

# The Precautionary Principle: Fragility and Black Swans from Policy Actions

Nassim Nicholas Taleb\*, Yaneer Bar-Yam†, Raphael Douady‡, Joseph Norman†, Rupert Read§

\*School of Engineering, New York University †New England Complex Systems Institute

‡ Institute of Mathematics and Theoretical Physics, C.N.R.S., Paris

§School of Philosophy, University of East Anglia

**Abstract**—The precautionary principle (PP) states that if an action or policy has a suspected risk of causing severe harm to the public domain (affecting general health or the environment globally), the action should not be taken in the absence of scientific near-certainty about its safety. Under these conditions, the burden of proof about absence of harm falls on those proposing an action, not those opposing it. PP is intended to deal with uncertainty and risk in cases where the absence of evidence and the incompleteness of scientific knowledge carries profound implications and in the presence of risks of "black swans", unforeseen and unforeseeable events of extreme consequence. Here we formalize PP, placing it within the statistical and probabilistic structure of "ruin" problems, in which a system is at risk of total failure, and in place of risk we use a formal "fragility" based approach. In these problems, what appear to be small and reasonable risks accumulate inevitably to certain irreversible harm. Traditional cost-benefit analyses, which seek to quantitatively weigh outcomes to determine the best policy option, do not apply, as outcomes may have infinite costs. Even high-benefit, high-probability outcomes do not outweigh the existence of low probability, infinite cost options—i.e. ruin. Uncertainties result in sensitivity analyses that are not mathematically well behaved. The PP is increasingly relevant due to man-made dependencies that propagate impacts of policies across the globe. In contrast, absent humanity the biosphere engages in natural experiments due to random variations with only local impacts. Our analysis makes clear that the PP is essential for a limited set of contexts and can be used to justify a limited set of actions. We discuss the implications for nuclear energy and GMOs. GMOs represent a public risk of global harm, while harm from nuclear energy is comparatively limited and better characterized. PP should be used to prescribe severe limits on GMOs.

July 24, 2014

## I. INTRODUCTION

The aim of the precautionary principle (PP) is to prevent decision makers from putting society as a whole—or a significant segment of it—at risk from the unexpected side effects of a certain type of decision. The PP states that if an action or policy has a suspected risk of causing severe harm to the public domain (such as general health or the environment), and in the absence of scientific near-certainty about the safety of the action, the burden of proof about absence of harm falls on those proposing the action. It is meant to deal with effects

of absence of evidence and the incompleteness of scientific knowledge in some risky domains.<sup>1</sup>

We believe that the PP should be evoked only in extreme situations: when the potential harm is systemic (rather than localized) and the consequences can involve total irreversible ruin, such as the extinction of human beings or all life on the planet.

The aim of this paper is to place the concept of precaution within a formal statistical and risk-analysis structure, grounding it in probability theory and the properties of complex systems. Our aim is to allow decision makers to discern which circumstances require the use of the PP and in which cases evoking the PP is inappropriate.

## II. DECISION MAKING AND TYPES OF RISK

Taking risks is necessary for individuals as well as for decision makers affecting the functioning and advancement of society. Decision and policy makers tend to assume all risks are created equal. This is not the case. Taking into account the structure of randomness in a given system can have a dramatic effect on which kinds of actions are, or are not, justified. Two kinds of potential harm must be considered when determining an appropriate approach to the role of risk in decision-making: 1) localized non-spreading impacts and 2) propagating impacts resulting in irreversible and widespread damage.

Traditional decision-making strategies focus on the case where harm is localized and risk is easy to calculate from past data. Under these circumstances, cost-benefit analyses and mitigation techniques are appropriate. The potential harm from miscalculation is bounded.

On the other hand, the possibility of irreversible and widespread damage raises different questions about the nature of decision making and what risks can be reasonably taken. This is the domain of the PP.

Criticisms are often levied against those who argue for caution portraying them as unreasonable and possibly even paranoid. Those who raise such criticisms are implicitly or explicitly advocating for a cost benefit analysis, and necessarily so. Critics of the PP have also expressed concern that

<sup>1</sup>The Rio Declaration on Environment and Development presents it as follows: "In order to protect the environment, the precautionary approach shall be widely applied by States according to their capabilities. Where there are threats of serious or irreversible damage, lack of full scientific certainty shall not be used as a reason for postponing cost-effective measures to prevent environmental degradation."

it will be applied in an overreaching manner, eliminating the ability to take reasonable risks that are needed for individual or societal gains. While indiscriminate use of the PP might constrain appropriate risk-taking, at the same time one can also make the error of suspending the PP in cases when it is vital.

Hence, a non-naive view of the precautionary principle is one in which it is only invoked when necessary, and only to prevent a certain variety of very precisely defined risks based on distinctive probabilistic structures. But, also, in such a view, the PP should never be omitted when needed.

The remainder of this section will outline the difference between the naive and non-naive approaches.

#### A. What we mean by a non-naive PP

Risk aversion and risk-seeking are both well-studied human behaviors. However, it is essential to distinguish the PP so that it is neither used naively to justify any act of caution, nor dismissed by those who wish to court risks for themselves or others.

The PP is intended to make decisions that ensure survival when statistical evidence is limited—because it has not had time to show up —by focusing on the adverse effects of "absence of evidence."

Table 1 encapsulates the central idea of the paper and shows the differences between decisions with a risk of harm (warranting regular risk management techniques) and decisions with a risk of total ruin (warranting the PP).

Standard Risk Management	Precautionary Approach
localized harm	systemic ruin
nuanced cost-benefit	avoid at all costs
statistical	fragility based
statistical	probabilistic non-statistical
variations	ruin
convergent probabilities	divergent probabilities
recoverable	irreversible
independent factors	interconnected factors
evidence based	precautionary
thin tails	fat tails
bottom-up, tinkering	top-down engineered
evolved	human-made

Table I: Two different types of risk and their respective characteristics compared

#### B. Harm vs. Ruin: When the PP is necessary

The purpose of the PP is to avoid a certain class of what, in probability and insurance, is called "ruin" problems [1]. A ruin problem is one where outcomes of risks have a non-zero probability of resulting in unrecoverable losses. An often-cited illustrative case is that of a gambler who loses his entire fortune and so cannot return to the game. In biology, an example would be a species that has gone extinct. For nature, "ruin" is ecocide: an irreversible termination of life at some scale, which could be planetwide. The large majority of variations

that occur within a system, even drastic ones, fundamentally differ from ruin problems: a system that achieves ruin cannot recover. As long as the instance is bounded, e.g. a gambler can work to gain additional resources, there may be some hope of reversing the misfortune. This is not the case when it is global.

Our concern is with public policy. While an individual may be advised to not "bet the farm," whether or not he does so is generally a matter of individual preferences. Policy makers have a responsibility to avoid catastrophic harm for society as a whole; the focus is on the aggregate, not at the level of single individuals, and on global-systemic, not idiosyncratic, harm. This is the domain of collective "ruin" problems.

### III. WHY RUIN IS SERIOUS BUSINESS

The risk of ruin is not sustainable. By the ruin theorems, if you incur a tiny probability of ruin as a "one-off" risk, survive it, then do it again (another "one-off" deal), you will eventually go bust with probability 1. Confusion arises because it may seem that the "one-off" risk is reasonable, but that also means that an additional one is reasonable. This can be quantified by recognizing that the probability of ruin approaches 1 as the number of exposures to individually small risks, say one in ten thousand, increases. For this reason a strategy of risk taking is not sustainable and we must consider *any* genuine risk of total ruin as if it were inevitable.

The good news is that some classes of risk can be deemed to be practically of probability zero: the earth survived trillions of natural variations daily over 3 billion years, otherwise we would not be here. By recognizing that normal risks are not in the category of ruin problems, we recognize also that it is not necessary or even normal to take risks that involve a possibility of ruin.

#### A. Harm vs. Ruin: When the PP is necessary

It is important to contrast and not conflate the PP and risk management. Risk management involves various strategies to make decisions based upon accounting for the effects of positive and negative outcomes and their probabilities, as well as seeking means to mitigate harm and offset losses. Risk management strategies are important for decision-making when ruin is not at stake. However, the only risk management strategy of importance in the case of the PP is ensuring that actions which can result in ruin are not taken, or equivalently, modifying potential choices of action so that ruin is not one of the possible outcomes.

More generally, we can identify three layers associated with strategies for dealing with uncertainty and risk. The first layer is the PP which addresses cases that involve potential global harm, whether probabilities are uncertain or known and whether they are large or small. The second is risk management which addresses the case of known probabilities of well-defined, bounded gains and losses. The third is risk aversion or risk-seeking behavior, which reflects quite generally the role of personal preferences for individual risks when uncertainty is present.

