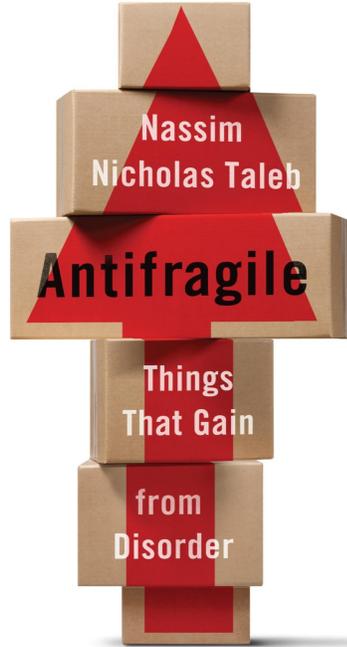


NEW YORK TIMES BESTSELLING AUTHOR OF  
THE BLACK SWAN



# بادشکننده

آنچه از بی‌نظمی بهره می‌برد



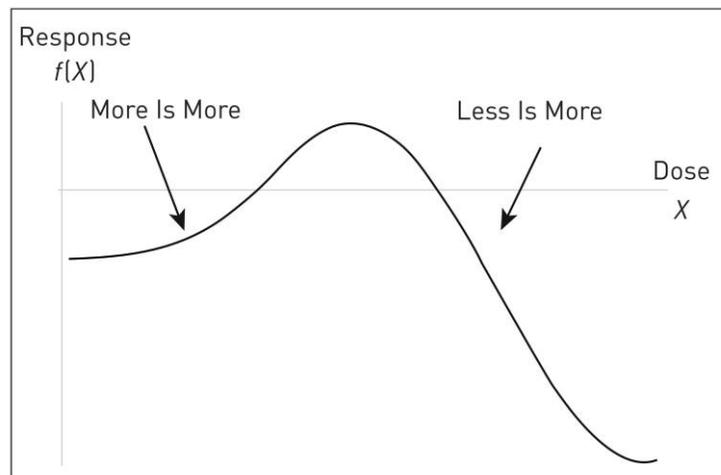
نسیم نیکلاس طالب  
ترجمه مینا صفری  
با ویراستاری تخصصی بهنام فلاح

Appendix I:  
**A GRAPHICAL TOUR OF THE BOOK**

---

For those nonliterary folks who like to see things in graphs, rather than words, and those only.

**NONLINEARITY AND LESS IS MORE (& PROCRUSTEAN BED)**



**FIGURE 19.** This graph explains both the nonlinear response and the “less is more” idea. As the dose increases beyond a certain point, benefits reverse. We saw that everything nonlinear is either convex, concave, or, as in this graph, mixed. Also shows how under nonlinearities, reductions fail: the Procrustean bed of words “good for you” or “bad” is severely distorting.

Also shows why tinkering-derived heuristics matter because they don’t take you into the danger zone—words and narratives do. Note how the “more is more” zone is convex, meaning accelerated initial benefits. (In Levantine Arabic, the zone beyond the saturation has a name: *كترتا محل قلنا* “more of it is like less of it.”)

Finally, it shows why competitive “sophistication” (rather, complication masked as sophistication) is harmful, as compared to the practitioner’s craving for optimal simplicity.

***Fragility Transfer Theorem:***

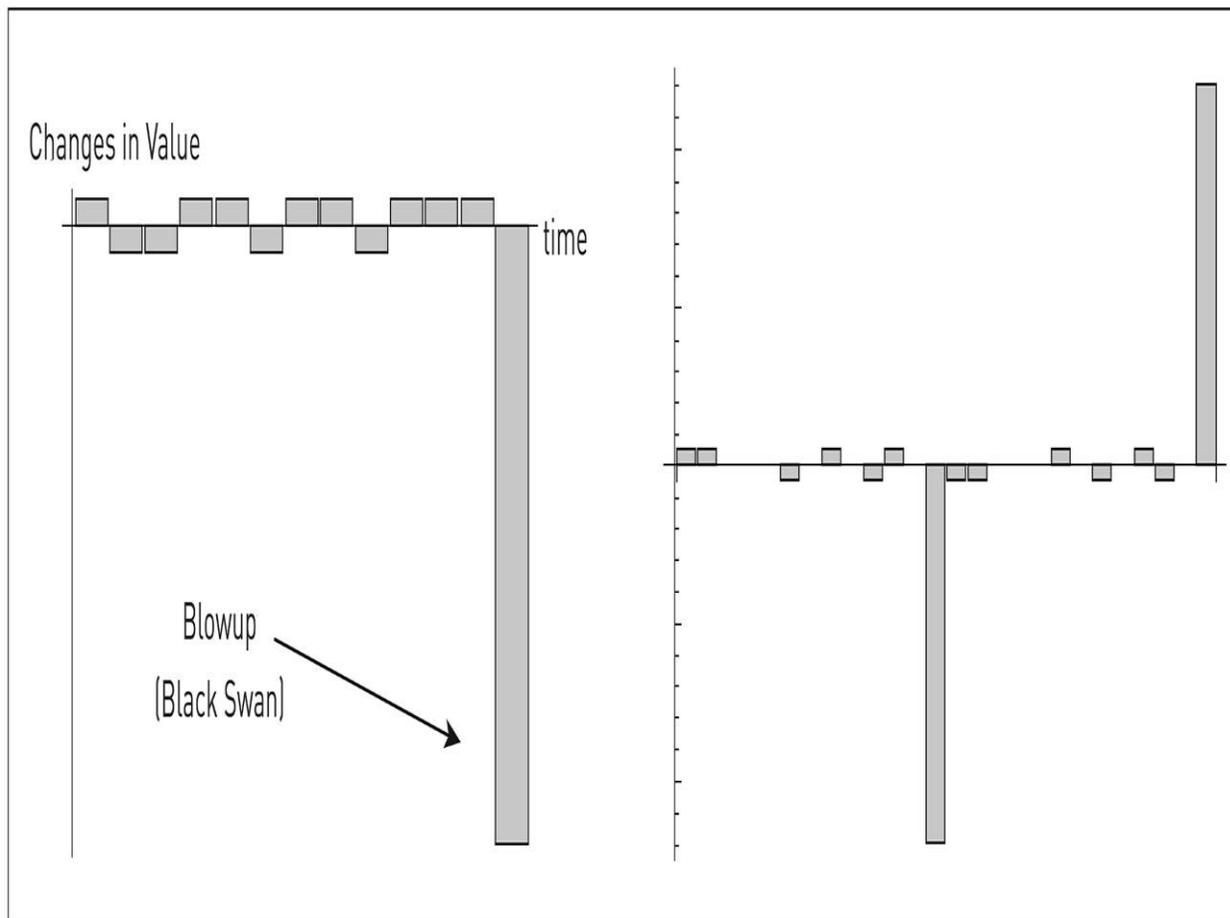
Note that by the Fragility Transfer Theorem,

CONVEX EXPOSURE [OVER SOME RANGE] ↔ LIKES VOLATILITY [UP TO SOME POINT]

**(volatility and other members of the disorder cluster), and**

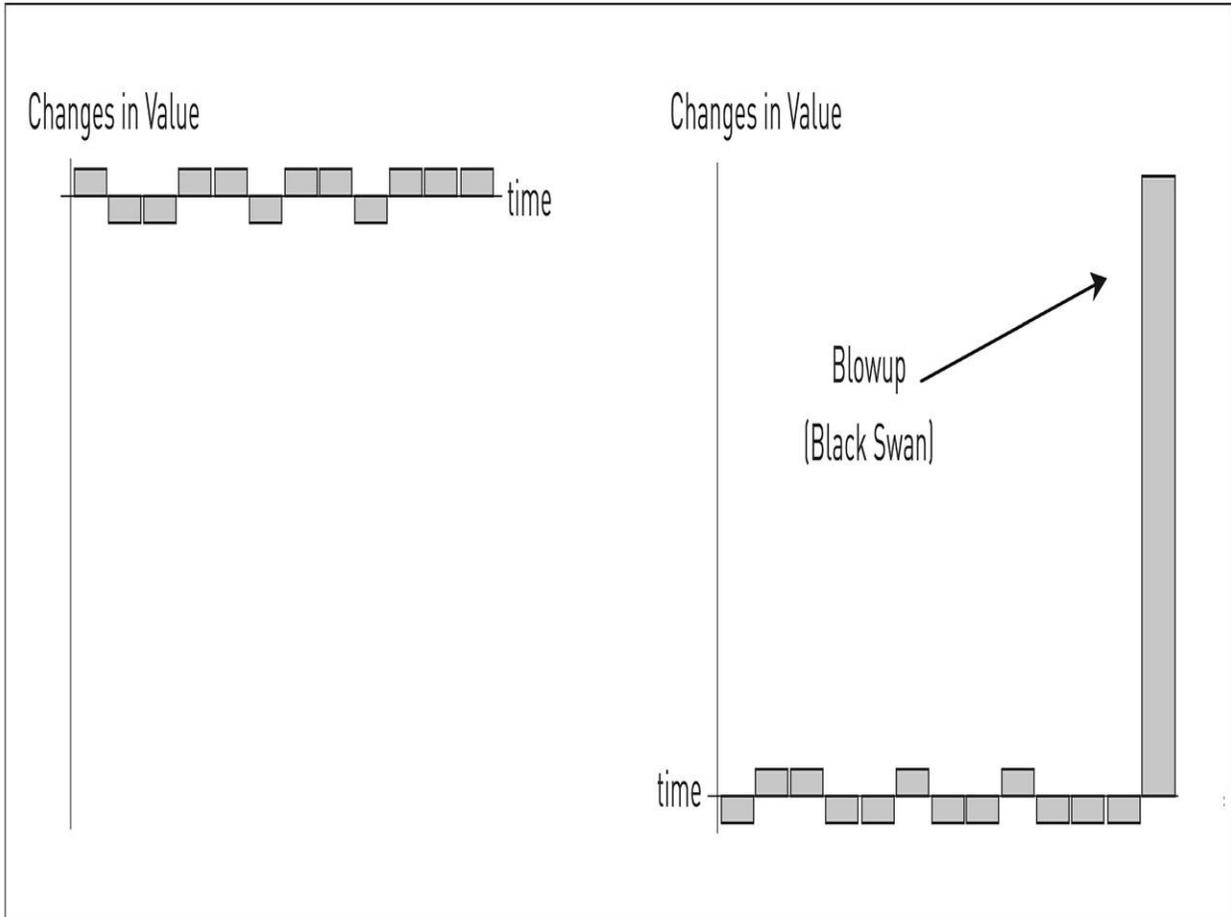
CONCAVE EXPOSURE ↔ DISLIKES VOLATILITY

### MAPPING OF FRAGILITIES



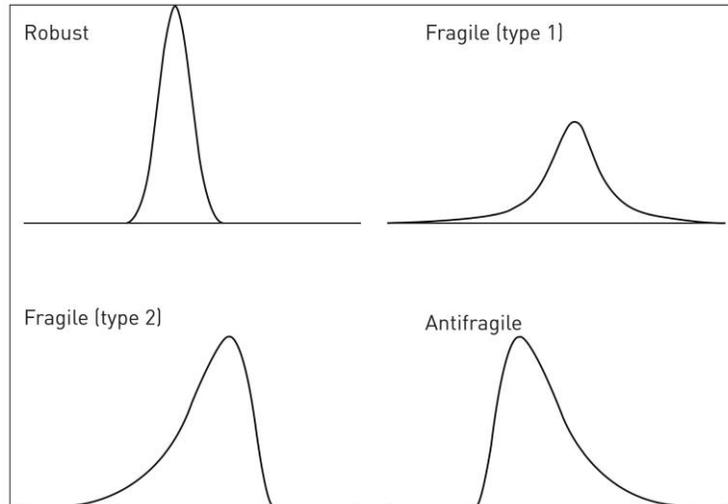
*In Time Series Space*

**FIGURE 20. Fragile variations through time, two types of fragilities.** A representative series. The horizontal axis shows time, the vertical one shows variations. This can apply to anything: a health indicator, changes in wealth, your happiness, etc. We can see small (or no) benefits and variations most of the time and occasional large adverse outcomes. Uncertainty can hit in a rather hard way. Notice that the loss can occur at any time and exceed the previous cumulative gains. Type 2 (left) and Type 1 (right) differ in that Type 2 does not experience large positive effects from uncertainty while Type 1 does.



