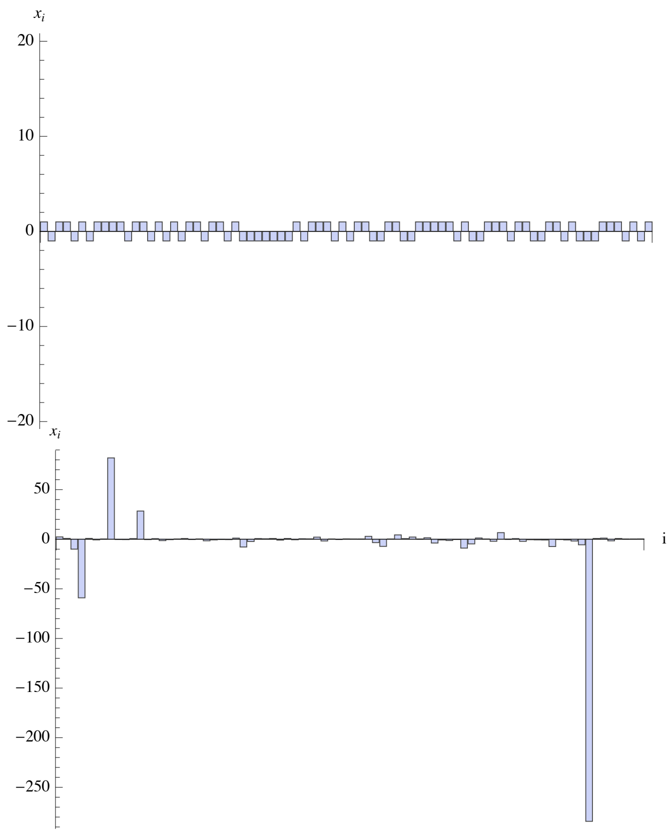

Fig. 1: Comparing payoff in classes $\mathfrak{P}_2$ to those in $\mathfrak{P}_3$ (top), or binaries to the vanilla. The vertical payoff shows x_i, $(x_1, x_2, ...)$ and the horizontal shows the index *i= (1,2,...)*, as *i* can be time, or any other form of classification. We assume in the first case payoffs of {-1,1}, and open-ended (or with a very remote and unknown bounds) in the second.

kernel, or integral transform.

The designation "vanilla" originates from definitions of financial contracts.[2]

Example 1 (Too much snow)**.** *The owner of a ski resort in the Lebanon, deploring lack of snow, deposited at a shrine of the Virgin Mary a $100 wishing for snow. Snow came, with such abundance, and avalanches, with people stuck in the cars, so the resort was forced to close, prompting the owner to quip "I should have only given $25". What the owner did is discover the notion of nonlinear exposure under tail events.*

Example 2 (Predicting the "Crisis" yet Blowing Up)**.** *The financial firm Morgan Stanley correctly predicted the onset of a subprime crisis, but they misdefined the event they called "crisis"; they had a binary hedge (for small drop) and ended up losing billions as the crisis ended up much deeper than predicted.*

As we will see, under fat tails, there is no such thing as a “typical event”, and nonlinearity widens the difference between verbalistic and precisely contractual definitions.

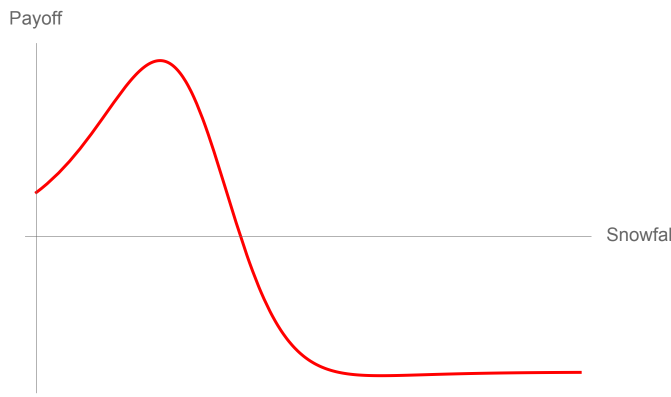

Fig. 2: The graph shows the payofd to the ski resort as a function of snowfall. So the discrete variable "snow" (vs "no snow") is not a random event for our purpose. Note that such a payoff is built via a convex/concave combinations of vanillas.

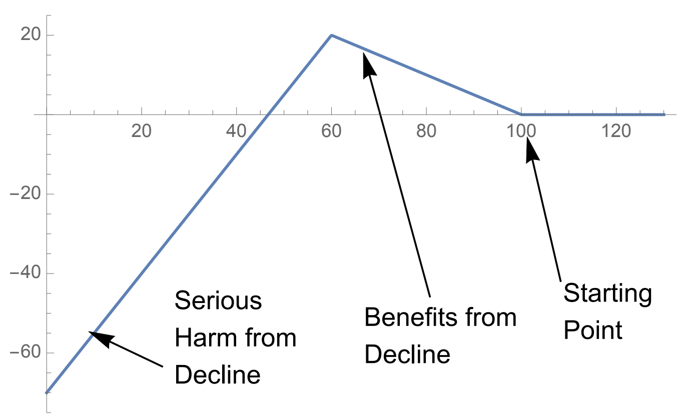

Fig. 3: A confusing story: mistaking a decline for an "event". This shows the Morgan Stanley error of defining a crisis as a binary event; they aimed at profiting from a decline and ended up structuring their exposure in a way to blow up from it. This exposure is called in derivatives traders jargon a "Christmas Tree", achieved in with $\mathfrak{P}_4$ through an addition of the following contracts $\Psi_3^-(K)_{1\leq i\leq 3}$ and quantitities q_1, q_2 and q_3 such that $q_1 > 0$, q_2 , $q_3 < 0$, and $q_1 < -q_2 < -q_3$, giving the toxic and highly nonlinear terminal payoff $\Psi_4 = q_1\Psi_3^-(K) + q_2\Psi_3^-(K - \Delta K) + q_3\Psi_3^-(K - k\Delta K)$, where $k > 1$. For convenience the figure shows K_2 triggered but not K_3 which kicks-in further in the tails.