## B. *Ruin is forever*

A way to formalize the ruin problem in terms of the destructive consequences of actions identifies harm as not about the amount of destruction, but rather a measure of the integrated level of destruction over the time it persists. When the impact of harm extends to all future times, i.e. forever, then the harm is infinite. When the harm is infinite, the product of any non-zero probability and the harm is also infinite, and it cannot be balanced against any potential gains, which are necessarily finite. This strategy for evaluation of harm as involving the duration of destruction can be used for localized harms for better assessment in risk management. Our focus here is on the case where destruction is complete for a system or an irreplaceable aspect of a system.

For example, for humanity global devastation cannot be measured on a scale in which harm is proportional to level of devastation. The harm due to complete destruction is not the same as 10 times the destruction of 1/10 of the system. As the percentage of destruction approaches 100%, the assessment of harm diverges to infinity (instead of converging to a particular number) due to the value placed on a future that ceases to exist.

Because the “cost” of ruin is effectively infinite, cost-benefit analysis (in which the potential harm and potential gain are multiplied by their probabilities and weighed against each other) is no longer a useful paradigm. Even if probabilities are expected to be zero but have a non-zero uncertainty, then a sensitivity analysis that considers the impact of that uncertainty results in infinities as well. The potential harm is so substantial that everything else in the equation ceases to matter. In this case, we must do everything we can to avoid the catastrophe.

## IV. SCIENTIFIC METHODS AND THE PP

How well can we know either the potential consequences of policies or their probabilities? What does science say about uncertainty? To be helpful in policy decisions, science has to encompass not just expectations of potential benefit and harm but also their probability and uncertainty.

Just as the imperative of analysis of decision-making changes when there is infinite harm for a small, non-zero risk, so is there a fundamental change in the ability to apply scientific methods to the evaluation of that harm. This influences the way we evaluate both the possibility of and the risk associated with ruin.

The idea of precaution is the avoidance of adverse consequences. This is qualitatively different from the idea of evidentiary action (from statistics). In the case of the PP, evidence may come too late. The non-naive PP bridges the gap between precaution and evidentiary action using the ability to evaluate the difference between local and global risks.

### A. *Precautionary vs. Evidentiary Action*

Statistical-evidentiary approaches to risk analysis and mitigation count the frequency of past events (robust statistics), or calibrate parameters of statistical distributions to generate probabilities of future events (parametric approach), or both.

Experimental evidentiary methods follow the model of medical trials, computing probabilities of harm from side effects of drugs or interventions by observing the reactions in a variety of animal and human models. Generally they assume that the risk itself (i.e. nature of harm and their probability) is adequately determined by available information. However, the level of risk may be hard to gauge as its probability may be uncertain, and, in the case of potential infinite harm, an uncertainty that allows for a non-zero probability results in infinities so that the problem is ill-defined mathematically.

While evidentiary approaches are often considered to reflect adherence to the scientific method in its purest form, it is apparent that these approaches do not apply to ruin problems. In an evidentiary approach to risk (relying on evidence-based methods), the existence of a risk or harm occurs when we experience that risk or harm. In the case of ruin, by the time evidence comes it will by definition be too late to avoid it. Nothing in the past may predict one fatal event as illustrated in Fig. 2. Thus standard evidence-based approaches cannot work.

More generally, evidentiary action is a framework based upon the quite reasonable expectation that we learn from experience. The idea of evidentiary action is embodied in the kind of learning from experience that is found in how people often react to disasters—after the fact. When a disaster occurs people prepare for the next one, but do not anticipate it in advance. For the case of ruin problems, such behavior guarantees extinction.

### B. *Invalid Empirical Arguments Against Ruin*

In the case of arguments about ruin problems, claims that experience thus far has not provided evidence for ruin, and thus it should not be considered, are not valid.

### C. *Unknowability, Uncertainty and Unpredictability*

It has been shown that the complexity of real world systems limits the ability of empirical observations to determine the outcomes of actions upon them [2]. This means that a certain class of systemic risks will remain inherently unknown. In some classes of complex systems, controlled experiments cannot evaluate all of the possible systemic consequences under real-world conditions. In these circumstances, efforts to provide assurance of the “lack of harm” are insufficiently reliable. This runs counter to both the use of empirical approaches (including controlled experiments) to evaluate risks, and to the expectation that uncertainty can be eliminated by any means.

### D. *Distinguishing Global and Local Risks*

Since there are mathematical limitations to predictability of outcomes in a complex system, the central issue to determine is whether the threat of harm is local (hence globally benign) or carries global consequences. Scientific analysis can robustly determine whether a risk is systemic, i.e. by evaluating the connectivity of the system to propagation of harm, without determining the specifics of such a risk. If the consequences

are systemic, the associated uncertainty of risks must be treated differently than if it is not. In such cases, precautionary action is not based on direct empirical evidence but on analytical approaches based upon the theoretical understanding of the nature of harm. It relies on probability theory without computing probabilities. The essential question is whether or not global harm is possible or not. Theory enables generalizing from experience in order to apply it to new circumstances. In the case of the PP, the existence of a robust way to generalize is essential.

The relevance of the precautionary principle today is greater than in the past, owing to the global connectivity of civilization that makes the spreading of effects to places previously insulated.

## V. FAT TAILS AND FRAGILITY

### A. Thin and Fat Tails

To figure out whether a given decision involves the risk of ruin and thus warrants the use of the PP, we must first understand the relevant underlying probabilistic structures.

There are two classes of probability distributions of events: one in which events are accompanied by well behaved, mild variations (e.g. Gaussian or thin tails), and the other where small probabilities are associated with large variations that have no characteristic scale (e.g. power law or fat tails). Allegorically these are illustrated by Mediocristan and Extremistan (Figs. 1 and 2), the former being typical of human weight distributions, and the latter of human wealth distributions. Given a series of events (a sequence of measurements of weight or wealth), in the case of thin tails the sum is proportional to the average, and in the case of fat tails a sum over them may be entirely dominated by a single one. Thus, while no human being can be heavier than, say, ten average adults (since weight is thin-tailed), a single individual can be richer than the poorest two billion humans (since wealth is fat tailed).

In thin tailed domains (Fig 1) harm comes from the collective effect of many, many events; no event alone can be consequential enough to affect the aggregate. It is practically impossible for a single day to account for 99% of all heart attacks in a given year (the probability is small enough to be practically zero), for an illustration). Statistical distributions that belong to the thin-tailed domain include: Gaussian, Binomial, Bernoulli, Poisson, Gamma, Beta and Exponential.

In fat tailed domains of risk (Fig. 2) harm comes from the largest single event. Examples of relevant statistical distributions include: Pareto, Levy-Stable distributions with infinite variance, Cauchy, and power law distributions, especially with larger exponents.

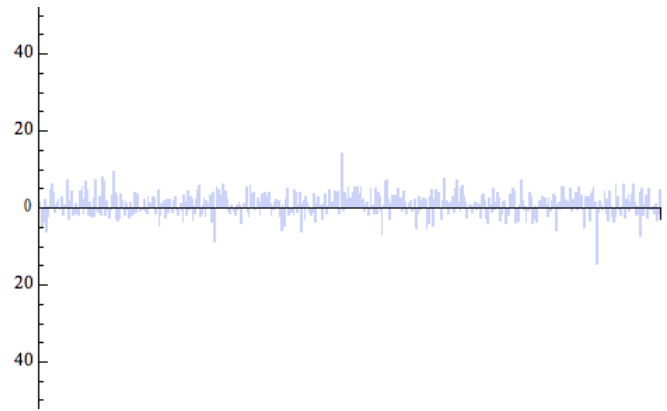


Figure 1: **Thin Tails from Tinkering, Bottom-Up, Evolution.** In nature no individual variation represents a large share of the sum of the variations. Natural boundaries prevent cascading effects from propagating globally. Mass extinctions arise from the rare cases where large impacts (meteorite hits and vulcanism) propagate across the globe through the atmosphere and oceans.

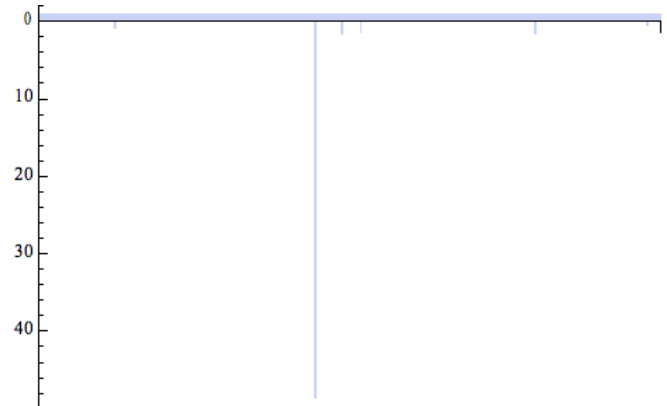


Figure 2: **Fat Tails from a Top-Down, Engineered Design** In human made variations the tightly connected global system implies a single deviation will eventually dominate the sum of their effects. Examples include pandemics, invasive species, financial crises and monoculture.