**FIGURE 21. The Just Robust (but not antifragile) (left):** It experiences small or no variations through time. Never large ones. **The Antifragile system (right):** Uncertainty benefits a lot more than it hurts—the exact opposite of the first graph in [Figure 20](#).

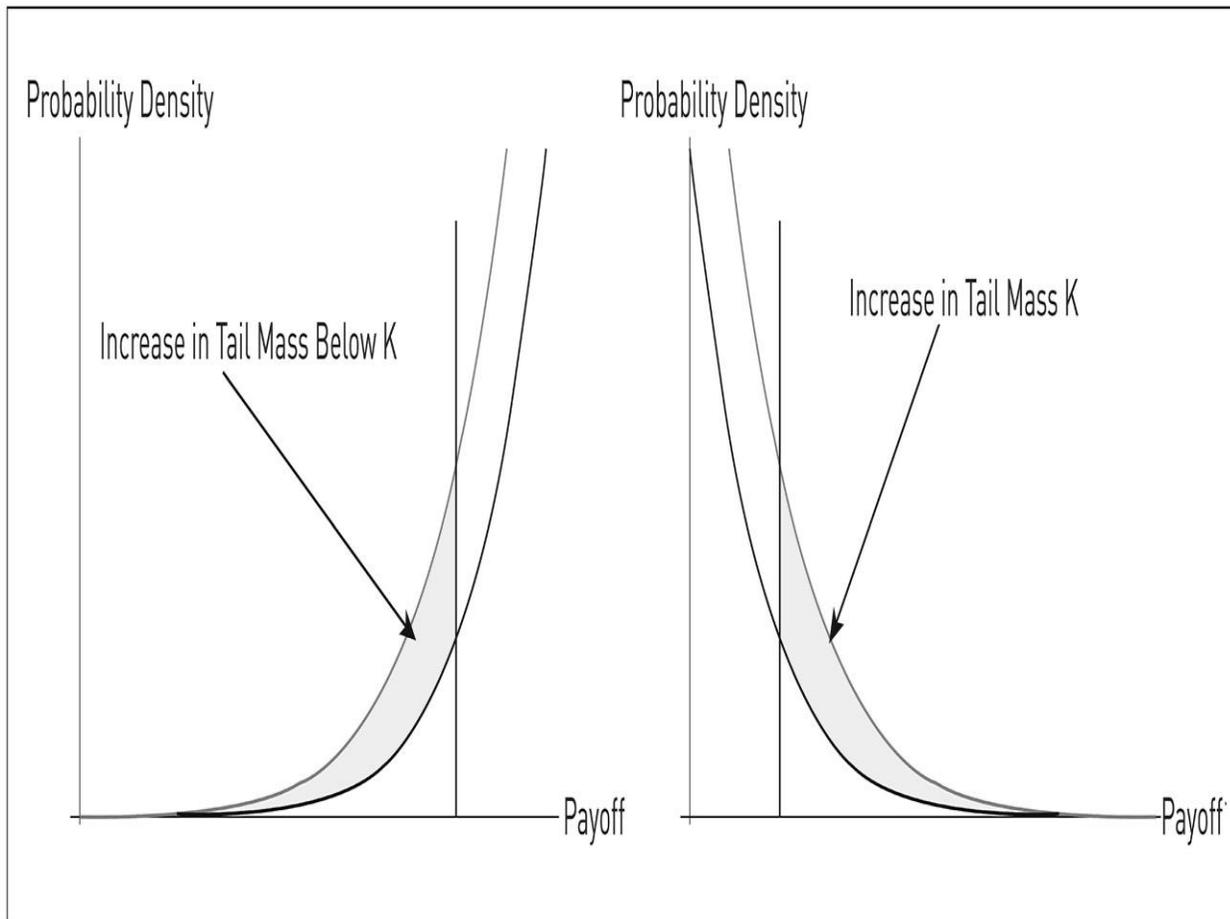
Seen in Probabilities



**FIGURE 22.** The horizontal axis represents outcomes, the vertical their probability (i.e., their frequency). **The Robust:** Small positive and negative outcomes. **The Fragile (Type 1, very rare):** Can deliver both large negative and large positive outcomes. Why is it rare? Symmetry is very, very rare empirically yet all statistical distributions tend to simplify by using it. **The Fragile (Type 2):** We see large improbable downside (often hidden and ignored), small upside. There is a possibility of a severe unfavorable outcome (left), much more than a hugely favorable one, as the left side is thicker than the right one. **The Antifragile:** Large upside, small downside. Large favorable outcomes are possible, large unfavorable ones less so (if not impossible). The right “tail,” for favorable outcomes, is larger than the left one.

<b>TABLE 9 • THE FOUR DIFFERENT CLASSES OF PAYOFFS</b>		
<i>Left Tail of the Distribution</i>	<i>Right Tail of the Distribution</i>	<i>Condition</i>
Thin	Thick	Antifragile
Thick	Thick	Fragile (Type 1) [Very Rare]
Thick	Thin	Fragile (Type 2)
Thin	Thin	Robust

Fragility has a left tail and, what is crucial, is therefore sensitive to perturbations of the left side of the probability distribution.



**FIGURE 23. Definition of Fragility** (left graph): Fragility is the shaded area, the increase in the mass in left tail below a certain level  $K$  of the target variable in response to any change in parameter of the source variable—mostly the “volatility” or something a bit more tuned. We subsume all these changes in  $s^-$ , about which later in the notes section (where I managed to hide equations).

For a **definition of antifragility** (right graph), which is not exactly symmetric, the same mirror image for right tail plus robustness in left tail. The parameter perturbed is  $s^+$ .

It is key that while we may not be able to specify the probability distribution with any precision, we can probe the response through heuristics thanks to the “transfer theorem” in Taleb and Douady (2012). In other words, we do not need to understand the future probability of events, but we can figure out the fragility to these events.

## BARBELL TRANSFORMATION IN TIME SERIES

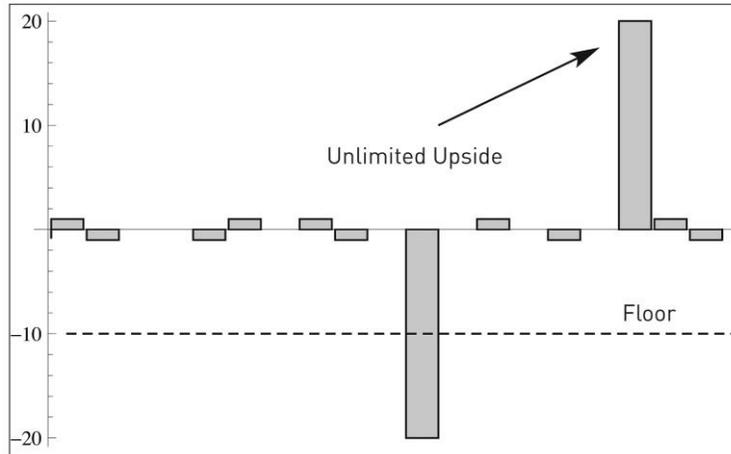
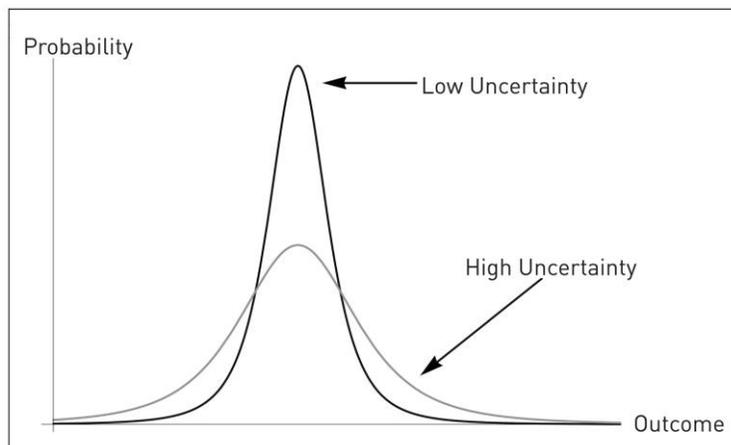