II. Payoff Classes $\mathfrak{P}_1$ through $\mathfrak{P}_4$

Let $x \equiv x_T$ be a (non necessarily) Markovian continuous state variables observed at period T, $T \in \mathbb{R}^+$; x has support $\mathcal{D} = (\mathcal{D}^-, \mathcal{D}^+)$. The state variable is one-tailed or two-tailed, that is bounded on no more than one side, so either $\mathcal{D}^+ = \infty$ or $\mathcal{D}^- = -\infty$, or both.

The "primitive" state variable x_t is continuously observed between discrete periods $T-\Delta t$ and T. The payoff or exposure function is $\Psi \mathbb{1}_{t>\tau}$ where $\tau = \{\inf(t) : x_t \notin A , t \leq T\}$,

[2]The “vanilla” designation comes from option exposures that are open-ended as opposed to the binary ones that are called “exotic”; it is fitting outside option trading because the exposures they designate are naturally occurring continuous variables, as opposed to the binary that which tend to involve abrupt institution-mandated discontinuities.

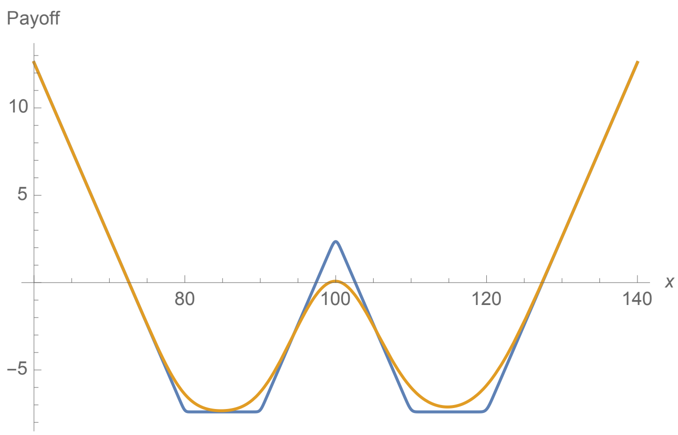

Fig. 4: Even more confusing: exposure to events –in class $\mathfrak{P}_4$ –that escape straightforward verbalistic descriptions. Option traders call this a "butterfly exposure" in the jargon.

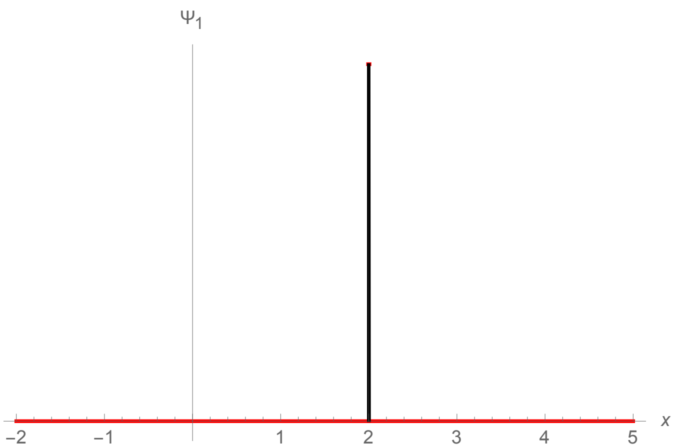

Fig. 5: Payoff Class $\mathfrak{P}_1$

a stopping-time conditional discretization of the continuously sampled time.[3]

The "payoff kernel" Ψ at time T is a member of the exhaustive and mutually exclusive following 4 classes. We write its probability distribution $\phi(\Psi)$ and characteristic function $\widehat{\phi}(t)$ (the distributions of the payoff under the law of state variable x between $T-\Delta t$ and T, Ψ itself taken as a random variable) at T, and $p(x)$ the probability law for x at T.

Note that the various layers are obtained by integration over the state variable x over segments of the domain $\mathcal{D}$:

$$\Psi_i = \int \Psi_{i-1}(x)\,\mathrm{d}x$$

A. Atomic Payoff $\mathfrak{P}_1$

Definition 1 (Class $\mathfrak{P}_1$, or Arrow-Debreu State Variable)**.** *$\Psi \equiv \Psi_1(x, K)$, which can be expressed as the Dirac Delta function:*

$$\Psi_1(x, K) = \delta(x - K)$$

[3]Without getting into details the stopping time does not have to be off the same primitive state variable x_t –even in dimention 1 –but can condition on any other state variable.