### B. Why interdependence brings fat tails

When variations lead to independent impacts locally, the aggregate effect of those variations is small according to the central limit theorem, guaranteeing thin-tailed distributions. When there is interdependence, the central limit theorem does not apply, and aggregate variations may become much more severe due to mutual reinforcement. Interdependence arises because of the coupling of behavior in different places. Under these conditions, cascades propagate through the system in a way that can cause large impacts. Whether components are independent or dependent clearly matters to systemic disasters such as pandemics and financial or other crises. Interdependence increases the probability of ruin, ultimately to the point of certainty.

Consider the global financial crash of 2008. As financial firms became increasingly interdependent during the latter part of the 20th century, small fluctuations during periods of calm masked the vulnerability of the system to cascading failures.

Instead of a local shock in an independent area of the system, we experienced a global shock with cascading effects. The crisis of 2008, in addition, illustrates the failure of evidentiary risk management. Since data from the time series beginning in the 1980s exhibited stability, causing the period to be dubbed "the great moderation," it deceived those relying on historical statistical evidence.

## VI. WHAT IS THE RISK OF HARM TO THE EARTH?

At the systemic largest scale on Earth, nature has thin tails, though tails may be fat at smaller length scales or sufficiently long time scales; occasional mass extinctions occur at very long time scales. This is characteristic of a bottom-up, local tinkering design process, where things change primarily locally and only mildly and iteratively on a global scale.

In recent years, it has been shown that natural systems often have fat tail (power law) behaviors associated with the propagation of shocks [3]. This, however, applies to selected systems that do not have barriers that limit those propagations. The earth has an intrinsic heterogeneity of oceans/continents, deserts, mountains, lakes, rivers and climate differences that limit the propagation of variations from one area to another. There are also smaller natural boundaries associated with organism sizes and those of local groups of organisms. Among the largest propagation events we commonly observe are forest fires, but even these are bounded in their impacts compared to a global scale. The various forms of barriers limit the propagation of cascades that enable large scale events.

At longer time scales of millions of years, mass extinctions can achieve a global scale. Connectivity of oceans and the atmosphere enables propagation of impacts, i.e. gas, ash and dust propagating through the atmosphere due to meteor impacts and volcanism, is considered a scenario for these extinction events [4]. The variability associated with mass extinctions can especially be seen in the fossil record of marine animal species; those of plants and land insects are comparatively robust. It is not known to what extent these events are driven extrinsically, by meteor impacts, geological events including volcanos, or cascading events of coupled species extinctions, or combinations of them. The variability associated with mass extinctions, however, indicates that there are fat tail events that can affect the global biosphere. The major extinction events during the past 500 million years occur at intervals of millions of years [5]. While mass extinctions occur, the extent of that vulnerability is driven by both sensitivity to external events and connectivity among ecosystems.

The greatest impact of human beings on this natural system connectivity is through dramatic increases in global transportation. The impact of invasive species and rapid global transmission of diseases demonstrates the role of human activity in connecting previously much more isolated natural systems. The role of transportation and communication in connecting civilization itself is apparent in economic interdependence manifest in cascading financial crises that were not possible even a hundred years ago. The danger we are facing today is that we as a civilization are globally connected, and the fat tail of the distribution of shocks extends globally, to our peril.

Had nature not imposed sufficiently thin-tailed variations in the aggregate or macro level, we would not be here today. A single one of the trillions, perhaps the trillions of trillions, of variations over evolutionary history would have terminated life on the planet. Figures 1 and 2 show the difference between the two separate statistical properties. While tails can be fat for subsystems, nature remains predominantly thin-tailed at the level of the planet [6]. As connectivity increases the risk of extinction increases dramatically and nonlinearly [7].

### A. Risk and Global Interventionism

Currently, global dependencies are manifest in the expressed concerns about policy maker actions that nominally appear to be local in their scope. In just recent months, headlines have been about Russia's involvement in Ukraine, the spread of Ebola in east Africa, expansion of ISIS control into Iraq, ongoing posturing in North Korea and Israeli-Palestinian conflict, among others. These events reflect upon local policy maker decisions that are justifiably viewed as having global repercussions. The connection between local actions and global risks compels widespread concern and global responses to alter or mitigate local actions. In this context, we point out that the broader significance and risk associated with policy actions that impact on global ecological and human survival is the essential point of the PP. Paying attention to the headline events without paying attention to these even larger risks is like being concerned about the wine being served on the Titanic.

## VII. FRAGILITY

The PP applies only to the largest scale impacts due to the inherent fragility of systems that maintain their structure. As the scale of impacts increases the harm increases non-linearly up to the point of destruction.

### A. Fragility as Nonlinear Response

Everything that has survived is necessarily non-linear to harm. If I fall from a height of 10 meters I am injured more than 10 times than if I fell from a height of 1 meter, or more than 1000 times than if I fell from a height of 1 centimeter, hence I am fragile. In general, every additional meter, up to the point of my destruction, hurts me more than the previous one.

Similarly, if I am hit with a big stone I will be harmed a lot more than if I were pelted serially with pebbles of the same total weight.

Everything that is fragile and still in existence (that is, unbroken), will be harmed more by a certain stressor of intensity  $X$  than by  $k$  times a stressor of intensity  $X/k$ , up to the point of breaking. If I were not fragile (susceptible to harm more than linearly), I would be destroyed by accumulated effects of small events, and thus would not survive. This non-linear response is central for everything on planet earth.

This explains the necessity of considering scale when invoking the PP. Polluting in a small way does not warrant the PP because it is essentially less harmful than polluting in large quantities, since harm is non-linear.

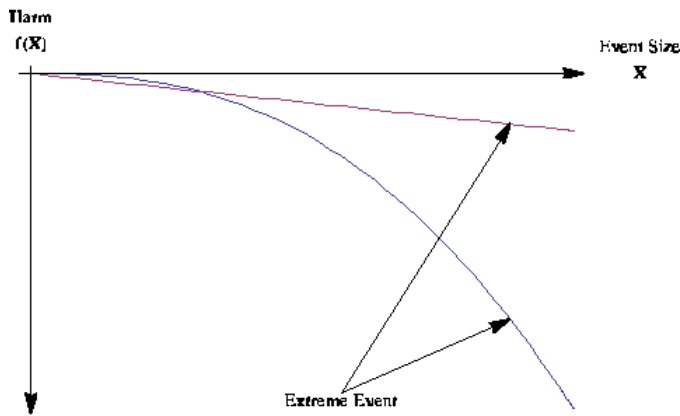


Figure 3: Nonlinear response compared to linear response. The PP should be evoked to prevent impacts that result in complete destruction due to the nonlinear response of natural systems, it is not needed for smaller impacts where risk management methods can be applied.

### B. Why is fragility a general rule?

The statistical structure of stressors is such that small variations are much, much more frequent than large ones. Fragility is intimately connected to the ability to withstand small impacts and recover from them. This ability is what makes a system retain its structure. Every system has a threshold of impact beyond which it will be destroyed, i.e. its structure is not sustained.

Consider a coffee cup sitting on a table: there are millions of recorded earthquakes every year; if the coffee cup were linearly sensitive to earthquakes and accumulated their effects as small deteriorations of its form, it would not persist even for a short time as it would have been broken down due to the accumulated impact of small vibrations. The coffee cup, however, is non-linear to harm, so that the small or remote earthquakes only make it wobble, whereas one large one would break it forever.

This nonlinearity is necessarily present in everything fragile.

Thus, when impacts extend to the size of the system, harm is severely exacerbated by non-linear effects. Small impacts, below a threshold of recovery, do not accumulate for systems that retain their structure. Larger impacts cause irreversible damage. We should be careful, however, of actions that may seem small and local but then lead to systemic consequences.

### C. Fragility, Dose response and the $1/n$ rule

Another area where we see non-linear responses to harm is the dose-response relationship. As the dose of some chemical or stressor increases, the response to it grows non-linearly. Many low-dose exposures do not cause great harm, but a single large-dose can cause irreversible damage to the system, like overdosing on painkillers.