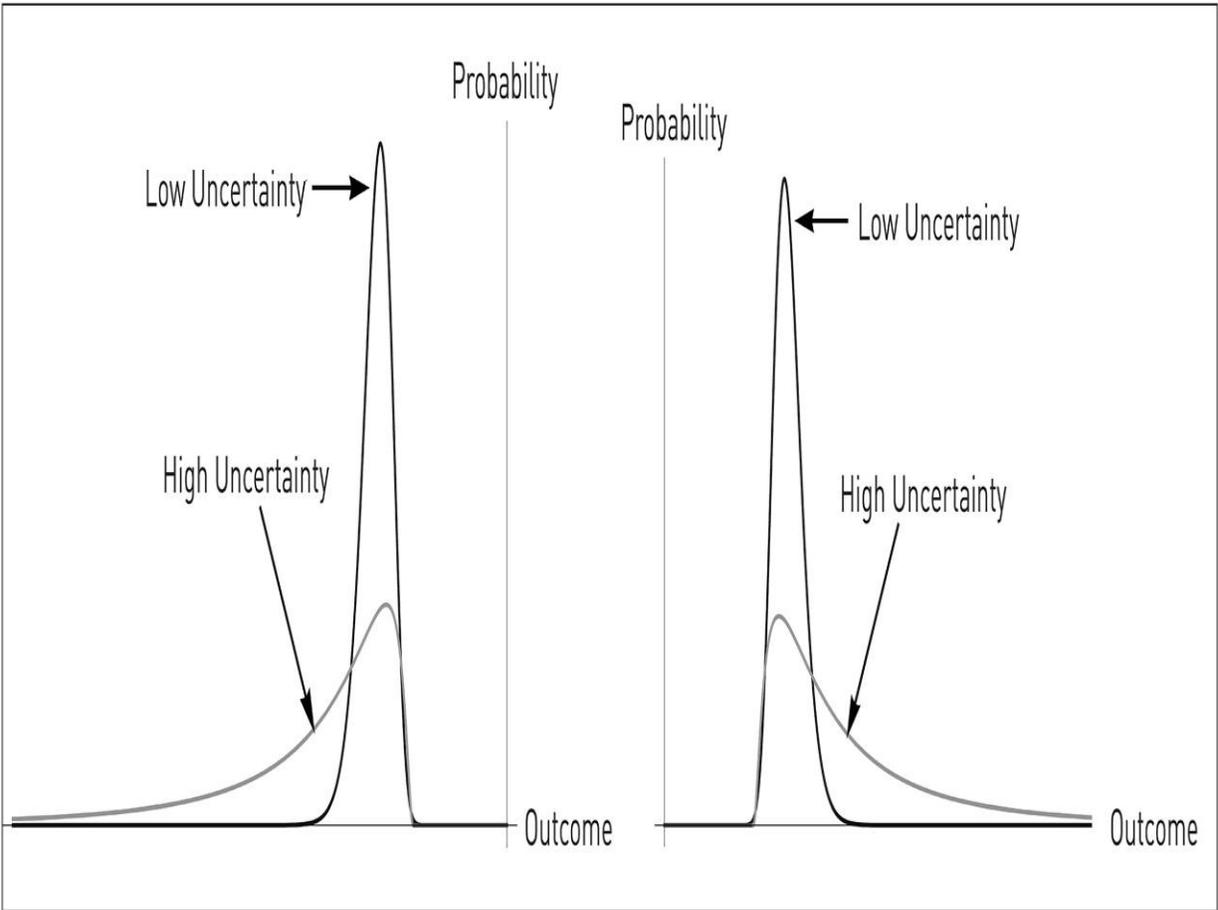
FIGURE 24. Barbell seen in time series space. Flooring payoffs while keeping upside.

## BARBELLS (CONVEX TRANSFORMATIONS) AND THEIR PROPERTIES IN PROBABILITY SPACE



A graphical expression of the barbell idea.

FIGURE 25. Case 1, **the Symmetric Case**. Injecting uncertainty into the system makes us move from one bell shape—the first, with narrow possible spate of outcomes—to the second, a lower peak but more spread out. So it causes an increase of both positive and negative surprises, both positive and negative Black Swans.



**FIGURE 26.** Case 2 (left): **Fragile.** Limited gains, larger losses. Increasing uncertainty in the system causes an augmentation of mostly (sometimes only) negative outcomes, just negative Black Swans. Case 3 (right): **Antifragile.** Increasing randomness and uncertainty in the system raises the probability of very favorable outcomes, and accordingly expand the expected payoff. It shows how discovery is, mathematically, exactly like an anti-airplane delay.

### TECHNICAL VERSION OF FAT TONY'S "NOT THE SAME 'TING,' " OR THE CONFLATION OF EVENTS AND EXPOSURE TO EVENTS

This note will also explain a "convex transformation."

$f(x)$  is *exposure* to the variable  $x$ .  $f(x)$  can equivalently be called "payoff from  $x$ ," "exposure to  $x$ ," even "utility of payoff from  $x$ " where we introduce in  $f$  a utility function.  $x$  can be anything.

**Example:**  $x$  is the intensity of an earthquake on some scale in some specific area,  $f(x)$  is the number of persons dying from it. We can easily see that  $f(x)$  can be made more predictable than  $x$  (if we force people to stay away from a specific area or build to some standards, etc.).

**Example:**  $x$  is the number of meters of my fall to the ground when someone pushes me from height  $x$ ,  $f(x)$  is a measure of my physical condition from the effect of the fall. Clearly I cannot predict  $x$  (who will push me, rather  $f(x)$ ).

**Example:**  $x$  is the number of cars in NYC at noon tomorrow,  $f(x)$  is travel time from point A to point B for a certain agent.  $f(x)$  can be made more predictable than  $x$  (take the subway, or, even better, walk).

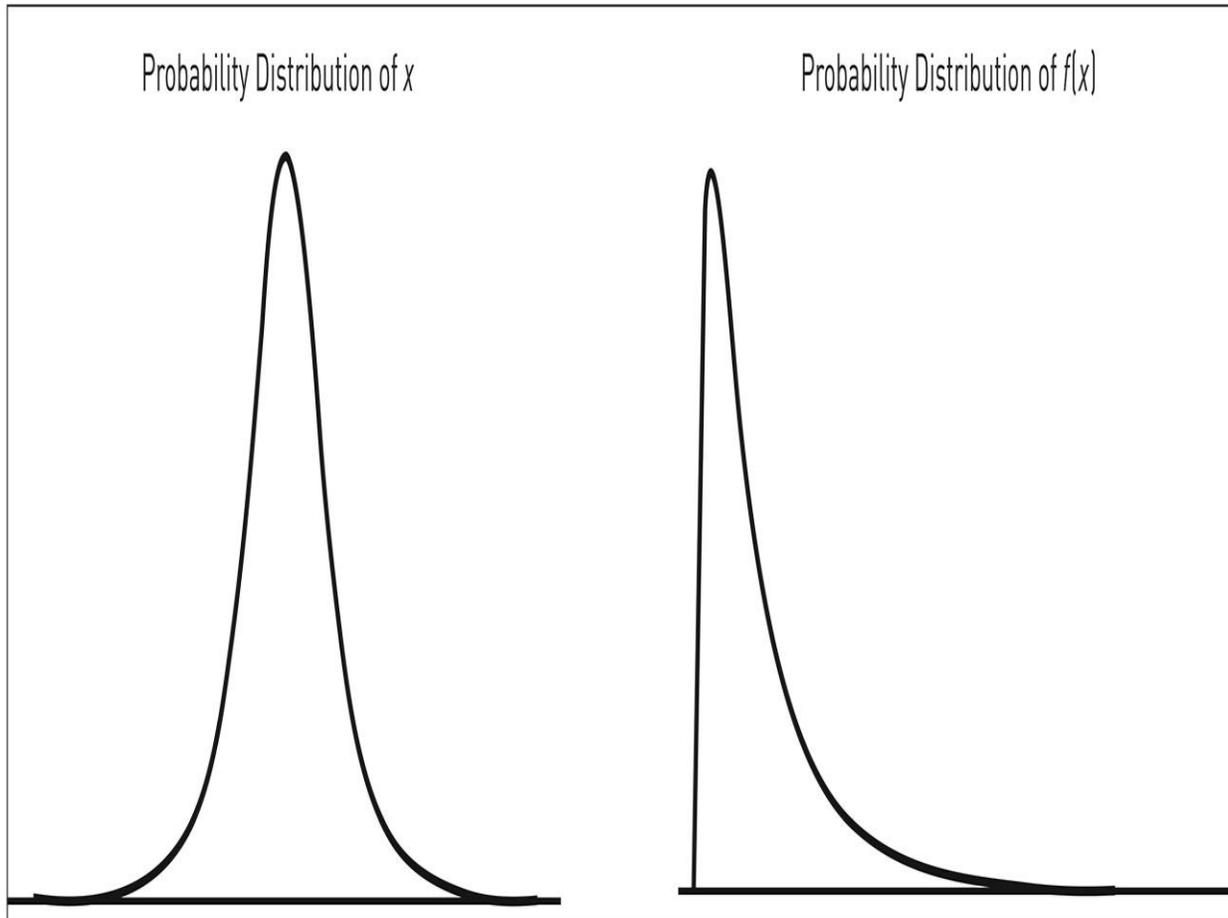
Some people talk about  $f(x)$  thinking they are talking about  $x$ . This is the problem of the **conflation of event and exposure**. This error present in Aristotle is virtually ubiquitous in the philosophy of probability (say, Hacking).

One can become antifragile to  $x$  without understanding  $x$ , through convexity of  $f(x)$ .

The answer to the question “what do you do in a world you don’t understand?” is, simply, work on the undesirable states of  $f(x)$ .

It is often easier to modify  $f(x)$  than to get better knowledge of  $x$ . (In other words, robustification rather than forecasting Black Swans.)

**Example:** If I buy an insurance on the market, here  $x$ , dropping more than 20 percent,  $f(x)$  will be independent of the part of the probability distribution of  $x$  that is below 20 percent and impervious to changes in its scale parameter. (This is an example of a barbell.)



**FIGURE 27. Convex Transformation ( $f(x)$  is a convex function of  $x$ ).** The difference between  $x$  and exposure to  $x$ . There is no downside risk in the second graph. The key is to modify  $f(x)$  in order to make knowledge of the properties of  $x$  on the left side of the distribution as irrelevant as possible. This operation is called convex transformation, nicknamed “barbell” here.

**Green lumber fallacy:** When one confuses  $f(x)$  for another function  $g(x)$ , one that has different nonlinearities.

**More technically:** If one is antifragile to  $x$ , then the variance (or volatility, or other measures of variation) of  $x$  benefit  $f(x)$ , since distributions that are skewed have their mean depend on the variance and when skewed right, their expectation increases with variance (the lognormal, for instance, has for mean a term that includes  $+1/2 \sigma^2$ ).

Further, the probability distribution of  $f(x)$  is markedly different from that of  $x$ , particularly in the presence of nonlinearities.

When  $f(x)$  is convex (concave) monotonically,  $f(x)$  is right (left) skewed.

When  $f(x)$  is increasing and convex on the left then concave to the right, the probability distribution of  $f(x)$  is thinner-tailed than that of  $x$ . For instance, in Kahneman-Tversky's prospect theory, the so-called utility of changes in wealth is more "robust" than that of wealth.

**Why payoff matters more than probability (technical):** Where  $p(x)$  is the density, the expectation, that is  $\int f(x)p(x)dx$ , will depend increasingly on  $f$  rather than  $p$ , and the more nonlinear  $f$ , the more it will depend on  $f$  rather than  $p$ .

**THE FOURTH QUADRANT (TALEB, 2009)**

The idea is that tail events are not computable (in fat-tailed domains), but we can assess our exposure to the problem. Assume  $f(x)$  is an increasing function, [Table 10](#) connects the idea to the notion of the Fourth Quadrant.