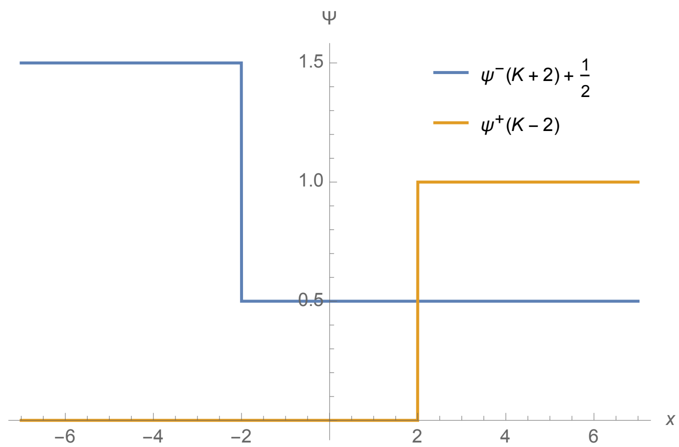

Fig. 6: Payoff Class $\mathfrak{P}_2$

where $\int_{K\in\mathcal{D}} \delta(x-K)\,\mathrm{d}x = 1$ and $\int_{K\notin\mathcal{D}} \delta(x-K)\,\mathrm{d}x = 0$ otherwise.

Remark 1 (Characteristic function invariance)**.** *The Characteristic function $\widehat{\phi}_1(t, K) = 1$ for all continuous probability distributions $p(x)$ of the primitive state variable x.*

Proof. $\int_{\mathfrak{D}} e^{i\,t\,\delta(x-K)} p(x)\mathrm{d}(x) = \int_{\mathfrak{D}} p(x)\mathrm{d}(x) = 1$ when K is in the domain of integration. □

Remark 2. *The expectation of Ψ_1 maps to a probability density at K for all continuous probability distributions.*

Proof. Consider that

$$\begin{aligned} i\frac{\partial}{\partial t}\widehat{\phi}_1(t,K) &= -i\frac{\partial}{\partial t}\int_{\mathfrak{D}} e^{(i\,t\,\delta(x-K))}p(x)\mathrm{d}x \\ &= \int_{\mathfrak{D}} e^{(i\,t\,\delta(x-K))}\delta(x-K)p(x)\mathrm{d}x \end{aligned} \tag{1}$$

Hence

$$\mathbb{E}(\Psi) = i\frac{\partial}{\partial t}\widehat{\phi}_1(t,K)|_{t=0} = p(K)$$

□

B. Binary Payoff Class $\mathfrak{P}_2$

Definition 2 ($\Psi \in \mathfrak{P}_2$, or Binary Payoffs)**.** *$\Psi \equiv \Psi_2(K)$ obtained by integration, so*

$$\Psi_2^{+}(K) = \int_{\mathcal{D}-}^{K} \Psi_1(x)\mathrm{d}x$$

which gives us, writing (for clarity) x for the state variable in the integrand and X for the observed one:

$$\Psi_2^{+}(X,K) = \begin{cases} 1 & \text{if } X \geq K; \\ 0 & \text{if } X < K. \end{cases}$$

and

$$\Psi_2^{-}(K) = \int_{K}^{\mathcal{D}^+} \Psi_1(x)\mathrm{d}x$$

giving us:

$$\Psi_2^{-}(X,K) = \begin{cases} 0 & \text{if } X > K; \\ 1 & \text{if } X \leq K. \end{cases}$$

which maps to the Heaviside θ function with known properties.

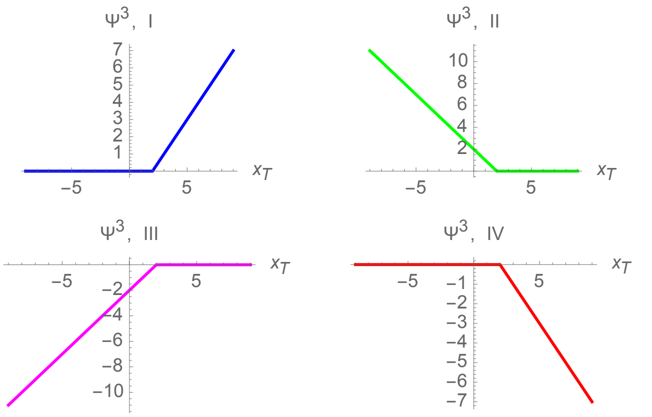

Fig. 7: Payoff Class $\mathfrak{P}_3$

Remark 3. *The class $\mathfrak{P}_2$ is closed under affine transformation $a_H \Psi + a_L$, for all combinations $\{a_H, a_L : a_H x + a_L \in \mathfrak{D}\}$. This is true for affine transformations of all payoff functions in $\Psi_{\geq 2}$, the unit of payoff becoming $a_H + a_L$ and the lower (upper) bound a_L (a_H).*

Proposition 1 (Binaries are Thin-Tailed)**.** *The probability distribution $\phi(\Psi_2)$, a "binary" payoff is a Bernouilli regardless of the underlying probability distribution over the state variable x.*

Proof. First consider that Ψ_2^+ can be written as $\Psi_2^+(x) = \frac{1}{2}(1 + \text{sgn}(x - K))$. Its characteristic function $\widehat{\phi}_2^+(t, K)$:

$$\widehat{\phi}_2^+(t, K) = \int_{\mathcal{D}} \mathbf{e}^{\frac{1}{2} i\, t\, (1+\text{sgn}(x-K))} p(x)\, \mathrm{d}x \tag{2}$$

$$= \int_{<K} p(x)\, \mathrm{d}x + \int_{\geq K} \mathbf{e}^{i\, t} p(x)\, \mathrm{d}x$$

So, with $\pi_K \equiv \mathbb{P}(X \geq K)$,

$$\widehat{\phi}_2^+(t, K) = (1 - \pi_K) + \mathbf{e}^{i\, t} \pi_K$$

which is the characteristic function of the Bernouilli distribution.