In decision theory, the  $1/n$  heuristic is a simple rule in which an agent invests equally across  $n$  funds (or sources of risk) rather than weighting their investments according to some optimization criterion such as mean-variance or Modern Portfolio Theory (MPT), which dictates some amount of concentration in order to increase the potential payoff. The

$1/n$  heuristic mitigates the risk of suffering ruin due to an error in the model; there is no single asset whose failure can bring down the ship. While the potential upside of the large payoff is dampened, ruin due to an error in prediction is avoided. This heuristic works best when the sources of variations are uncorrelated and, in the presence of correlation or dependence between the various sources of risk, the total exposure needs to be reduced.

Hence, because of non-linearities, it is preferable to diversify our effect on the planet, e.g. distinct types of pollutants, across the broadest number of uncorrelated sources of harm, rather than concentrate them. In this way, we avoid the risk of an unforeseen, disproportionately harmful response to a pollutant deemed "safe" by virtue of responses observed only in relatively small doses.

## VIII. THE LIMITATION OF TOP-DOWN ENGINEERING IN COMPLEX ENVIRONMENTS

In considering the limitations of risk-taking, a key question is whether or not we can analyze the potential outcomes of interventions and, knowing them, identify the associated risks. Can't we just "figure it out?" With such knowledge we can gain assurance that extreme problems such as global destruction will not arise.

Since the same issue arises for any engineering effort, we can ask what is the state-of-the-art of engineering? Does it enable us to know the risks we will encounter? Perhaps it can just determine the actions we should, or should not, take. There is justifiably widespread respect for engineering because it has provided us with innovations ranging from infrastructure to electronics that have become essential to modern life. What is not as well known by the scientific community and the public, is that engineering approaches fail in the face of complex challenges and this failure has been extensively documented by the engineering community itself [8]. The underlying reason for the failure is that complex environments present a wide range of conditions. Which conditions will actually be encountered is uncertain. Engineering approaches involve planning that requires knowledge of the conditions that will be encountered. Planning fails due to the inability to anticipate the many conditions that will arise.

This problem arises particularly for "real-time" systems that are dealing with large amounts of information and have critical functions in which lives are at risk. A classic example is the air traffic control system. An effort to modernize that system by traditional engineering methods cost \$3-6 billion and was abandoned without changing any part of the system because of the inability to evaluate the risks associated with its implementation.

Significantly, the failure of traditional engineering to address complex challenges has led to the adoption of innovation strategies that mirror evolutionary processes, creating platforms and rules that can serve as a basis for safely introducing small incremental changes that are extensively tested in their real world context [8]. This strategy underlies the approach used by highly-successful, modern, engineered-evolved, complex systems ranging from the Internet, to Wikipedia, to iPhone App communities.

## IX. WHY SHOULD GMOs BE BANNED BUT NOT NUCLEAR ENERGY?

As examples that are relevant to the discussion of the different types of strategies, we consider the differences between concerns about nuclear energy and GM crops.

### A. Nuclear energy

Many are justifiably concerned about nuclear energy. It is known that the potential harm due to radiation release, core meltdowns and waste can be large. At the same time, the nature of these risks has been extensively studied, and the risks from local uses of nuclear energy have a scale that is much smaller than global. Thus, even though some uncertainties remain, it is possible to formulate a cost benefit analysis of risks for local decision-making. The large potential harm at a local scale means that decisions about whether, how and how much to use nuclear energy, and what safety measures to use, should be made carefully so that decision makers and the public can rely upon them. Risk management is a very serious matter when potential harm can be large and should not be done casually or superficially. Those who perform the analysis must not only do it carefully, they must have the trust of others that they are doing it carefully. Nevertheless, the known statistical structure of the risks and the absence of global systemic consequences makes the cost benefit analysis meaningful. Decisions can be made in the cost-benefit context—evoking the PP is not appropriate for small amounts of nuclear energy, as the local nature of the risks is not indicative of the circumstances to which the PP applies.

In large quantities, we should worry about an unseen risk from nuclear energy and invoke the PP. In small quantities, it may be OK—how small we should determine by direct analysis, making sure threats never cease to be local.

In addition to the risks from nuclear energy use itself, we must keep in mind the longer term risks associated with the storage of nuclear waste, which are compounded by the extended length of time they remain hazardous. The problems of such longer term “lifecycle” effects is present in many different industries. It arises not just for nuclear energy but also for fossil fuels and other sources of pollution, though the sheer duration of toxicity effects for nuclear waste, enduring for hundreds of thousands of years in some cases, makes this problem particularly intense for nuclear power.

Weighing different options is an important part of policy decision-making. The general idea is that we should limit pollution to small sources, and allow for significant errors in the estimates of harm, even if proponents and experts deem them safe.

### B. GMOs

Genetically Modified Organisms (GMOs) and their risk are currently the subject of debate [9]. Here we argue that they fall squarely under the PP because their risk is systemic. There are two aspects of systemic risk, the widespread impact on the ecosystem and the widespread impact on health.

Ecologically, in addition to intentional cultivation, GMOs have the propensity to spread uncontrollably, and thus their

risks cannot be localized. The cross-breeding of wild-type plants with genetically modified ones prevents their disentangling, leading to irreversible system-wide effects with unknown downsides. The ecological implications of releasing modified organisms into the wild are not tested empirically before release.

Health wise, the modification of crops impacts everyone. Corn, one of the primary GMO crops, is not only eaten fresh or as cereals, but is also a major component of processed foods in the form of high-fructose corn syrup, corn oil, corn starch and corn meal. In 2014 in the US almost 90% of corn and 94% of soybeans are GMO [11]. Foods derived from GMOs are not tested in humans before they are marketed.

The widespread impacts of GMOs on ecologies and human health imply they are in the domain of the PP. This should itself compel policy makers to take extreme caution. However, there is a difficulty for many in understanding the abstract nature of the engagement in risks and imagining the many possible ways that harm can be caused. Thus, we summarize further the nature of the risks that are involved.

### C. GMOs in detail

The systemic global impacts of GMOs arise from a combination of (1) engineered genetic modifications, (2) monoculture—the use of single crops over large areas. Global monoculture itself is of concern for potential global harm, but the evolutionary context of traditional crops provides important assurances. Invasive species are frequently a problem but one might at least argue that the long term evolutionary testing of harmful impacts of organisms on local ecological systems mitigates if not eliminates the largest potential risks. Monoculture in combination with genetic engineering dramatically increases the risks being taken. Instead of a long history of evolutionary selection, these modifications rely not just on naive engineering strategies that do not appropriately consider risk in complex environments, but also explicitly reductionist approaches that ignore unintended consequences and employ very limited empirical testing.

Ironically, at a time when engineering is adopting evolutionary approaches due to the failure of top-down strategies, biologists and agronomists are adopting top-down engineering strategies and taking global systemic risks in introducing organisms into the wild.

One argument in favor of GMOs is that they are no more “unnatural” than the selective farming our ancestors have been doing for generations. In fact, the ideas developed in this paper show that this is not the case. Selective breeding over human history is a process in which change still happens in a bottom-up way, and can be expected to result in a thin-tailed distribution. If there is a mistake, some harmful variation, it will not spread throughout the whole system but end up dying out due to local experience over time. Human experience over generations has chosen the biological organisms that are relatively safe for consumption. There are many that are not, including parts of and varieties of the crops we do cultivate [12]. Introducing rapid changes in organisms is inconsistent with this process. There is a limited rate at which variations can be introduced and selection will be effective [13].

There is no comparison between tinkering with the selective breeding of genetic components of organisms that have previously undergone extensive histories of selection and the top-down engineering of taking a gene from a fish and putting it into a tomato. Saying that such a product is natural misses the process of natural selection by which things become "natural." While there are claims that all organisms include transgenic materials, those genetic transfers that are currently present were subject to selection over long times and survived. The success rate is tiny. Unlike GMOs, in nature there is no immediate replication of mutated organisms to become a large fraction of the organisms of a species. Indeed, any one genetic variation is unlikely to become part of the long term genetic pool of the population. Instead, just like any other genetic variation or mutation, transgenic transfers are subject to competition and selection over many generations before becoming a significant part of the population. A new genetic transfer engineered today is not the same as one that has survived this process of selection.

An example of the effect of transfer of biologically evolved systems to a different context is that of zoonotic diseases. Even though pathogens consume their hosts, they evolve to be less harmful than they would otherwise be. Pathogens that cause highly lethal diseases are selected against because their hosts die before they are able to transmit to others. This is the underlying reason for the greater dangers associated with zoonotic diseases—caused by pathogens that shift from the host that they evolved in to human beings, including HIV, Avian and Swine flu that transferred from monkeys (through chimpanzees), birds and hogs, respectively.