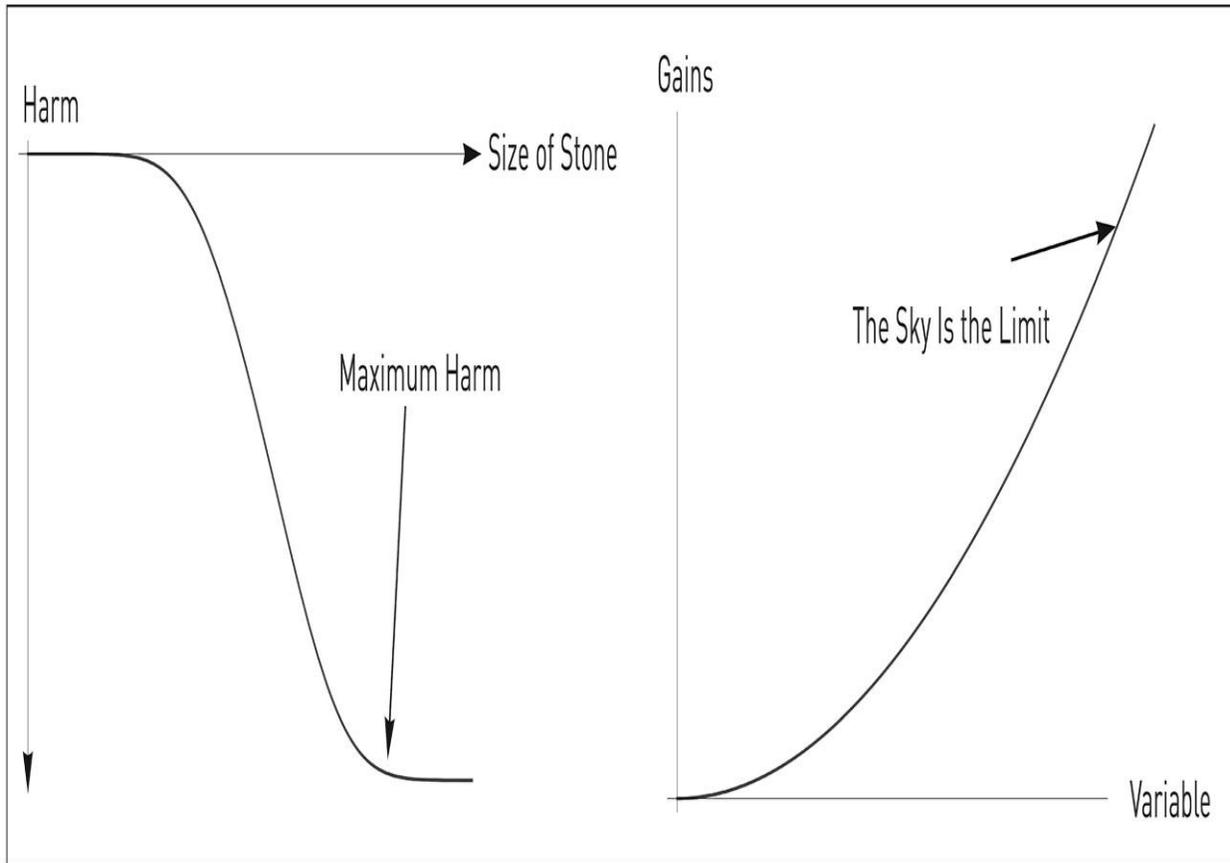
<b>TABLE 10</b>		
	<i>THIN-TAILED DISTRIBUTION FOR X</i>	<i>FAT-TAILED DISTRIBUTION FOR X</i>
$f(x)$ "mitigating" by clipping extreme outcomes, i.e., convex-concave	Very robust outcome	Quite robust outcome
$f(x)$ concave-convex, exacerbates remote outcomes	Robust outcome (sort of)	FOURTH QUADRANT Fragile (if $f(x)$ is concave) or antifragile

**LOCAL AND GLOBAL CONVEXITIES (TECHNICAL)**

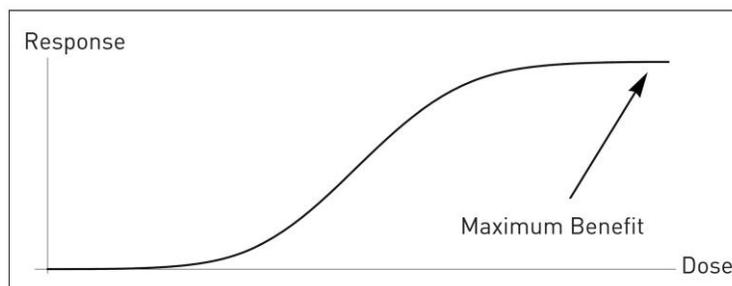
Nothing is open-ended in nature—death is a maximum outcome for a unit. So things end up convex on one end, concave on the other.

In fact, there is maximum harm at some point in things biological. Let us revisit the concave figure of the stone and pebbles in [Chapter 18](#): by widening the range we see that

boundedness of harm brings convexities somewhere. Concavity was dominant, but local. [Figure 28](#) looks at the continuation of the story of the stone and pebbles.



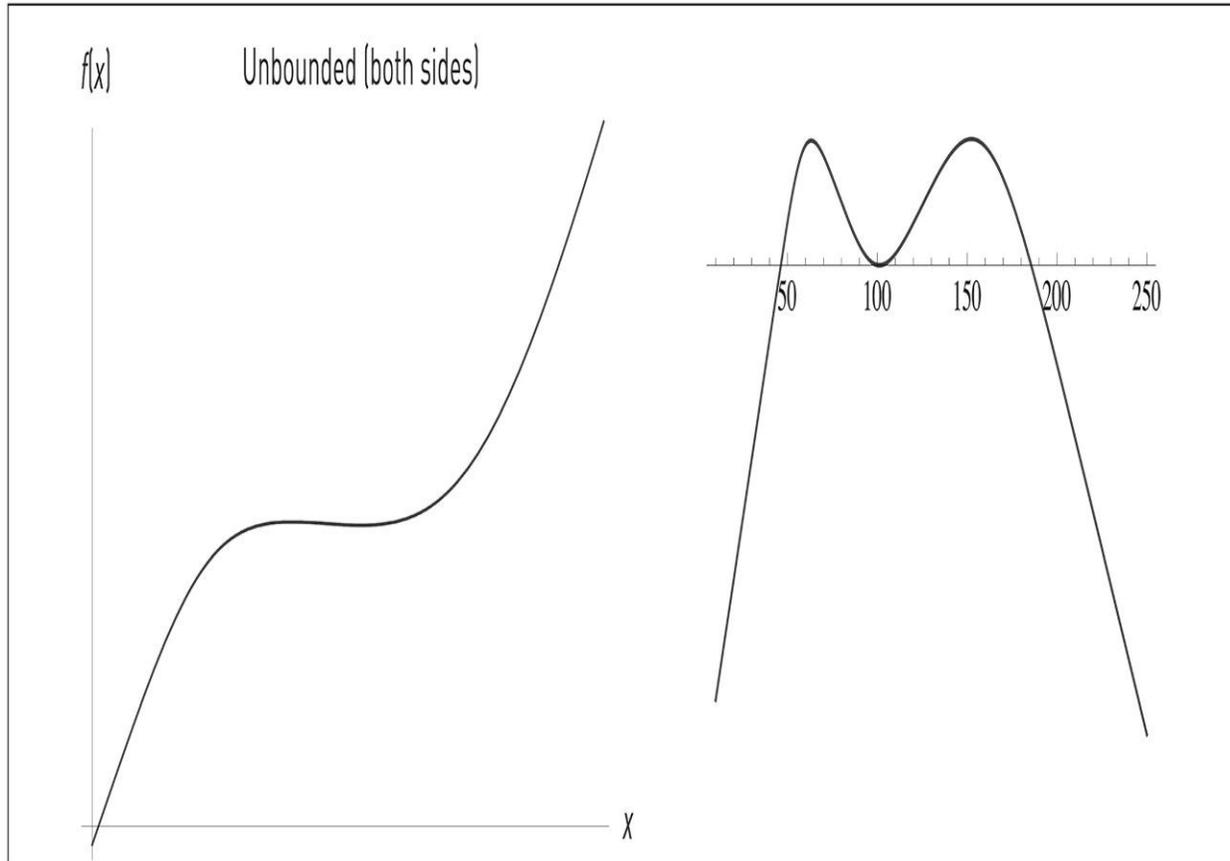
**FIGURE 28.** The left graph shows a broader range in the story of the stone and pebbles in [Chapter 18](#). At some point, the concave turns convex as we hit maximum harm. The right graph shows strong antifragility, with no known upper limit (leading to Extremistan). These payoffs are only available in economic variables, say, sales of books, or matters unbounded or near-unbounded. I am unable to find such an effect in nature.



**FIGURE 29. Weak Antifragility (Mediocristan), with bounded maximum.** Typical in nature.

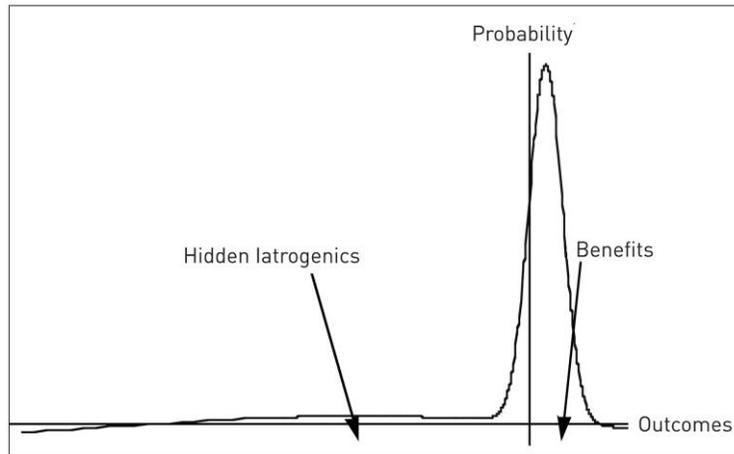
## FREAK NONLINEARITIES (VERY TECHNICAL)

The next two types of nonlinearities are almost never seen outside of economic variables; they are particularly limited to those caused by derivatives.



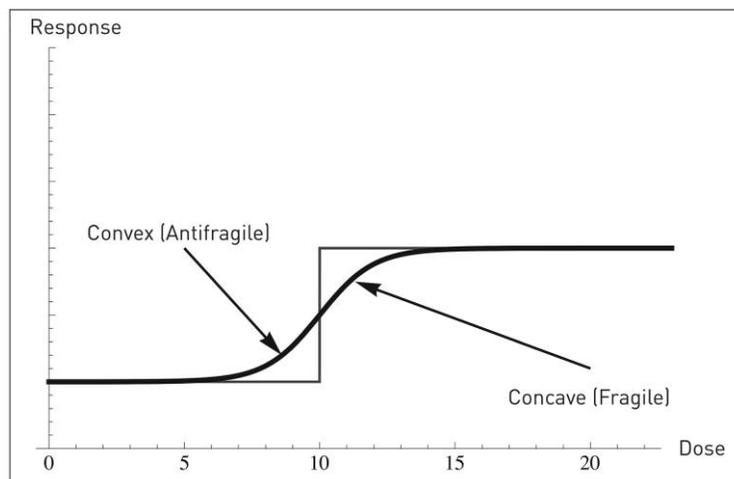
**FIGURE 30.** The left graph shows a convex-concave increasing function, the opposite of the bounded dose-response functions we see in nature. It leads to Type 2, Fragile (very, very fat tails). The right graph shows the most dangerous of all: pseudoconvexity. Local antifragility, global fragility.