□

Note that we proved that Ψ_2 is subgaussian as defined in [3] regardless of $p(x)$ the probability distribution of the state variable, even if $p(x)$ has no moments.

C. Vanilla Payoff Class $\mathfrak{P}_3$, building blocks for regular exposures.

Definition 3 ($\Psi \in \mathfrak{P}_3$, or Vanilla Payoff)**.** $\Psi \equiv \Psi_3(X, K)$ *obtained by integration, so*

$$\Psi_3^+(X, K) = \int_{\mathcal{D}^-}^{X} \Psi_2(x - K)\mathrm{d}x$$

which gives us:

$$\Psi_3^+(X, K) = \begin{cases} X - K & \text{if } X \geq K; \\ 0 & \text{if } X < K. \end{cases}$$

and

$$\Psi_3^-(X, K) = \int_{X}^{\mathcal{D}^+} \Psi_2(x)\mathrm{d}x$$

giving us:

$$\Psi_3^-(X, K) = \begin{cases} K - X & \text{if } X \leq K; \\ 0 & \text{if } X > K. \end{cases}$$

Assume the support spans the real line. The characteristic function $\phi(t, K)$ can be expressed as:

$$\phi(t, K) = \int_{-\infty}^{\infty} p(X) e^{\frac{1}{2} it(X-K)(\text{sgn}(X-k)+1)}\, \mathrm{d}X$$

which becomes

$$\phi(t, K) = (1 - \pi_K) + e^{-itK} \int_K^{\infty} e^{itx} p(x)\mathrm{d}x \tag{3}$$

Proposition 2 (Impossibility)**.** *It is possible to build a composite/sigmoidal payoff using the limit of sums of vanillas with strikes K, and $K + \Delta K$, but not possible to obtain vanillas using binaries.*

Proof. The Fourier transform of the binary does not integrate into that of the vanilla as one need K struck at infinity. The sum requires open-ended payoffs on at least one side. □

For many distributions of the state variable the characteristic function allows explicit inversion (we can of course get numerical effects). Of some interest is the expectation that becomes:

$$\mathcal{E}(\Psi_3^+) = \int_K^{\infty} x\, p(x)\, dx - K\, \pi_K \tag{4}$$

which maps to common derivatives pricing such as the Bachelier approach[4] or it Lognormal generalizations popularized with [5].

As we can see Eq. 4 doesn't depend on the portion of the tail of the distribution below K. Of interest is the "stub" part of the pricing, which represents the difference between the vanilla and the binary of same strike K:

$$\Delta^+(K) \equiv E(\Psi_3^+ - K\Psi_2^+) = \int_K^{\infty} x\, p(x)\, \mathrm{d}x \tag{5}$$

The Δ contract has the convenience of sensitivity to fat tails (or other measures of uncertainty such as the scale of the distribution), as it extracts the "tail", segment of the distribution above (below) K.

The idea is to compare $\int_K^{\infty} x\, p(x)\, \mathrm{d}x$ and $\int_K^{\infty} p(x)\, \mathrm{d}x$ and see how they react in opposite directions to certain parameters that control the fatness of tails.

Remark 4 (Symmetry/Skewness Problem)**.** *There exists a nondegenerate distribution p*(x) with $\mathbb{E}^{p^*}(X) = \mathbb{E}^{p}(X)$ and $\mathbb{E}^{p^*}(|X|^s) = \mathbb{E}^{p}(|X|^s)$ for $s \leq 2$ such that:*

$$\begin{aligned} &\textit{sgn}\left(\int_K^{\infty} x\, p^*(x)\, \mathrm{d}x - \int_K^{\infty} x\, p(x)\, \mathrm{d}x\right) \\ &= -\textit{sgn}\left(\int_K^{\infty} p^*(x)\, \mathrm{d}x - \int_K^{\infty} p(x)\, \mathrm{d}x\right) \end{aligned} \tag{6}$$

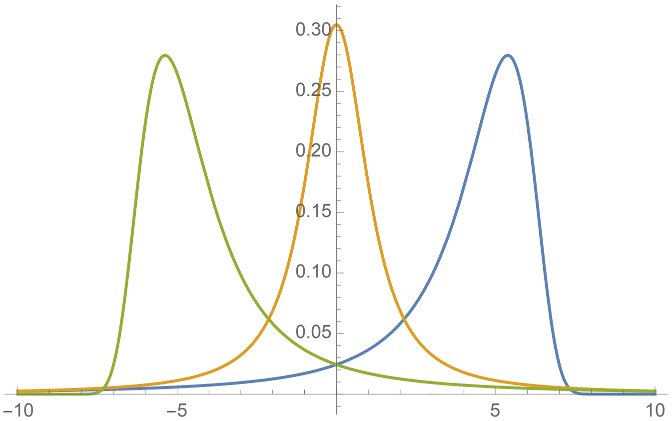

Fig. 8: Stable Distributions: remarkably the three have exactly the same mean and mean deviation, but different β symmetry parameter.