More generally, engineered modifications to ecological systems (through GMOs) are categorically and statistically different from bottom up ones. Bottom-up modifications do not remove the crops from their long term evolutionary context, enabling the push and pull of the ecosystem to locally extinguish harmful mutations. Top-down modifications that bypass this evolutionary pathway unintentionally manipulate large sets of interdependent factors at the same time, with dramatic risks of unintended consequences. They thus result in fat-tailed distributions and place a huge risk on the food system as a whole.

For the impact of GMOs on health, the evaluation of whether the genetic engineering of a particular chemical (protein) into a plant is OK by the FDA is based upon considering limited existing knowledge of risks associated with that protein. The number of ways such an evaluation can be in error is large. The genetic modifications are biologically significant as the purpose is to strongly impact the chemical functions of the plant, modifying its resistance to other chemicals such as herbicides or pesticides, or affecting its own lethality to other organisms—i.e. its antibiotic qualities. The limited existing knowledge generally does not include long term testing of the exposure of people to the added chemical, even in isolation. The evaluation is independent of the ways the protein affects the biochemistry of the plant, including interactions among the various metabolic pathways and regulatory systems—and the impact of the resulting changes in biochemistry on health of consumers. The evaluation is independent of its

farm-ecosystem combination (i.e. pesticide resistant crops are subject to increased use of pesticides, which are subsequently present in the plant in larger concentrations and cannot be washed away). Rather than recognizing the limitations of current understanding, poorly grounded perspectives about the potential damage with unjustified assumptions are being made. Limited empirical validation of both essential aspects of the conceptual framework as well as specific conclusions are being used because testing is recognized to be difficult.

We should exert the precautionary principle here – our non-naive version – because we do not want to discover errors after considerable and irreversible environmental and health damage.

#### *D. Red herring: How about the risk of famine without GMOs?*

An argument used by those who advocate for GMOs is that they will reduce the hunger in the world. Invoking the risk of famine as an alternative to GMOs is a deceitful strategy, no different from urging people to play Russian roulette in order to get out of poverty.

The evocation of famine also prevents clear thinking about not just GMOs but also about global hunger. The idea that GMO crops will help avert famine ignores evidence that the problem of global hunger is due to poor economic and agricultural policies. Those who care about the supply of food should advocate for an immediate impact on the problem by reducing the amount of corn used for ethanol in the US, which burns food for fuel consuming over 40% of the US crop that could provide enough food to feed 2/3 of a billion people [14].

One of the most extensively debated cases for GMOs is a variety of rice—"golden rice"—to which has been added a precursor of vitamin A as a potential means to alleviate this nutritional deficiency, which is a key medical condition affecting impoverished populations. Since there are alternatives, including traditional vitamin fortification, one approach is to apply a cost benefit analysis comparing these approaches. Counter to this approach stands both the largely unknown risks associated with the introduction of GMOs, and the need and opportunities for more systemic interventions to alleviate not just malnutrition but poverty and hunger worldwide. While great attention should be placed on immediate needs, neglecting the larger scale risks is unreasonable [10]. Here science should adopt an unyielding rigor for both health benefit and risk assessment, including careful application of the PP. Absent such rigor, advocacy by the scientific community not only fails to be scientific, but also becomes subject to challenge for short term interests, not much different from corporate endorsers. Thus, cutting corners on tests, including tests without adequate consent or approvals performed on Chinese children [15], undermines scientific claims to humanitarian ideals. Given the promotion of "golden rice" by the agribusiness that also promote biofuels, their interest in humanitarian impacts versus profits gained through wider acceptance of GMO technology can be legitimately questioned [16].

#### *E. GMOs in summary*

In contrast to nuclear energy (which, as discussed in section IX-A above, may or may not fall under the PP, depending on



how and where (how widely) it is implemented), Genetically Modified Organisms, GMOs, fall squarely under the PP because of their systemic risk. The understanding of the risks is very limited and the scope of the impacts are global both due to engineering approach replacing an evolutionary approach, and due to the use of monoculture.

Labeling the GMO approach “scientific” betrays a very poor—indeed warped—understanding of probabilistic payoffs and risk management. A lack of observations of explicit harm does not show absence of hidden risks. Current models of complex systems only contain the subset of reality that is accessible to the scientist. Nature is much richer than any model of it. To expose an entire system to something whose potential harm is not understood because extant models do not predict a negative outcome is not justifiable; the relevant variables may not have been adequately identified.

Given the limited oversight that is taking place on GMO introductions in the US, and the global impact of those introductions, we are precisely in the regime of the ruin problem. A rational consumer should say: We do not wish to pay—or have our descendants pay—for errors made by executives of Monsanto, who are financially incentivized to focus on quarterly profits rather than long term global impacts. We should exert the precautionary principle—our non-naive version—simply because we otherwise will discover errors with large impacts only after considerable damage.

## X. PRECAUTION AS POLICY AND NAIVE INTERVENTION

When there is a risk of ruin, obstructionism and policy inaction are important strategies, impeding the rapid headlong experimentation with global ruin by those with short-term, self-centered incentives and perspectives. Two approaches for policy action are well justified. In the first, actions that avoid the inherent sensitivity of the system to propagation of harm can be used to free the system to enable local decision-making and exploration with only local harm. This involves introducing boundaries, barriers and separations that inhibit propagation of shocks, preventing ruin for overly connected systems. In the second, where such boundaries don’t exist or cannot be introduced due to other effects, there is a need for actions that are adequately evaluated as to their global harm. Scientific analysis of such actions, meticulously validated, is needed to prevent small risks from causing ruin.

What is not justified, and dangerous, are actions that are intended to prevent harm by additional intervention. The reason is that indirect effects are likely to create precisely the risks that one is intending to avoid.

When existing risks are perceived as having the potential for ruin, it may be assumed that any preventive measure is justified. There are at least two problems with such a perspective. First, localized harm is often mistaken for ruin, and the PP is wrongly invoked where risk management techniques should be employed. When a risk is not systemic, overreaction will typically cause more harm than benefits, like undergoing dangerous surgery to remove a benign growth. Second, even if the threat of ruin is real, taking specific (positive) action in order to ward off the perceived threat may introduce new

systemic risks. It is often wiser to reduce or remove activity that is generating or supporting the threat and allow natural variations to play out in localized ways.

Preventive action should be limited to correcting situations by removing threats *via negativa* in order to bring them back in line with a statistical structure that avoids ruin. It is often better to remove structure or allow natural variation to take place rather than to *add* something additional to the system.

When one takes the opposite approach, taking specific action designed to diminish some perceived threat, one is almost guaranteed to induce unforeseen consequences. Even when there appears to be a direct link from a specific action to a specific preventive outcome, the web of causality extends in complex ways with consequences that are far from the intended goal. These unintended consequences may generate new vulnerabilities or strengthen the harm one is hoping to diminish. Thus, when possible, limiting fragilizing dependencies is better than imposing additional structure that increases the fragility of the system as a whole.

## XI. FALLACIOUS ARGUMENTS AGAINST PP

In this section we respond to a variety of arguments that have been made against the PP.

### A. *Crossing the road (the paralysis fallacy)*

Many have countered the invocation of the PP with “nothing is ever totally safe.” “I take risks crossing the road every day, so according to you I should stay home in a state of paralysis.” The answer is that we don’t cross the street blindfolded, we use sensory information to mitigate risks and reduce exposure to extreme shocks.

Even more importantly in the context of the PP, the probability distribution of death from road accidents at the population level is thin-tailed; I do not incur the risk of generalized human extinction by crossing the street—a human life is bounded in duration and its unavoidable termination is an inherent part of the bio-social system [17]. The error of my crossing the street at the wrong time and meeting an untimely demise in general does not cause others to do the same; the error does not spread. If anything, one might expect the opposite effect, that others in the system benefit from my mistake by adapting their behavior to avoid exposing themselves to similar risks. Equating risks a person takes with his or her own life with risking the existence of civilization is an inappropriate ego trip. In fact, the very idea of the PP is to avoid such a frivolous focus.

The paralysis argument is often used to present the PP as incompatible with progress. This is untrue: tinkering, bottom-up progress where mistakes are bounded is *how* progress has taken place in history. The non-naive PP simply asserts that the risks we take as we innovate must not extend to the entire system; local failure serves as information for improvement. Global failure does not.

### B. *The Psychology of Risk and Thick Tailed Distributions*

One concern about the utility of the PP is that its evocation may become commonplace because of risk aversion. Is it true

that people overreact to small probabilities and the PP would feed into human biases? While we have carefully identified the scope of the domain of applicability of the PP, it is also helpful to review the evidence of risk aversion, which we find not to be based upon sound studies.