**MEDICAL NONLINEARITIES AND THEIR PROBABILITY CORRESPONDENCE ([CHAPTERS 21 & 22](#))**



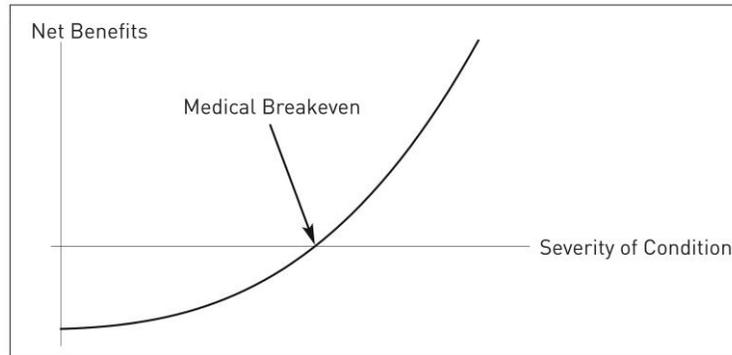
**FIGURE 31. Medical Iatrogenics:** Case of small benefits and large Black Swan–style losses seen in probability space. Iatrogenics occurs when we have small identifiable gains (say, avoidance of small discomfort or a minor infection) and exposure to Black Swans with delayed invisible large side effects (say, death). These concave benefits from medicine are just like selling a financial option (plenty of risk) against small tiny immediate gains while claiming “evidence of no harm.”

In short, for a healthy person, there is a small probability of disastrous outcomes (discounted because unseen and not taken into account), and a high probability of mild benefits.



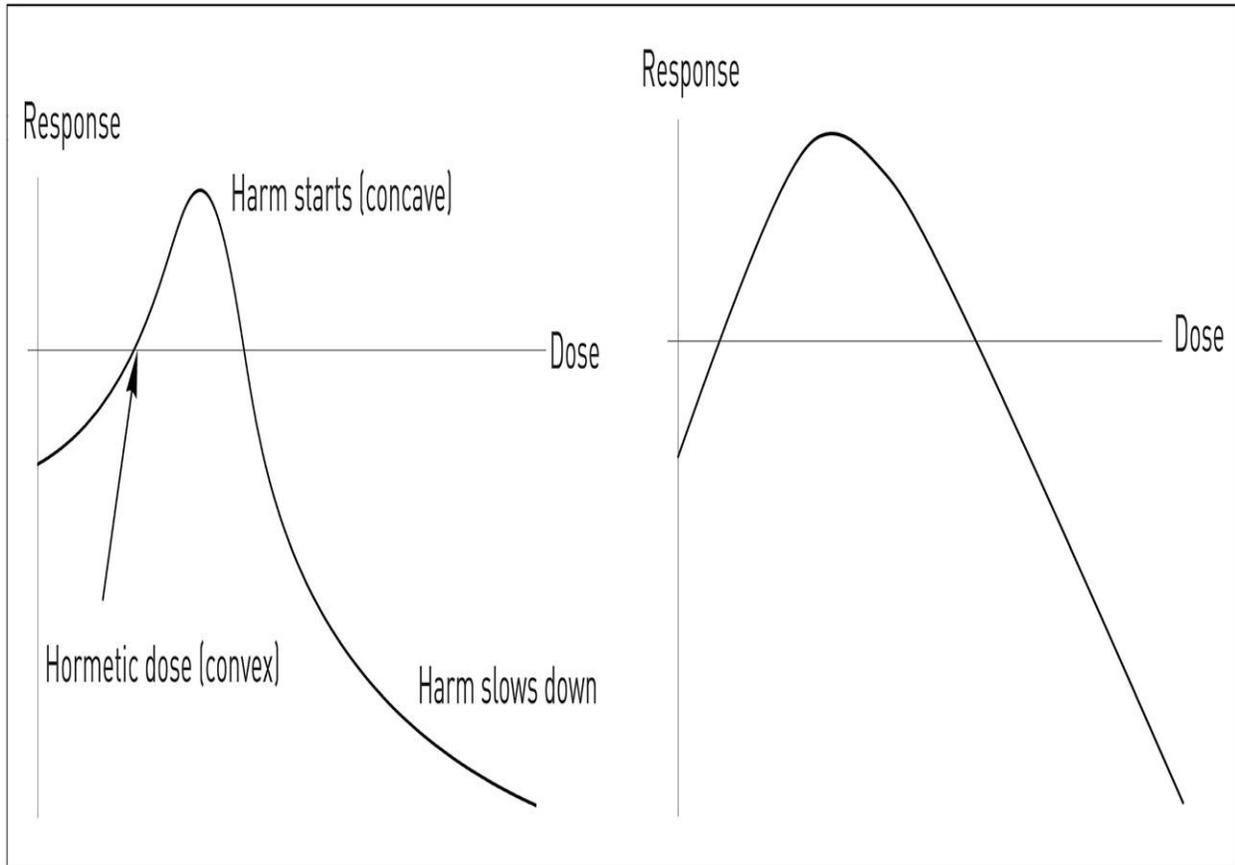
**FIGURE 32. Nonlinearities in biology.** The shape convex-concave necessarily flows from anything increasing (monotone, i.e., never decreasing) and bounded, with maximum and minimum values, i.e., does not reach infinity from either side. At low levels, the dose response is convex (gradually more and more effective). Additional doses tend to become gradually ineffective or start hurting. The same can apply to anything consumed in too much regularity. This type of graph necessarily applies to any situation bounded on both sides, with a known minimum and maximum (saturation), which includes happiness.

For instance, if one considers that there exists a maximum level of happiness and unhappiness, then the general shape of this curve with convexity on the left and concavity on the right has to hold for happiness (replace “dose” with “wealth” and “response” with “happiness”). Kahneman-Tversky prospect theory models a similar shape for “utility” of changes in wealth, which they discovered empirically.



**FIGURE 33.** Recall the hypertension example. On the vertical axis, we have the benefits of a treatment, on the horizontal, the severity of the condition. The arrow points at the level where probabilistic gains match probabilistic harm. Iatrogenics disappears nonlinearly as a function of the severity of the condition. This implies that when the patient is very ill, the distribution shifts to antifragile (thicker right tail), with large benefits from the treatment over possible iatrogenics, little to lose.

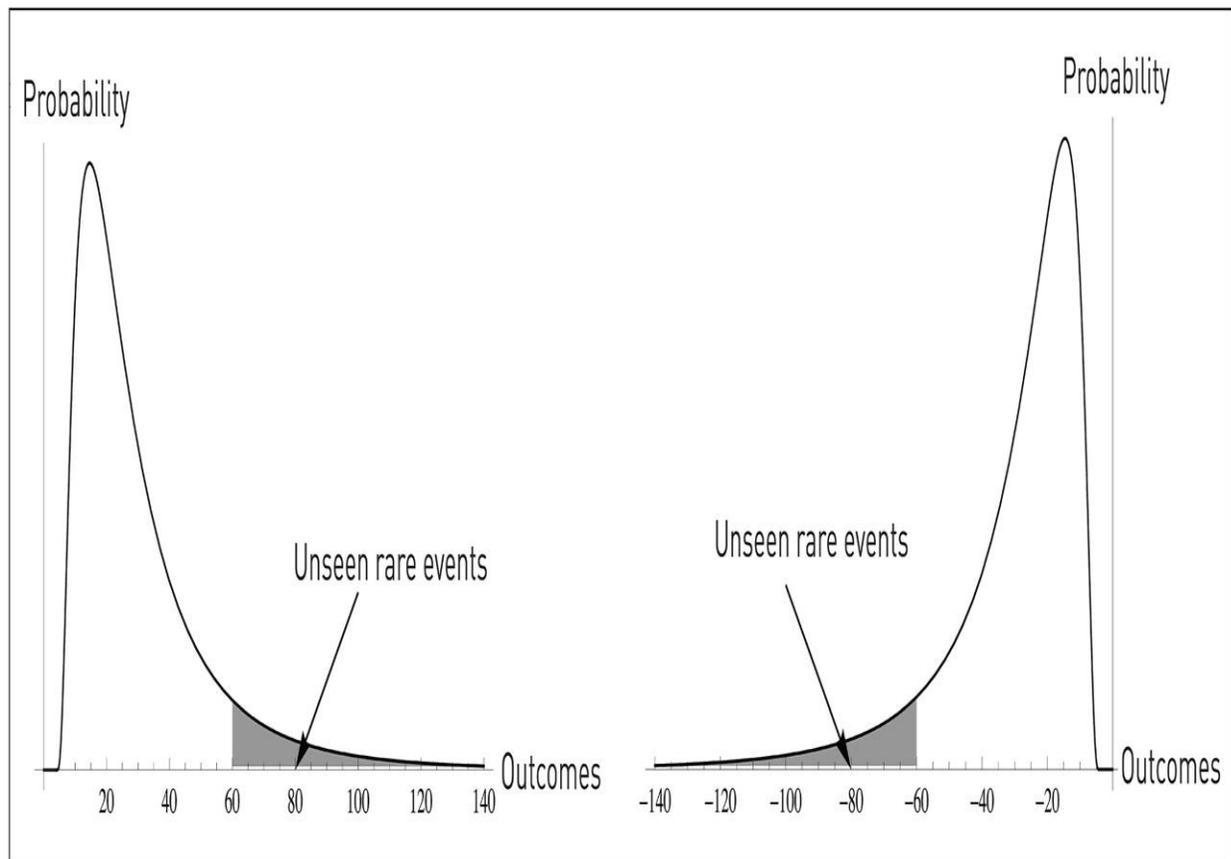
Note that if you increase the treatment you hit concavity from maximum benefits, a zone not covered in the graph—seen more broadly, it would look like the preceding graph.



**FIGURE 34.** The left graph shows hormesis for an organism (similar to [Figure 19](#)): we can see a stage of benefits as the dose increases (initially convex) slowing down into a phase of harm as we increase the dose a bit further (initially concave); then we see things flattening out at the level of maximum harm (beyond a certain point, the organism is dead so there is such a thing as a bounded and known worst case scenario in biology). To the right, a wrong graph of hormesis in medical textbooks showing initial concavity, with a beginning that looks linear or slightly concave.