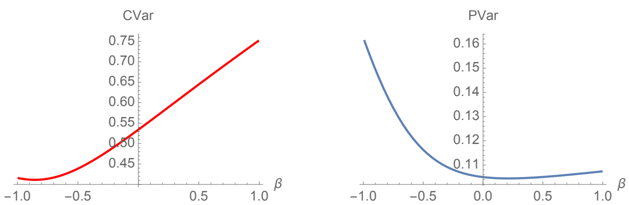

Fig. 9: Stable Distribution. As we decrease skewness, with all other properties invariant, the CVar rises and the PVar (probability associated with VaR) declines.

Proof. The sketch of a proof is as follows. Just consider two "mirror" asymmetric distributions, p_1 and p_2, with equal left and right side expectations.

With $\mathbb{P}^+_{p_1} \equiv \int_0^\infty p_1(x)\,\mathrm{d}x$ and $\mathbb{P}^-_{p_2} \equiv \int_{-\infty}^0 p_2(x)\,\mathrm{d}x$, we assumed $\mathbb{P}^+_{p_1} = \mathbb{P}^-_{p_2}$. This is sufficient to have all moments the exact the same (should these exist) and all other attributes in L^1 as well: the distributions are identical except for the "mirror" of positive and negative values for attributes that are allowed to have a negative sign.

We write $\mathbb{E}^+_{p_1} \equiv \int_0^\infty x\, p_1(x)\,\mathrm{d}x$ and $\mathbb{E}^+_{p_1} \equiv -\int_{-\infty}^0 x\, p_2(x)\,\mathrm{d}x$. Since $\mathbb{E}^+_{p_1} = -\mathbb{E}^-_{p_2}$ we can observe that all changes in the expectation of the positive (negative) side of p_2 around the origin need to be offset by a change in the cumulative probability over the same domain in opposite sign. □

The argument is easily explored with discrete distributions or mixing Gaussians, but we can make it more general with the use of continuous non-mixed ones: the α-Stable offers the remarkable property of allowing changes in the symmetry parameter while retaining others (mean, scale, mean deviation) invariant, unlike other distribution such as the Skew-Normal distribution that have a skew parameter that affects the mean.[4] In addition to the skewness, the stable can also thus show us precisely how we can fatten the tails while preserving other properties.

Example 3 (Mirror Stable distributions)**.** *Consider two mirror α-stable distributions as shown in Figure 9, $S_{\alpha,\beta,\mu,\sigma}$ with tail exponent $\alpha = \frac{3}{2}$ and $\beta = \pm 1$, centering at $\mu = 0$ to simplify;*

$$p_1(x) = -\sqrt[3]{2}\; e^{\frac{(\mu-x)^3}{27\sigma^3}} \frac{\left(\frac{\sqrt[3]{3}(\mu-x)Ai\left(\frac{(\mu-x)^2}{3\ 2^{2/3}\sqrt[3]{3}\sigma^2}\right)}{\sigma} + 3\sqrt[3]{2}Ai'\left(\frac{(\mu-x)^2}{3\ 2^{2/3}\sqrt[3]{3}\sigma^2}\right)\right)}{3\ 3^{2/3}\sigma}$$

$$p_2(x) = -\sqrt[3]{2}\; e^{\frac{(\mu-x)^3}{27\sigma^3}} \frac{\left(\frac{\sqrt[3]{3}(\mu-x)Ai\left(\frac{(\mu-x)^2}{3\ 2^{2/3}\sqrt[3]{3}\sigma^2}\right)}{\sigma} + 3\sqrt[3]{2}Ai'\left(\frac{(x-\mu)^2}{3\ 2^{2/3}\sqrt[3]{3}\sigma^2}\right)\right)}{3\ 3^{2/3}\sigma}$$

$$\mathbb{E}^+_{p_1} = \frac{\sqrt[3]{2}\sigma}{\Gamma\left(\frac{2}{3}\right)},\ \mathbb{E}^-_{p_1} = -\frac{\sqrt[3]{2}\sigma}{\Gamma\left(\frac{2}{3}\right)}$$

$$\mathbb{E}^+_{p_2} = \frac{\sqrt[3]{2}\sigma}{\Gamma\left(\frac{2}{3}\right)},\ \mathbb{E}^-_{p_2} = -\frac{\sqrt[3]{2}\sigma}{\Gamma\left(\frac{2}{3}\right)}$$

$$\mathbb{P}^+_{p_1} = \frac{1}{3},\ \mathbb{P}^+_{p_1} = \frac{2}{3}$$

$$\mathbb{P}^+_{p_2} = \frac{2}{3},\ \mathbb{P}^+_{p_1} = \frac{1}{3}$$

Moving the beta parameter which controls symmetry (and, only symmetry) to change the distribution have the effect of moving probabilities without altering expectations.

1) Stochastic Volatility Divergence: Let s be the scale of the distribution with density $p_s(x)$. Consider the ratio of densities;

$$\exists \lambda : \forall K > \lambda, 0 < \delta < 1, \frac{1}{2}\frac{(p_{s-\delta s}(K) + p_{s+\delta s}(K))}{p_s(K)} > 1$$

which is satisfied for continuous distributions with semi-concave densities.

> We will ferret out situations in which $\int_K^\infty x\, p(x)\,\mathrm{d}x$ (the "Cvar" or conditional value at risk) and $\int_K^\infty p(x)\,\mathrm{d}x$ (the Probability associated with "VaR" or value-at-risk) react to tail fattening situations in opposite manner.