Certain empirical studies appear to support the existence of a bias toward risk aversion, claiming evidence that people choose to avoid risks that are beneficial, inconsistent with cost-benefit analyses. The relevant experiments ask people questions about single probability events, showing that people overreact to small probabilities. However, those researchers failed to include the consequences of the associated events which humans underestimate. Thus, this empirical strategy as a way of identifying effectiveness of response to risk is fundamentally flawed [18].

The proper consideration of risk involves both probability and consequence, which should be multiplied together. Consequences in many domains have thick tails, i.e. much larger consequences can arise than are considered in traditional statistical approaches. Overreacting to small probabilities is not irrational when the effect is large, as the product of probability and harm is larger than expected from the traditional treatment of probability distributions.

### C. The Loch Ness fallacy

Many have countered that we have no evidence that the Loch Ness monster doesn't exist, and, to take the argument of *evidence of absence* being different from *absence of evidence*, we should act as if the Loch Ness monster existed. The argument is a corruption of the absence of evidence problem and certainly not part of the PP.

The relevant question is whether the existence of the Loch Ness monster has implications for decisions about actions that are being taken. We are not considering a decision to swim in the Loch Ness. If the Loch Ness monster did exist, there would still be no reason to invoke the PP, as the harm he might cause is limited in scope to Loch Ness itself, and does not present the risk of ruin.

### D. The fallacy of misusing the naturalistic fallacy

Some people invoke "the naturalistic fallacy," a philosophical concept that is limited to the moral domain. According to this critique, we should not claim that natural things are necessarily good; human innovation can be equally valid. We do not claim to use nature to derive a notion of how things "ought" to be organized. Rather, as scientists, we respect nature for the extent of its experimentation. The high level of statistical significance given by a very large sample cannot be ignored. Nature may not have arrived at the best solution to a problem we consider important, but there is reason to believe that it is smarter than our technology based only on statistical significance.

The question about what kinds of systems *work* (as demonstrated by nature) is different than the question about what working systems ought to do. We can take a lesson from nature—and time—about what kinds of organizations are robust against, or even benefit from, shocks, and in that

sense systems should be structured in ways that allow them to function. Conversely, we cannot derive the structure of a functioning system from what we believe the outcomes *ought* to be.

To take one example, Cass Sunstein—who has written an article critical of the PP [19]—claims that there is a "false belief that nature is benign." However, his conceptual discussion fails to distinguish between thin and fat tails, local harm and global ruin. The method of analysis misses both the statistical significance of nature and the fact that it is not necessary to believe in the perfection of nature, or in its "benign" attributes, but rather in its track record, its sheer statistical power as a risk evaluator and as a risk manager in avoiding ruin.

### E. The "Butterfly in China" fallacy

The statement "if I move my finger to scratch my nose, by the butterfly-in-China effect, owing to non-linearities, I may terminate life on earth," is known to be flawed. The explanation is not widely understood. The fundamental reason arises because of the existence of a wide range in levels of predictability and the presence of a large number of fine scale degrees of freedom for every large scale one [20]. Thus, the traditional deterministic chaos, for which the butterfly effect was named, applies specifically to low dimensional systems with a few variables in a particular regime. High dimensional systems, like the earth, have large numbers of fine scale variables for every large scale one. Thus, it is apparent that not all butterfly wing flaps can cause hurricanes. It is not clear that any one of them can, and, if small perturbations can influence large scale events, it happens only under specific conditions where amplification occurs.

Empirically, our thesis rebuts the butterfly fallacy with the argument that, in the aggregate, nature has experienced trillions of small variations and yet it survives. Therefore, we know that the effects of scratching one's nose fall into the thin tailed domain and thus do not warrant the precautionary principle.

As described previously, barriers in natural systems lead to subsystems having a high-degree of independence. Understanding how modern systems with a high-degree of connectivity have cascading effects is essential for understanding when it is and isn't appropriate to use the PP.

### F. The potato fallacy

Many species were abruptly introduced into the Old World starting in the 16th Century that did not cause environmental disasters (perhaps aside from diseases affecting Native Americans). Some use this observation in defense of GMOs. However, the argument is fallacious at two levels:

First, by the fragility argument, potatoes, tomatoes and similar "New World" goods were developed locally through progressive, bottom-up tinkering in a complex system in the context of its interactions with its environment. Had they had an impact on the environment, it would have caused adverse consequences that would have prevented their continual spread.

Second, a counterexample is not evidence in the risk domain, particularly when the evidence is that taking a similar action previously did not lead to ruin. Lack of ruin due to several or even many trials does not indicate safety from ruin in the next one. This is also the Russian roulette fallacy, detailed below.

#### *G. The Russian roulette fallacy (the counterexamples in the risk domain)*

The potato example, assuming potatoes had not been generated top-down by some engineers, would still not be sufficient. Nobody says "look, the other day there was no war, so we don't need an army," as we know better in real-life domains. Nobody argues that a giant Russian roulette with many barrels is "safe" and a great money making opportunity because it didn't blow up someone's brains last time.

There are many reasons a previous action may not have led to ruin while still having the *potential* to do so. If you attempt to cross the street with a blindfold and earmuffs on, you may make it across, but this is not evidence that such an action carries no risk.

More generally, one needs a large sample for claims of *absence of risk* in the presence of a small probability of ruin, while a single " $n = 1$ " example would be sufficient to counter the claims of safety—this is the Black Swan argument [?]. Simply put, systemic modifications require a very long history in order for the evidence of lack of harm to carry any weight.

#### *H. The Carpenter fallacy*

Risk managers skeptical of the understanding of risk of biological processes, such as GMOs, by the experts are sometimes asked "are you a biologist?". But nobody asks a probabilist dealing with roulette sequences if he is a carpenter. To understand the gambler's ruin problem by roulette betting, we know to ask a probabilist, not a carpenter. No amount of expertise in carpentry can replace rigor in understanding the properties of long sequences of small probability bets. Likewise, no amount of expertise in the details of biological processes can be a substitute for probabilistic rigor.

Moreover, the track record of the experts in understanding biological and medical risks has been extremely poor, and we need policies to be robust to their miscalculations. The "expert problem" is manifest in a very poor record historically of risks taken with innovations in biological products, ranging from biofuels to transfat to nicotine, etc. Consider the recent major drug recalls such as Thalidomide, Fen-Phen, Tylenol and Vioxx—all of these show chronic blindness on the part of the specialist to large scale risks of Black Swan events. Yet most of these risks were local and not systemic (with the exception of biofuel impacts on global hunger and social unrest): with the systemic the recall happens too late, which is why we need this strong version of the PP.

#### *I. The technological salvation fallacy*

Iatrogenics is harm done by a healer despite positive intentions, see Appendix A for a list of innovations in care that

have extensive documentation of adverse consequences. Each of these underwent best practices testing that did not reveal the iatrogenic consequences prior to widespread application. The controlled tests that are used to evaluate innovations for potential harm cannot replicate the large number of conditions in which interventions are applied in the real world. Adverse consequences are exposed only by extensive experience with the combinatorial number of real world conditions. Natural, i.e. evolutionary, selection implements as a strategy the use of selection of lack of harm under such conditions in a way that bounds the consequences because the number of replicates is increased only gradually during the process in which success is determined. In contrast, traditional engineering of technological solutions does not. Thus, the more technological a solution to a current problem—the more it departs from solutions that have undergone evolutionary selection—the more exposed one becomes to iatrogenics owing to combinatorial branching of conditions with adverse consequences.

Our concern here isn't mild iatrogenics, but the systemic case.

#### *J. The pathologization fallacy*

Today many mathematical or conceptual models that are claimed to be rigorous are based upon unvalidated and incorrect assumptions. Such models are rational in the sense that they are logically derived from their assumptions, except that it is the modeler who is using an incomplete representation of the reality. Often the modelers are not familiar with the dynamics of complex systems or use Gaussian statistical methods that do not take into account fat-tails and make inferences that would not be acceptable under different classes of probability distributions. Many biases, such as the ones used by Cass Sunstein (mentioned above), about the overestimation of the probabilities of rare events in fact correspond to the testers using a bad probability model that is thin-tailed. See Ref. [6] for a deeper discussion.

It has become popular to claim irrationality for GMO and other skepticism on the part of the general public—not realizing that there is in fact an "expert problem" and such skepticism is healthy and even necessary for survival. For instance, in *The Rational Animal*, the authors pathologize people for not accepting GMOs although "the World Health Organization has never found evidence of ill effects," a standard confusion of evidence of absence and absence of evidence. Such pathologizing is similar to behavioral researchers labeling hyperbolic discounting as "irrational" when in fact it is largely the researcher who has a very narrow model and richer models make the "irrationality" go away.