## THE INVERSE TURKEY PROBLEM

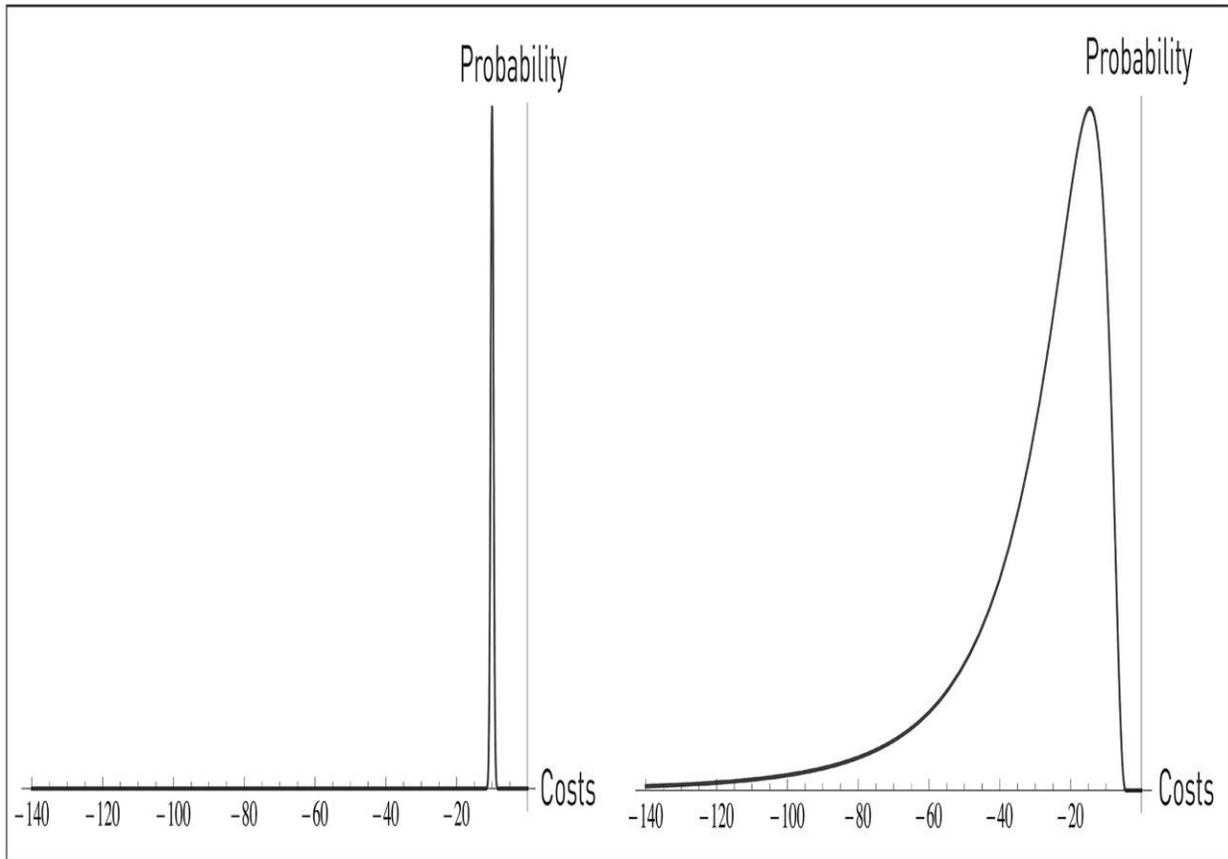
# THE INVERSE TURKEY PROBLEM



**FIGURE 35. Antifragile, Inverse Turkey Problem:** The unseen rare event is positive. When you look at a positively skewed (antifragile) time series and make inferences about the unseen, you miss the good stuff and underestimate the benefits (the Pisano, 2006a, 2006b, mistake). On the right, the other Harvard problem, that of Froot (2001). The filled area corresponds to what we do not tend to see in small samples, from insufficiency of points. Interestingly the shaded area increases with model error. The more technical sections call this zone  $\omega_B$  (turkey) and  $\omega_C$  (inverse turkey).

## DIFFERENCE BETWEEN POINT ESTIMATES AND DISTRIBUTIONS

Let us apply this analysis to how planners make the mistakes they make, and why deficits tend to be worse than planned:



**FIGURE 36. The gap between predictions and reality:** probability distribution of outcomes from costs of projects in the minds of planners (left) and in reality (right). In the first graph they assume that the costs will be both low and quite certain. The graph on the right shows outcomes to be both worse and more spread out, particularly with higher possibility of unfavorable outcomes. Note the fragility increase owing to the swelling left tail.

*This misunderstanding of the effect of uncertainty applies to government deficits, plans that have IT components, travel time (to a lesser degree), and many more. We will use the same graph to show model error from underestimating fragility by assuming that a parameter is constant when it is random. This is what plagues bureaucrat-driven economics (next discussion).*

## Appendix II (Very Technical):

### WHERE MOST ECONOMIC MODELS FRAGILIZE AND BLOW PEOPLE UP

---

When I said “technical” in the main text, I may have been fibbing. Here I am not.

**The Markowitz incoherence:** Assume that someone tells you that the probability of an event is exactly zero. You ask him where he got this from. “Baal told me” is the answer. In such case, the person is coherent, but would be deemed unrealistic by non-Baalists. But if on the other hand, the person tells you “I *estimated* it to be zero,” we have a problem. The person is both unrealistic and inconsistent. Something estimated needs to have an estimation error. So probability cannot be zero if it is estimated, its lower bound is linked to the estimation error; the higher the estimation error, the higher the probability, up to a point. As with Laplace’s argument of total ignorance, an infinite estimation error pushes the probability toward  $1/2$ .

We will return to the implication of the mistake; take for now that anything estimating a parameter and then putting it into an equation is different from estimating the equation across parameters (same story as the health of the grandmother, the average temperature, here “estimated” is irrelevant, what we need is average health across temperatures). And Markowitz showed his incoherence by starting his “semi-nal” paper with “Assume you know  $E$  and  $V$ ” (that is, the expectation and the variance). At the end of the paper he accepts that they need to be estimated, and what is worse, with a combination of statistical techniques and the “judgment of practical men.” Well, if these parameters need to be estimated, with an error, then the derivations need to be written differently and, of course, we would have no paper—and no Markowitz paper, no blowups, no modern finance, no fragilistas teaching junk to students.... Economic models are extremely fragile to assumptions, in the sense that a slight alteration in these assumptions can, as we will see, lead to extremely consequential differences in the results. And, to make matters worse, many of these models are “back-fit” to assumptions, in the sense that the hypotheses are selected to make the math work, which makes them ultrafragile and ultrafragilizing.

**Simple example:** Government deficits.

We use the following deficit example owing to the way calculations by governments and government agencies currently miss convexity terms (and have a hard time accepting it). Really, they don’t take them into account. The example illustrates:

- (a) missing the stochastic character of a variable known to affect the model but deemed deterministic (and fixed), and

(b)  $F$ , the function of such variable, is convex or concave with respect to the variable.

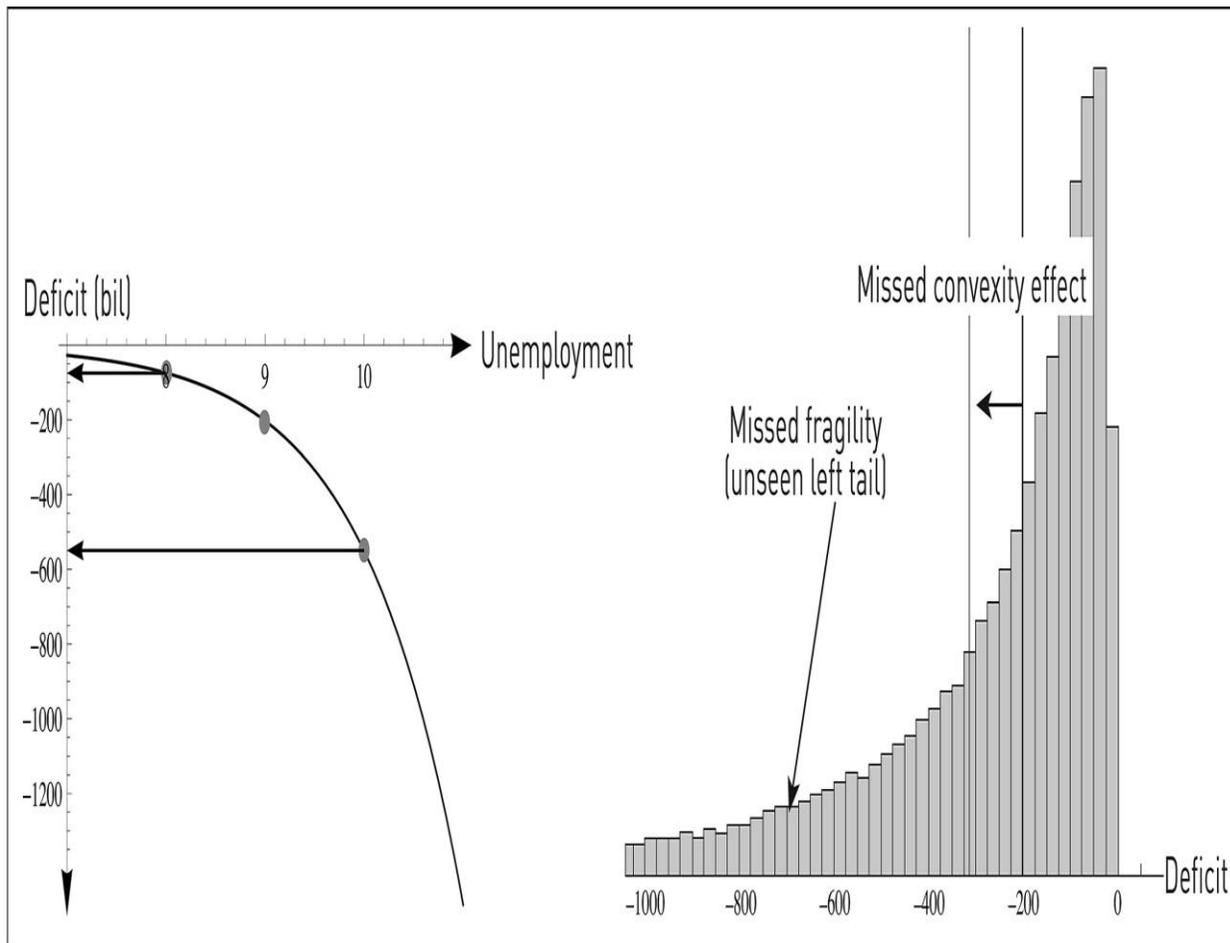
Say a government estimates unemployment for the next three years as averaging 9 percent; it uses its econometric models to issue a forecast balance  $B$  of a two-hundred-billion deficit in the local currency. But it misses (like almost everything in economics) that unemployment is a stochastic variable. Employment over a three-year period has fluctuated by 1 percent on average. We can calculate the effect of the error with the following:

*Unemployment at 8%, Balance  $B(8\%) = -75$  bn (improvement of 125 bn)*

*Unemployment at 9%, Balance  $B(9\%) = -200$  bn*

*Unemployment at 10%, Balance  $B(10\%) = -550$  bn (worsening of 350 bn)*

The concavity bias, or negative convexity bias, from underestimation of the deficit is  $-112.5$  bn, since  $\frac{1}{2} \{B(8\%) + B(10\%)\} = -312$  bn, not  $-200$  bn. This is the exact case of the **inverse philosopher's stone**.



**FIGURE 37. Nonlinear transformations allow the detection of both model convexity bias and fragility.** Illustration of the example: histogram from Monte Carlo simulation of government deficit as a left-tailed random variable simply as a result of randomizing unemployment, of which it is a concave function. The method of point estimate would assume a Dirac stick at  $-200$ , thus underestimating both the **expected** deficit ( $-312$ ) and the tail fragility of it. (From Taleb and Douady, 2012).