[4]For instance, the Skew-Normal $N(\mu,\sigma,\beta;x)$, where $\beta \in \mathbb{R}$ controls the skewness, with PDF $\frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}\operatorname{erfc}\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)}{\sqrt{2\pi}\sigma}$, has mean $\frac{\sqrt{\frac{2}{\pi}}\beta\sigma}{\sqrt{\beta^2+1}} + \mu$ and standard deviation $\sqrt{1-\frac{2\beta^2}{\pi(\beta^2+1)}}\sigma$, meaning the manipulation of β leads to change in expectation and scale. The same applies to the mirrored Lognormal (where skewness and expectation depends on variance) and the Pareto Distribution (where the tail exponent controls the variance and the mean deviation if these exist.

D. *Composite/Sigmoidal Payoff Class* $\mathfrak{P}_4$

Definition 4 (P_4, or Composite Payoff)**.** *Pieced together sums of n payoffs weighted by Ω_j:*

$$\Psi_4 = \sum_{j=1}^{n} \Omega_j^+ \Phi_{i>1}^+(K_j) + \Omega_j^- \Phi_{i>1}^-(K_j)$$

This is the standard arithmetically decomposable composite payoff class, if we assume no conditions for stopping time – the ones encountered in regular exposures without utility taken into account, as a regular exposure can be expressed as the difference of two, more precisely $\Psi_2^+(K) - \Psi_2^-(K)\,, \forall K \in \mathcal{D}$.

Remark 5. *The class $\mathfrak{P}_4$ is closed under addition.*

III. Achieving Nonlinearity through $\mathfrak{P}_4$

IV. Main Errors in the Literature

The main errors are as follows.

- Binaries always belong to the class of thin-tailed distributions, because of boundedness, while the vanillas don't. This means the law of large numbers operates very rapidly there. Extreme events wane rapidly in importance: for instance, as we will see further down in the discussion of the Chernoff bound, the probability of a series of 1000 bets to diverge more than 50% from the expected average is less than 1 in 10^{18}, while the vanillas can experience wilder fluctuations with a high probability, particularly in fat-tailed domains. Comparing one to another can be a lunacy.
- The research literature documents a certain class of biases, such as "dread risk" or "long shot bias", which is the overestimation of some classes of rare events, but derived from binary variables, then falls for the severe mathematical mitake of extending the result to vanillas exposures. If ecological exposures in the real world tends to have vanillas, not binary properties, then much of these results are invalid.

Let us return to the point that the variations of vanillas are not bounded. The consequence is that the prediction of the vanilla is marred by Black Swan effects and need to be considered from such a viewpoint. For instance, a few prescient observers saw the potential for war among the Great Power of Europe in the early 20th century but virtually everyone missed the second dimension: that the war would wind up killing an unprecedented twenty million persons.

V. The Applicability of Some Psychological Biases

VI. Misfitness of Prediction Markets

A. *The Black Swan is Not About Probability But Payoff*

In short, the vanilla has another dimension, the payoff, in addition to the probability, while the binary is limited to the probability. Ignoring this additional dimension is equivalent to living in a 3-D world but discussing it as if it were 2-D, promoting the illusion to all who will listen that such an analysis captures all worth capturing.

TABLE II: True and False Biases in the Psychology Literature

Alleged Bias Derived in $\mathfrak{P}_2$	**Misspecified domain**	**Justified domain**
Dread Risk	Comparing Terrorism to fall from ladders	Comparing risks of driving vs flying
Overestimation of small probabilities	Open-ended payoffs in fat-tailed domains	Bounded bets in laboratory setting
Long shot bias	Convex financial payoffs	Lotteries

TABLE III: Adequate and inadequade decision domains

Application	**Questionable domain**	**Justified domain**
Prediction markets	Revolutions	Elections
Prediction markets	"Crashes" in Natural Markets (Finance)	Sports
Forecasting	Judging by frequency in venture capital and other winner take all domains;	Judging by frequency in finite bets

Now the Black Swan problem has been misunderstood. We are saying neither that there must be more volatility in our complexified world nor that there must be more outliers. Indeed, we may well have fewer such events but it has been shown that, under the mechanisms of “fat tails”, their “impact” gets larger and larger and more and more unpredictable.

Two points.

1) Binary predictions are more tractable than standard ones: First, binary predictions tend to work; we can learn to be pretty good at making them (at least on short timescales and with rapid accuracy feedback that teaches us how to distinguish signals from noise —all possible in forecasting tournaments as well as in electoral forecasting — see Silver, 2012). Further, these are mathematically tractable: your worst mistake is bounded, since probability is defined on the interval between 0 and 1. But the applications of these binaries tend to be restricted to manmade things, such as the world of games (the “ludic” domain).

It is important to note that, ironically, not only do Black Swan effects not impact the binaries, but they even make them more mathematically tractable, as will see further down.

2) Binary predictions are often taken as a substitute for standard ones: Second, most non-decision makers tend to confuse the binary and the vanilla. And well-intentioned efforts to improve performance in binary prediction tasks can have the unintended consequence of rendering us oblivious to catastrophic vanilla exposure.