These researchers fail to understand that humans may have precautionary principles against systemic risks, and can be skeptical of the untested consequences of policies for deeply rational reasons.

## XII. CONCLUSIONS

This formalization of the two different types of uncertainty about risk (local and systemic) makes clear when the precautionary principle is, and when it isn't, appropriate. The

examples of GMOs and nuclear energy help to elucidate the application of these ideas. We hope this will help decision makers to avoid ruin in the future.

#### ACKNOWLEDGMENTS

Gloria Origgi, Maya Bialik, David Boxenhorn, Jessica Woolley, ...

#### CONFLICTS OF INTEREST

One of the authors (Taleb) reports having received monetary compensation for lecturing on risk management and Black Swan risks by the Institute of Nuclear Power Operations, INPO, the main association in the United States, in 2011, in the wake of the Fukushima accident.

#### REFERENCES

- [1] Asmussen, S., & Albrecher, H., 2010, *Ruin probabilities* (Vol. 14). World Scientific.
- [2] Bar-Yam, Y., 2013, *The Limits of Phenomenology: From Behaviorism to Drug Testing and Engineering Design*, arXiv 1308.3094
- [3] Bak, P., 2009, *How nature works*. Copernicus.
- [4] Schulte, P., Alegret, L., Arenillas, I., Arz, J. A., Barton, P. J., Bown, P. R., ... & Willumsen, P. S., 2010. The Chicxulub asteroid impact and mass extinction at the Cretaceous-Paleogene boundary. *Science*, 327(5970), 1214-1218.
- [5] Alroy, J., 2008. Dynamics of origination and extinction in the marine fossil record. *Proceedings of the National Academy of Sciences*, 105(Supplement 1), 11536-11542.
- [6] Taleb, N.N., 2014, *Silent Risk: Lectures on Fat Tails, (Anti)Fragility, and Asymmetric Exposures*, SSRN
- [7] Rauch, E.M. and Y. Bar-Yam, 2006, *Long-range interactions and evolutionary stability in a predator-prey system*, *Physical Review E* 73, 020903
- [8] Bar-Yam, Y., 2003, *When Systems Engineering Fails — Toward Complex Systems Engineering in International Conference on Systems, Man & Cybernetics Vol. 2*, IEEE Press, Piscataway, NJ, 2003, pp. 2021- 2028.
- [9] Thompson, P.B. (Ed.), 2007. *Food biotechnology in ethical perspective* (Vol. 10). Springer.
- [10] Read, R., Hutchinson, P., 2014. *What is Wrong With GM Food?*, *Philosophers' Magag*.
- [11] *Recent Trends in GE Adoption, Adoption of Genetically Engineered Crops in the U.S.*, USDA Economics Research Service,
- [12] See e.g. *List of poisonous plants*, Wikipedia
- [13] Nowak, M., Schuster, P., 1989. Error thresholds of replication in finite populations mutation frequencies and the onset of Muller's ratchet. *Journal of Theoretical Biology*, 137, 375-395.
- [14] Albino, D.K., Bertrand, K.Z., Bar-Yam, Y., 2012, *Food for fuel: The price of ethanol*. arXiv:1210.6080.
- [15] Qiu, J., 2012, *China sacks officials over Golden Rice controversy*. *Nature News*, 10.
- [16] Harmon, A., 2013, *Golden Rice: Lifesaver?*
- [17] Taleb, N.N., 2007, *Black swans and the domains of statistics*. *The American Statistician*, 61, 198-200.
- [18] Taleb, N.N. and Tetlock, P.E., 2014, *On the Difference between Binary Prediction and True Exposure with Implications for Forecasting Tournaments and Decision Making Research* <http://dx.doi.org/10.2139/ssrn.2284964>
- [19] Sunstein, C.R., *Beyond the Precautionary Principle* (January 2003). U Chicago Law & Economics, Olin Working Paper No. 149; U of Chicago, Public Law Working Paper No. 38.
- [20] Bar-Yam, Y., *Complex Systems: The Science of Prediction*,

<b>Medical Intervention</b>	<b>Intended Effects</b>	<b>Unintended Effects</b>
Rofecoxib (Vioxx, Ceoxx, Ceeoxx)	relieve osteoarthritis, dysmenorrhoea	myocardial infarctions [10]
Thalidomide (Immunoprin, Talidex, Talizer, Thalomid)	sedative	severe birth defects [13]
Fen-phen (Pondimin)	weight loss	valvular heart disease, pulmonary hypertension [7]
Diethylstilbestrol (Distilbene, Stilbestrol, Stilbetin)	reduce miscarriage	cancerous tumors in daughters exposed in utero [9]
Cerivastatin (Baycol, Lipobay)	lower cholesterol, reduce cardiovascular disease	Rhabdomyolysis leading to renal failure [8]
lobotomy	improve mental disorder	loss of personality, intellect [5]
Troglitazone (Rezulin, Resulin, Romozin, Noscald)	antidiabetic, antiinflammatory	drug-induced hepatitis [22]
Terfenadine (Seldane, Triludan, Teldane)	antihistamine	cardiac arrhythmia [18]
Phenylpropanolamine (Accutrim)	appetite suppressant, stimulant, decongestant	increased stroke [12]
hospitalization	patient treatment and monitoring	nosocomial infection; medication errors [11]
antibiotics	clear bacterial infections	treatment-resistant bacteria [2]
antidepressants	relieve depression	increased suicide risk [6]
Encainide (Enkaid), flecainide (Tambocor)	reduced arrhythmia	increased mortality [3]
Acetaminophen (Tylenol)	pain relief	liver damage [14]
coronary angioplasty	increased blood flow	increased risk of death/myocardial infarction [16]
cosmetic surgery	improved aesthetics	infection, death, deformity, other malfunction [15]
obsessive hygiene	keeping safe from 'germs'	autoimmune disorders [19]
ear-tubes	otitis media with effusion	tympanosclerosis [21]

Table II: Examples of iatrogenics in the medical field. The upper portion of the table shows medications and treatments whose use has been significantly reduced or completely discontinued due to their undesired effects (which were discovered only after significant damage had been done). The lower portion of the table lists examples where unintended side effects are significant but treatment continues to be applied due to expected benefits.

APPENDIX A  
A SAMPLE OF IATROGENICS, ERRORS DISCOVERED TOO LATE

## APPENDIX B

### DEFINITION OF FAT TAILS AND DISTINCTION BETWEEN MEDIOCRISTAN AND EXTREMISTAN

Different classes of distributions range between extreme thin-tailed (Bernoulli) and extreme fat tailed [6]: 1) Compact but not degenerate support, 2) Subgaussian, 3) Gaussian, 4) subexponential, 5) Power Laws with exponent greater than 3, 6) Power Laws with Exponent less than or equal to 3 and greater than 2, 7) Power law tails with exponents less than or equal to 2.

But there is a clear cutpoint between Mediocristan and Extremistan which we put here, not at the powerlaw, or the "infinite variance" as is often made in the literature, but at the exponential class which has the following central property:

Let  $X = (x_i)_{1 \leq i \leq n}$  be realizations of i.i.d. random variables in  $\mathbb{R}^+$ , with cumulative distribution function  $F$ ; then by the Teugels (1975)[20] and Pitman [17] (1980) definition:

$$\lim_{x \rightarrow \infty} \frac{1 - F^2(x)}{1 - F(x)} = 2$$

where  $F^2$  is the convolution of  $x$  with itself.

Note that  $X$  does not have to be limited to  $\mathbb{R}^+$ ; we can split the variables in positive and negative domain for the analysis, and replace  $x \rightarrow \infty$  with  $x \rightarrow -\infty$

More generally, distributions are called subexponential when the exceedance probability declines more slowly in the tails than the exponential.

For a one-tailed random variable,

a)  $\lim_{x \rightarrow \infty} \frac{P_{X > \Sigma x}}{P_{X > x}} = n$ , (Christyakov, 1964, [1]), which is equivalent to

b)  $\lim_{x \rightarrow \infty} \frac{P_{X > \Sigma x}}{P_{(X > \max(x))}} = 1$ , (Embrecht and Goldie, 1980,[4] ).

The sum is of the same order as the maximum (positive) value, another way of saying that the tails play a large role.

Clearly  $F$  has to have no exponential moment:

$$\int_0^{\infty} e^{\epsilon x} dF(x) = \infty$$

for all  $\epsilon > 0$ .