**Application: Ricardian Model and Left Tail—The Price of Wine Happens to Vary**

For almost two hundred years, we’ve been talking about an idea by the economist David Ricardo called “comparative advantage.” In short, it says that a country should have a certain policy based on its comparative advantage in wine or clothes. Say a country is good at both wine and clothes, better than its neighbors with whom it can trade freely. Then the visible *optimal* strategy would be to specialize in either wine or clothes, whichever fits the best and minimizes opportunity costs. Everyone would then be happy. The analogy by the economist Paul Samuelson is that if someone happens to be the best doctor in town and, at the same time, the best secretary, then it would be preferable to be the higher-earning doctor—as it would minimize opportunity losses—and let someone else be the secretary and buy secretarial services from him.

I agree that there are benefits in *some* form of specialization, but not from the models used to prove it. The flaw with such reasoning is as follows. True, it would be inconceivable for a doctor to become a part-time secretary just because he is good at it. But, at the same time, we can safely assume that being a doctor insures some professional stability: People will not cease to get sick and there is a higher social status associated with the profession than that of secretary, making the profession more desirable. But assume now that in a two-country world, a country specialized in wine, hoping to sell its specialty in the market to the other country, and that *suddenly the price of wine drops precipitously*. Some change in taste caused the price to change. Ricardo’s analysis assumes that both the market price of wine and the costs of production remain constant, and there is no “second order” part of the story.

Double-tap or move your cursor over the table and click to zoom.

**TABLE 11 • RICARDO’S ORIGINAL EXAMPLE  
(COSTS OF PRODUCTION PER UNIT)**

	<i>CLOTH</i>	<i>WINE</i>
<i>Britain</i>	100	110
<i>Portugal</i>	90	80

**The logic:** The table above shows the cost of production, normalized to a selling price of one unit each, that is, assuming that these trade at equal price (1 unit of cloth for 1 unit of wine). What looks like the paradox is as follows: that Portugal produces cloth cheaper than Britain, but should buy cloth from there instead, using the gains from the sales of wine. In the absence of transaction and transportation costs, it is efficient for Britain to produce just cloth, and Portugal to only produce wine.

The idea has always attracted economists because of its paradoxical and counterintuitive aspect. For instance, in an article “Why Intellectuals Don’t Understand Comparative Advantage” (Krugman, 1998), Paul Krugman, who fails to understand the concept himself, as this essay and his technical work show him to be completely innocent of tail events and risk management, makes fun of other intellectuals such as S. J. Gould who understand tail events albeit intuitively rather than analytically. (Clearly one cannot talk about returns and gains without discounting these benefits by the offsetting risks.) The article shows Krugman falling into the critical and dangerous mistake of confusing function of average and average of function. (Traditional Ricardian analysis assumes the variables are endogenous, but does not add a layer of stochasticity.)

Now consider the price of wine and clothes *variable*—which Ricardo did not assume—with the numbers above the unbiased average long-term value. Further assume that they follow a fat-tailed distribution. Or consider that their costs of production vary according to a fat-tailed distribution.

If the price of wine in the international markets rises by, say, 40 percent, then there are clear benefits. But should the price drop by an equal percentage, –40 percent, then massive harm would ensue, in magnitude larger than the benefits should there be an equal rise. There are concavities to the exposure—severe concavities.

And clearly, should the price drop by 90 percent, the effect would be disastrous. Just imagine what would happen to your household should you get an instant and unpredicted 40 percent pay cut. Indeed, we have had problems in history with countries specializing in some goods, commodities, and crops that happen to be not just volatile, but extremely volatile. And disaster does not necessarily come from variation in price, but problems in production: suddenly, you can’t produce the crop because of a germ, bad weather, or some other hindrance.

A bad crop, such as the one that caused the Irish potato famine in the decade around 1850, caused the death of a million and the emigration of a million more (Ireland’s entire population at the time of this writing is only about six million, if one includes the northern part). It is very hard to reconvert resources—unlike the case in the doctor-typist story, countries don’t have the ability to change. Indeed, monoculture (focus on a single crop) has turned out to be lethal in history—one bad crop leads to devastating famines.

The other part missed in the doctor-secretary analogy is that countries don’t have family and friends. A doctor has a support community, a circle of friends, a collective that takes care of him, a father-in-law to borrow from in the event that he needs to

reconvert into some other profession, a state above him to help. Countries don't. Further, a doctor has savings; countries tend to be borrowers.

So here again we have fragility to second-order effects.

**Probability Matching:** The idea of comparative advantage has an analog in probability: if you sample from an urn (with replacement) and get a black ball 60 percent of the time, and a white one the remaining 40 percent, the optimal strategy, according to textbooks, is to bet 100 percent of the time on black. The strategy of betting 60 percent of the time on black and 40 percent on white is called “probability matching” and considered to be an error in the decision-science literature (which I remind the reader is what was used by Triffat in [Chapter 10](#)). People’s instinct to engage in probability matching appears to be sound, not a mistake. In nature, probabilities are unstable (or unknown), and probability matching is similar to redundancy, as a buffer. So if the probabilities change, in other words if there is another layer of randomness, then the optimal strategy is probability matching.

**How specialization works:** The reader should not interpret what I am saying to mean that specialization is not a good thing—only that one should establish such specialization after addressing fragility and second-order effects. Now I do believe that Ricardo is ultimately right, but not from the models shown. Organically, systems without top-down controls would specialize progressively, slowly, and over a long time, through trial and error, get the right amount of specialization—not through some bureaucrat using a model. To repeat, systems make small errors, design makes large ones.

So the imposition of Ricardo’s insight-turned-model by some social planner would lead to a blowup; letting tinkering work slowly would lead to efficiency—true efficiency. The role of policy makers should be to, *via negativa* style, allow the emergence of specialization by preventing what hinders the process.

#### **A More General Methodology to Spot Model Error**

**Model second-order effects and fragility:** Assume we have the right model (which is a very generous assumption) but are uncertain about the parameters. As a generalization of the deficit/employment example used in the previous section, say we are using  $f$ , a simple function:  $f(x|\bar{\alpha})$ , where  $\bar{\alpha}$  is supposed to be the average expected input variable, where we take  $\varphi$  as the distribution of  $\alpha$  over its domain  $\mathcal{P}_\alpha$ ,  
$$\bar{\alpha} = \int_{\mathcal{P}_\alpha} \alpha \varphi(\alpha) d\alpha .$$

**The philosopher’s stone:** The mere fact that  $\alpha$  is uncertain (since it is estimated) might lead to a bias if we perturbate from the *inside* (of the integral), i.e., stochasticize the parameter deemed fixed. Accordingly, the convexity bias is easily measured as the difference between (a) the function  $f$  integrated across values of potential  $\alpha$ , and (b)  $f$  estimated for a single value of  $\alpha$  deemed to be its average. The convexity bias (philosopher’s stone)  $\omega_A$  becomes:<sup>1</sup>

$$\omega_A \equiv \int_{\varnothing_x} \int_{\varnothing_\alpha} f(x | \alpha) \varphi(\alpha) d\alpha dx - \int_{\varnothing_x} f(x) \left( \int_{\varnothing_\alpha} \alpha \varphi(\alpha) d\alpha \right) dx$$

**The central equation:** Fragility is a partial philosopher's stone below  $K$ , hence  $\omega_B$  the missed fragility is assessed by comparing the two integrals below  $K$  in order to capture the effect on the left tail:

$$\omega_B(K) \equiv \int_{-\infty}^K \int_{\varnothing_\alpha} f(x | \alpha) \varphi(\alpha) d\alpha dx - \int_{-\infty}^K f(x) \left( \int_{\varnothing_\alpha} \alpha \varphi(\alpha) d\alpha \right) dx$$

which can be approximated by an interpolated estimate obtained with two values of  $\alpha$  separated from a midpoint by  $\Delta\alpha$  its mean deviation of  $\alpha$  and estimating

$$\omega_B(K) \equiv \int_{-\infty}^K \frac{1}{2} (f(x | \bar{\alpha} + \Delta\alpha) + f(x | \bar{\alpha} - \Delta\alpha)) dx - \int_{-\infty}^K f(x | \bar{\alpha}) dx$$

Note that antifragility  $\omega_C$  is integrating from  $K$  to infinity. We can probe  $\omega_B$  by point estimates of  $f$  at a level of  $X \leq K$

$$\omega'_B(X) = \frac{1}{2} (f(X | \bar{\alpha} + \Delta\alpha) + f(X | \bar{\alpha} - \Delta\alpha)) - f(X | \bar{\alpha})$$

so that

$$\omega_B(K) = \int_{-\infty}^K \omega'_B(x) dx$$

which leads us to the fragility detection heuristic (Taleb, Canetti, et al., 2012). In particular, if we assume that  $\omega'_B(X)$  has a constant sign for  $X \leq K$ , then  $\omega_B(K)$  has the same sign. The detection heuristic is a perturbation in the tails to probe fragility, by checking the function  $\omega'_B(X)$  at any level  $X$ .

Double-tap or move your cursor over the table and click to zoom.

**TABLE 12**

<i>MODEL</i>	<i>SOURCE OF FRAGILITY</i>	<i>REMEDY</i>
<i>Portfolio theory, mean-variance, etc.</i>	Assuming knowledge of the parameters, not integrating models across parameters, relying on (very unstable) correlations. Assumes $\omega_A$ (bias) and $\omega_B$ (fragility) = 0	1/n (spread as large a number of exposures as manageable), barbells, progressive and organic construction, etc.
<i>Ricardian comparative advantage</i>	Missing layer of randomness in the price of wine may imply total reversal of allocation. Assumes $\omega_A$ (bias) and $\omega_B$ (fragility) = 0	Natural systems find their own allocation through tinkering
<i>Samuelson optimization</i>	Concentration of sources of randomness under concavity of loss function. Assumes $\omega_A$ (bias) and $\omega_B$ (fragility) = 0	Distributed randomness
<i>Arrow-Debreu lattice state-space</i>	Ludic fallacy: assumes exhaustive knowledge of outcomes and knowledge of probabilities. Assumes $\omega_A$ (bias), $\omega_B$ (fragility), and $\omega_C$ (antifragility) = 0	Use of metaprobabilities changes entire model implications
<i>Dividend cash flow models</i>	Missing stochasticity causing convexity effects. Mostly considers $\omega_C$ (antifragility) = 0	Heuristics

**Portfolio fallacies:** Note one fallacy promoted by Markowitz users: *portfolio theory entices people to diversify, hence it is better than nothing*. Wrong, you finance fools: it pushes them to optimize, hence overallocate. It does not drive people to take less risk based on diversification, but causes them to take more open positions owing to perception of offsetting statistical properties—making them vulnerable to model error, and especially vulnerable to the underestimation of tail events. To see how, consider two investors facing a choice of allocation across three items: cash, and securities  $A$  and  $B$ . The investor who does not know the statistical properties of  $A$  and  $B$  and knows he doesn't know will allocate, say, the portion he does not want to lose to cash, the rest into  $A$  and  $B$ —according to whatever heuristic has been in traditional use. The investor who thinks he knows the statistical properties, with parameters  $\sigma_A$ ,  $\sigma_B$ ,  $\rho_{A,B}$ , will allocate  $\omega_A$ ,  $\omega_B$  in a way to put the total risk at some target level (let us ignore the expected return for this). The lower his perception of the correlation  $\rho_{A,B}$ , the worse his exposure to model error. Assuming he thinks that the correlation  $\rho_{A,B}$  is 0, he will be overallocated by  $\frac{1}{3}$  for extreme events. But if the poor investor has the illusion that the correlation is  $-1$ , he will be maximally overallocated to his  $A$  and  $B$  investments. If the investor uses leverage, we end up with the story of Long-Term Capital Management, which turned out to be fooled by the parameters. (In real life, unlike in economic papers, things tend to change; for Baal's sake, they change!) We can repeat the idea for each parameter  $\sigma$  and see how lower perception of this  $\sigma$  leads to overallocation.