Remark:*More technically, for a heavy tailed distribution (defined as part of the subexponential family), with at least one unbounded side to the random variable (one-tailedness), the variable prediction record over a long series will be of the same order as the best or worst prediction, whichever in largest in absolute value, while no single outcome can change the record of the binary.*

B. Chernoff Bound

The binary is subjected to very tight bounds. Let $(X_i)_{1<i\leq n}$ bea sequence independent Bernouilli trials taking values in the set $\{0,1\}$, with $\mathbb{P}(X=1])=p$ and $\mathbb{P}(X=0)=1-p$, Take the sum $S_n=\sum_{1<i\leq n} X_i$. with expectation $\mathbb{E}(S_n)= np=\mu$. Taking δ as a "distance from the mean", the Chernoff bounds gives:
For any $\delta>0$

$$\mathbb{P}(S\geq(1+\delta)\mu)\leq\left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu$$

and for $0<\delta\leq 1$

$$\mathbb{P}(S\geq(1+\delta)\mu)\leq 2e^{-\frac{\mu\delta^2}{3}}$$

Let us compute the probability of coin flips n of having 50% higher than the true mean, with p= $\frac{1}{2}$ and $\mu=\frac{n}{2}$: $\mathbb{P}\left(S\geq\left(\frac{3}{2}\right)\frac{n}{2}\right)\leq 2e^{-\frac{\mu\delta^2}{3}}=e^{-n/24}$
which for $n=1000$ happens every 1 in 1.24×10^{18}.

C. Fatter tails lower the probability of remote events (the binary) and raise the value of the vanilla.

The following intuitive exercise will illustrate what happens when one conserves the variance of a distribution, but "fattens the tails" by increasing the kurtosis. The probability of a certain type of intermediate and large deviation drops, but their impact increases. Counterintuitively, the possibility of staying within a band increases.

Let x be a standard Gaussian random variable with mean 0 (with no loss of generality) and standard deviation σ. Let $P_{>1\sigma}$ be the probability of exceeding one standard deviation. $P_{>1\sigma}$= $1-\frac{1}{2}\text{erfc}\left(-\frac{1}{\sqrt{2}}\right)$, where erfc is the complementary error function, so $P_{>1\sigma}$ = $P_{<1\sigma}$ $\simeq$15.86% and the probability of staying within the "stability tunnel" between $\pm$ 1 σ is 1− $P_{>1\sigma}$− $P_{<1\sigma}$ $\simeq$ 68.3%.

Let us fatten the tail in a variance-preserving manner, using the "barbell" standard method of linear combination of two Gaussians with two standard deviations separated by $\sigma\sqrt{1+a}$ and $\sigma\sqrt{1-a}$, a $\in$(0,1), where a is the "vvol" (which is variance preserving, technically of no big effect here, as a standard deviation-preserving spreading gives the same qualitative result). Such a method leads to the immediate raising of the standard Kurtosis by $\left(1+a^2\right)$ since $\frac{\mathbb{E}(x^4)}{\mathbb{E}(x^2)^2}=3\left(a^2+1\right)$, where $\mathbb{E}$ is the expectation operator.

$$\begin{aligned}P_{>1\sigma}&=P_{<1\sigma}\\&=1-\frac{1}{2}\text{erfc}\left(-\frac{1}{\sqrt{2}\sqrt{1-a}}\right)-\frac{1}{2}\text{erfc}\left(-\frac{1}{\sqrt{2}\sqrt{a+1}}\right)\end{aligned}\tag{7}$$

So then, for different values of a in Eq. 1 as we can see in Figure 2, the probability of staying inside 1 sigma rises, "rare" events become less frequent.

Note that this example was simplified for ease of argument. In fact the "tunnel" inside of which fat tailedness increases probabilities is between$-\sqrt{\frac{1}{2}\left(5-\sqrt{17}\right)}\sigma$ and $\sqrt{\frac{1}{2}\left(5-\sqrt{17}\right)}\sigma$ (even narrower than 1 σ in the example, as it numerically corresponds to the area between -.66 and .66), and the outer one is $\pm\sqrt{\frac{1}{2}\left(5+\sqrt{17}\right)}\sigma$, that is the area beyond $\pm$2.13 σ.

D. The law of large numbers works better with the binary than the variable

Getting a bit more technical, the law of large numbers works much faster for the binary than the variable (for which it may never work, see Taleb, 2013). The more convex the payoff, the more observations one needs to make a reliable inference. The idea is as follows, as can be illustrated by an extreme example of very tractable binary and intractable variable.

Let x_t be the realization of the random variable $X\in$ (-∞, ∞) at period t, which follows a Cauchy distribution with p.d.f. $f\left(x_t\right)\equiv\frac{1}{\pi\left(\left(x_0-1\right)^2+1\right)}$. Let us set $x_0=0$ to simplify and make the exposure symmetric around 0. The variable exposure maps to the variable x_t and has an expectation $\mathbb{E}\left(x_t\right)=\int_{-\infty}^{\infty}x_t f(x)dx$, which is undefined (i.e., will never converge to a fixed value). A bet at x_0 has a payoff mapped by as a Heaviside Theta Function $\theta_{>x_0}(x_t)$ paying 1 if $x_t>x_0$and 0 otherwise. The expectation of the payoff is simply $\mathbb{E}(\theta(x))=\int_{-\infty}^{\infty}\theta_{>x_0}(x)f(x)dx$= $\int_{x_0}^{\infty}f(x)dx$, which is simply $P(x>0)$. So long as a distribution exists, the binary exists and is Bernouilli distributed with probability of success and failure p and $1—p$ respectively .

The irony is that the payoff of a bet on a Cauchy, admittedly the worst possible distribution to work with since it lacks both mean and variance, can be mapped by a Bernouilli distribution, about the most tractable of the distributions. In this case the variable is the hardest thing to estimate, and the binary is the easiest thing to estimate.