I noticed as a trader—and obsessed over the idea—that correlations were never the same in different measurements. Unstable would be a mild word for them: 0.8 over a long period becomes  $-0.2$  over another long period. A pure sucker game. At times of stress, correlations experience even more abrupt changes—without any reliable regularity, in spite of attempts to model “stress correlations.” Taleb (1997) deals with the effects of stochastic correlations: One is only safe shorting a correlation at 1, and buying it at  $-1$ —which seems to correspond to what the  $1/n$  heuristic does.

**Kelly Criterion vs. Markowitz:** In order to implement a full Markowitz-style optimization, one needs to know the entire joint probability distribution of all assets for the entire future, plus the exact utility function for wealth at all future times. And without errors! (We saw that estimation errors make the system explode.) Kelly's method, developed around the same period, requires no joint distribution or utility function. In practice one needs the ratio of expected profit to worst-case return—dynamically adjusted to avoid ruin. In the case of barbell transformations, the worst case is guaranteed. And model error is much, much milder under Kelly criterion. Thorp (1971, 1998), Haigh (2000).

The formidable Aaron Brown holds that Kelly's ideas were rejected by economists—in spite of the practical appeal—because of their love of general theories for all asset prices.

Note that bounded trial and error is compatible with the Kelly criterion when one has an idea of the potential return—even when one is ignorant of the returns, if losses are bounded, the payoff will be robust and the method should outperform that of Fragilista Markowitz.

**Corporate Finance:** In short, corporate finance seems to be based on point projections, not distributional projections; thus if one perturbs cash flow projections, say, in the Gordon valuation model, replacing the fixed—and known—growth (and other parameters) by continuously varying jumps (particularly under fat-tailed distributions), companies deemed “expensive,” or those with high growth, but low earnings, could markedly increase in expected value, something the market prices heuristically but without explicit reason.

**Conclusion and summary:** Something the economics establishment has been missing is that having the right model (which is a very generous assumption), but being uncertain about the parameters will invariably lead to an increase in fragility in the presence of convexity and nonlinearities.

## FUHGETABOUT SMALL PROBABILITIES

Now the meat, beyond economics, the more general problem with probability and its mismeasurement.

*How Fat Tails (Extremistan) Come from  
Nonlinear Responses to Model Parameters*

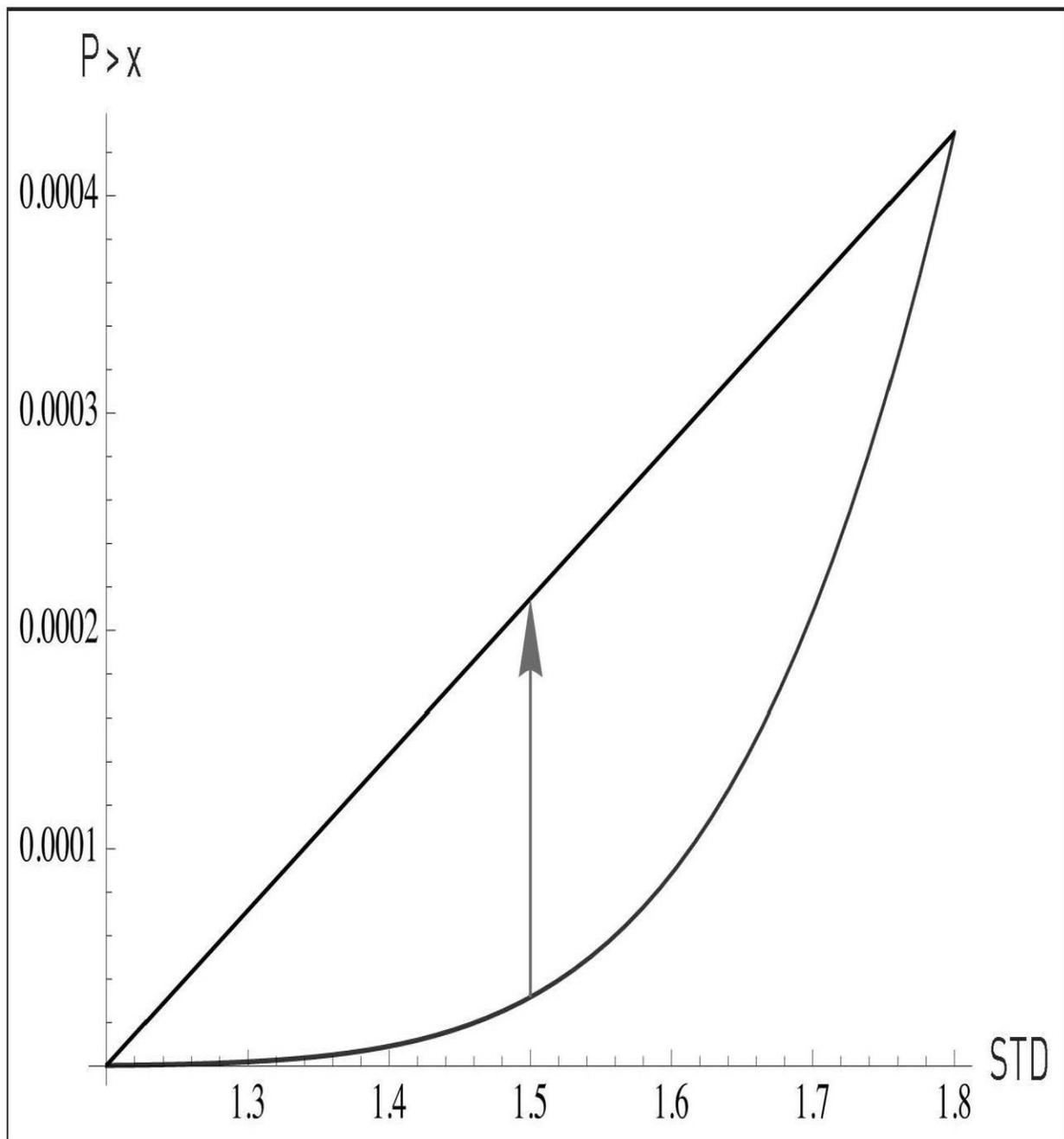
Rare events have a certain property—missed so far at the time of this writing. We deal with them using a model, a mathematical contraption that takes input parameters and outputs the probability. The more parameter uncertainty there is in a model designed to compute probabilities, the more small probabilities tend to be underestimated. Simply, small probabilities are convex to errors of computation, as an airplane ride is concave to errors and disturbances (remember, it gets longer, not shorter). The more sources of disturbance one forgets to take into account, the longer the airplane ride compared to the naive estimation.

We all know that to compute probability using a standard Normal statistical distribution, one needs a parameter called *standard deviation*—or something similar that characterizes the scale or dispersion of outcomes. But uncertainty about such standard deviation has the effect of making the small probabilities rise. For instance, for a deviation that is called “three sigma,” events that should take place no more than one in 740 observations, the probability rises by 60% if one moves the standard deviation up by 5%, and drops by 40% if we move the standard deviation down by 5%. So if your error is on average a tiny 5%, the underestimation from a naive model is about 20%. Great asymmetry, but nothing yet. It gets worse as one looks for more deviations, the “six sigma” ones (alas, chronically frequent in economics): a rise of five times more. The rarer the event (i.e., the higher the “sigma”), the worse the effect from small uncertainty about what to put in the equation. With events such as ten sigma, the difference is more than a billion times. We can use the argument to show how smaller and smaller probabilities require more precision in computation. The smaller the probability, the

more a small, very small rounding in the computation makes the asymmetry massively insignificant. For tiny, very small probabilities, you need near-infinite precision in the parameters; the slightest uncertainty there causes mayhem. They are very convex to perturbations. This in a way is the argument I've used to show that small probabilities are incomputable, even if one has the right model—which we of course don't.

The same argument relates to deriving probabilities nonparametrically, from past frequencies. If the probability gets close to  $1/\text{sample size}$ , the error explodes.

This of course explains the error of Fukushima. Similar to Fannie Mae. To summarize, small probabilities increase in an accelerated manner as one changes the parameter that enters their computation.



**FIGURE 38.** The probability is convex to standard deviation in a Gaussian model. The plot shows the STD effect on  $P > x$ , and compares  $P > 6$  with an STD of 1.5 compared to  $P > 6$  assuming a linear combination of 1.2 and 1.8 (here  $a(1)=1/5$ ).

The worrisome fact is that a perturbation in  $\sigma$  extends well into the tail of the distribution in a convex way; the risks of a portfolio that is sensitive to the tails would explode. That is, we are still here in the Gaussian world! Such explosive uncertainty isn't

the result of natural fat tails in the distribution, merely small imprecision about a future parameter. It is just epistemic! So those who use these models while admitting parameters uncertainty are necessarily committing a severe inconsistency.<sup>2</sup>

Of course, uncertainty explodes even more when we replicate conditions of the non-Gaussian real world upon perturbing tail exponents. Even with a powerlaw distribution, the results are severe, particularly under variations of the tail exponent as these have massive consequences. Really, fat tails mean incomputability of tail events, little else.

### *Compounding Uncertainty (Fukushima)*

Using the earlier statement that *estimation implies error*, let us extend the logic: errors have errors; these in turn have errors. Taking into account the effect makes all small probabilities rise regardless of model—even in the Gaussian—to the point of reaching fat tails and powerlaw effects (even the so-called infinite variance) when higher orders of uncertainty are large. Even taking a Gaussian with  $\sigma$  the standard deviation having a proportional error  $a(1)$ ;  $a(1)$  has an error rate  $a(2)$ , etc. Now it depends on the higher order error rate  $a(n)$  related to  $a(n-1)$ ; if these are in constant proportion, then we converge to a very thick-tailed distribution. If proportional errors decline, we still have fat tails. In all cases mere error is not a good thing for small probability.

The sad part is that getting people to accept that every measure has an error has been nearly impossible—the event in Fukushima held to happen once per million years would turn into one per 30 if one percolates the different layers of uncertainty in the adequate manner.

<sup>1</sup> The difference between the two sides of Jensen's inequality corresponds to a notion in information theory, the Bregman divergence. Briys, Magdalou, and Nock, 2012.

<sup>2</sup> This further shows the defects of the notion of "Knightian uncertainty," since *all tails* are uncertain under the slightest perturbation and their effect is severe in fat-tailed domains, that is, economic life.