Set $S_n=\frac{1}{n}\sum_{i=1}^n x_{t_i}$ the average payoff of a variety of variable bets x_{t_i}across periods t_i, and $S^\theta{}_n=\frac{1}{n}\sum_{i=1}^n\theta_{>x_0}\left(x_{t_i}\right)$. No matter how large n, $\lim_{n\to\infty}S^\theta{}_n$ has the same properties — the exact same probability distribution —as S_1. On the other hand $\lim_{n\to\infty}S^\theta{}_{n=p}$; further the presaymptotics of $S^\theta{}_n$ are tractable since it converges to $\frac{1}{2}$ rather quickly, and the standard deviations declines at speed $\sqrt{n}$, since $\sqrt{V\left(S^\theta{}_n\right)}=\sqrt{\frac{V\left(S^\theta{}_1\right)}{n}}$ =$\sqrt{\frac{(1-p)p}{n}}$ (given that the moment generating function for the average is $M(z)$ =$\left(pe^{z/n}-p+1\right)^n$).

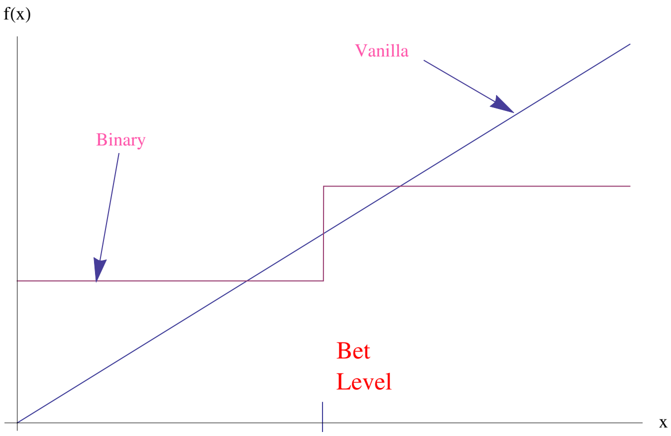

Fig. 10: The different classes of payoff f(x) seen in relation to an event x. (When considering options, the vanilla can start at a given bet level, so the payoff would be continuous on one side, not the other).

The binary has necessarily a thin-tailed distribution, regardless of domain

More, generally, for the class of heavy tailed distributions, in a long time series, the sum is of the same order as the maximum, which cannot be the case for the binary:

$$\lim_{X\to\infty} \frac{P\left(X > \sum_{i=1}^{n} x_{t_i}\right)}{P\left(X > \max\left(x_{t_i}\right)_{i\leq 2\leq n}\right)} = 1 \tag{8}$$

Compare this to the binary for which

$$\lim_{X\to\infty} P\left(X > \max\left(\theta(x_{t_i})\right)_{i\leq 2\leq n}\right) = 0 \tag{9}$$

The binary is necessarily a thin-tailed distribution, regardless of domain.

We can assert the following:

- The sum of binaries converges at a speed faster or equal to that of the variable.
- The sum of binaries is never dominated by a single event, while that of the variable can be.

How is the binary more robust to model error?

In the more general case, the expected payoff of the variable is expressed as $\int_A x dF(x)$ (the unconditional shortfall) while that of the binary= $\int_A dF(x)$, where A is the part of the support of interest for the exposure, typically A≡[K,∞), or (−∞,K]. Consider model error as perturbations in the parameters that determine the calculations of the probabilities. In the case of the variable, the perturbation's effect on the probability is multiplied by a larger value of x.

As an example, define a slighly more complicated variable than before, with option-like characteristics, $V(\alpha, K) \equiv \int_K^\infty x\ p_\alpha(x) dx$ and $B(\alpha, K) \equiv \int_K^\infty p_\alpha(x)\, dx$, where V is the expected payoff of variable, B is that of the binary, K is the "strike" equivalent for the bet level, and with x∈[1, ∞) let $p_\alpha(x)$ be the density of the Pareto distribution with minimum value 1 and tail exponent α, so $p_\alpha(x) \equiv \alpha x^{-\alpha-1}$.

Set the binary at .02, that is, a 2% probability of exceeding a certain number K, corresponds to an α=1.2275 and a K=24.2, so the binary is expressed as B(1.2, 24.2). Let us perturbate α, the tail exponent, to double the probability from .02 to .04. The result is $\frac{B(1.01,24.2)}{B(1.2,24.2)}$ = 2. The corresponding effect on the variable is $\frac{V(1.01,24.2)}{V(1.2,24.2)}$ = 37.4. In this case the variable was ∼18 times more sensitive than the binary.

—-

E. My of Tail Overestimation in Psychology of Tail Events

1) Long Shot Bias, Dread Risk, a Brief Survey of the Literature:

2) Anectodalism with Cass Sunstein's Critique of the Precautionary Principle:

References

[1] J. M. Keynes, "A treatise on probability," 1921.
[2] B. De Finetti, *Theory of Probability, volumes I and 2*. Wiley, 1977.
[3] J.-P. Kahane, *Some random series of functions*, 2nd ed., ser. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 1993.
[4] L. Bachelier, *Théorie de la spéculation*. Gauthier-Villars, 1900.
[5] F. Black and M. Scholes, "The pricing of options and corporate liabilities," *The journal of political economy*, pp. 637–654, 1973.