## APPENDIX C

### MATHEMATICAL DERIVATIONS OF FRAGILITY

The following offers a formal definition of fragility as "vega", negative expected response from uncertainty. It also shows why this is necessarily linked to accelerated response, how "size matters". The derivations explain, among other things"

- How spreading risks are dangerous compared to limited one; the derivations show the notion of risk spreading as a non-concave response.
- Why error is a problem in the presence of nonlinearity.
- Why polluting "a little" is qualitatively different from polluting "a lot".

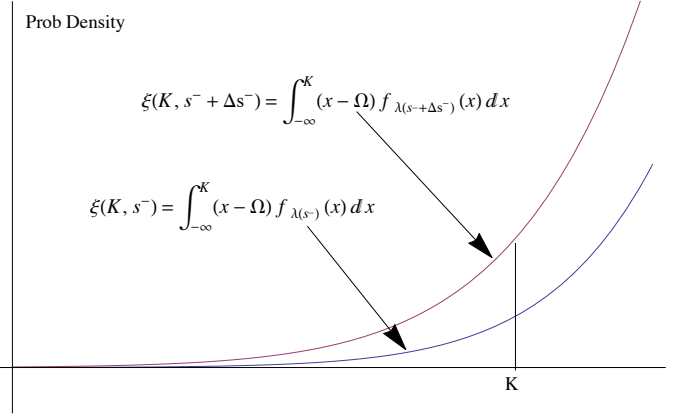


Figure 4: A definition of fragility as left tail-vega sensitivity; the figure shows the effect of the perturbation of the lower semi-deviation  $s^-$  on the tail integral  $\xi$  of  $(x - \Omega)$  below  $K$ ,  $\Omega$  being a centering constant. Our detection of fragility does not require the specification of  $f$  the probability distribution.

- Eventually, why fat tails arise from accelerating response.

**Intrinsic and Inherited Fragility:** Our definition of fragility is two-fold. First, of concern is the intrinsic fragility, the shape of the probability distribution of a variable and its sensitivity to  $s^-$ , a parameter controlling the left side of its own distribution. But we do not often directly observe the statistical distribution of objects, and, if we did, it would be difficult to measure their tail-vega sensitivity. Nor do we need to specify such distribution: we can gauge the response of a given object to the volatility of an external stressor that affects it. For instance, an option is usually analyzed with respect to the scale of the distribution of the "underlying" security, not its own; the fragility of a coffee cup is determined as a response to a given source of randomness or stress; that of a house with respect of, among other sources, the distribution of earthquakes. This fragility coming from the effect of the underlying is called inherited fragility. The transfer function, which we present next, allows us to assess the effect, increase or decrease in fragility, coming from changes in the underlying source of stress.

**Transfer Function:** A nonlinear exposure to a certain source of randomness maps into tail-vega sensitivity (hence fragility). We prove that

Inherited Fragility  $\Leftrightarrow$  Concavity in exposure on the left side of the distribution

and build  $H$ , a transfer function giving an exact mapping of tail vega sensitivity to the second derivative of a function. The transfer function will allow us to probe parts of the distribution and generate a fragility-detection heuristic covering both physical fragility and model error.

Taking  $z$  as a stress level and  $\Pi(z)$  the harm function, it suffices to see that, with  $n > 1$ ,

$$\Pi(nz) < n\Pi(z) \text{ for all } 0 < nz < Z^*$$

where  $Z^*$  is the level (not necessarily specified) at which the item is broken. Such inequality leads to  $\Pi(z)$  having a

negative second derivative at the initial value  $z$ .

So if a coffee cup is less harmed by  $n$  times a stressor of intensity  $Z$  than once a stressor of  $nZ$ , then harm (as a negative function) needs to be concave to stressors up to the point of breaking; such stricture is imposed by the structure of survival probabilities and the distribution of harmful events, and has nothing to do with subjective utility or some other figments. Just as with a large stone hurting more than the equivalent weight in pebbles, if, for a human, jumping one millimeter caused an exact linear fraction of the damage of, say, jumping to the ground from thirty feet, then the person would be already dead from cumulative harm. Actually a simple computation shows that he would have expired within hours from touching objects or pacing in his living room, given the multitude of such stressors and their total effect. The fragility that comes from linearity is immediately visible, so we rule it out because the object would be already broken and the person already dead. The relative frequency of ordinary events compared to extreme events is the determinant. In the financial markets, there are at least ten thousand times more events of 0.1% deviations than events of 10%. There are close to 8,000 micro-earthquakes daily on planet earth, that is, those below 2 on the Richter scale —about 3 million a year. These are totally harmless, and, with 3 million per year, you would need them to be so. But shocks of intensity 6 and higher on the scale make the newspapers. Accordingly, we are necessarily immune to the *cumulative* effect of small deviations, or shocks of very small magnitude, which implies that these affect us disproportionately less (that is, nonlinearly less) than larger ones.

Model error is not necessarily mean preserving.  $s^-$ , the lower absolute semi-deviation does not just express changes in overall dispersion in the distribution, such as for instance the “scaling” case, but also changes in the mean, i.e. when the upper semi-deviation from  $\Omega$  to infinity is invariant, or even decline in a compensatory manner to make the overall mean absolute deviation unchanged. This would be the case when we shift the distribution instead of rescaling it. Thus the same vega-sensitivity can also express sensitivity to a stressor (dose increase) in medicine or other fields in its effect on either tail. Thus  $s^-(l)$  will allow us to express the sensitivity to the “disorder cluster” in *Antifragile*: i) uncertainty, ii) variability, iii) imperfect, incomplete knowledge, iv) chance, v) chaos, vi) volatility, vii) disorder, viii) entropy, ix) time, x) the unknown, xi) randomness, xii) turmoil, xiii) stressor, xiv) error, xv) dispersion of outcomes.

#### A. Tail Sensitivity to Uncertainty

We construct a measure of “vega”, that is, the sensitivity to uncertainty, in the left tails of the distribution that depends on the variations of  $s$  the semi-deviation below a certain level  $W$ , chosen in the  $L^1$  norm in order to ensure its existence under “fat tailed” distributions with finite first semi-moment. In fact  $s$  would exist as a measure even in the case of undefined moments to the right side of  $W$ .

Let  $X$  be a random variable, the distribution of which is one among a one-parameter family of pdf  $f_\lambda$ ,  $\lambda \in \mathbb{I} \subset \mathbb{R}$ . We

consider a fixed reference value  $\Omega$  and, from this reference, the left-semi-absolute deviation:

$$s^-(\lambda) = \int_{-\infty}^{\Omega} (\Omega - x) f_\lambda(x) dx \quad (1)$$

We assume that  $\lambda \rightarrow s^-(\lambda)$  is continuous, strictly increasing and spans the whole range  $\mathbb{R}_+ = [0, +\infty)$ , so that we may use the left-semi-absolute deviation  $s^-$  as a parameter by considering the inverse function  $\lambda(s) : \mathbb{R}_+ \rightarrow \mathbb{I}$ , defined by  $s^-(\lambda(s)) = s$  for  $s \in \mathbb{R}_+$ .

This condition is for instance satisfied if, for any given  $x < \Omega$ , the probability is a continuous and increasing function of  $\lambda$ . Indeed, denoting

$$F_\lambda(x) = P_{f_\lambda}(X < x) = \int_{-\infty}^x f_\lambda(t) dt, \quad (2)$$

an integration by parts yields:

$$s^-(\lambda) = \int_{-\infty}^{\Omega} F_\lambda(x) dx$$

This is the case when  $\lambda$  is a scaling parameter, i.e.,  $X \sim \Omega + \lambda(X_1 - \Omega)$  indeed one has in this case

$$F_\lambda(x) = F_1\left(\Omega + \frac{x - \Omega}{\lambda}\right),$$

$$\frac{\partial F_\lambda}{\partial \lambda}(x) = \frac{\Omega - x}{\lambda^2} f_\lambda(x) \text{ and } s^-(\lambda) = \lambda s^-(1).$$

It is also the case when  $\lambda$  is a shifting parameter, i.e.  $X \sim X_0 - \lambda$ , indeed, in this case  $F_\lambda(x) = F_0(x + \lambda)$  and  $\frac{\partial s^-}{\partial \lambda}(x) = F_\lambda(\Omega)$ .

For  $K < \Omega$  and  $s \in \mathbb{R}^+$ , let:

$$\xi(K, s^-) = \int_{-\infty}^K (\Omega - x) f_{\lambda(s^-)}(x) dx \quad (3)$$

In particular,  $\xi(\Omega, s^-) = s^-$ . We assume, in a first step, that the function  $\xi(K, s^-)$  is differentiable on  $(-\infty, \Omega] \times \mathbb{R}_+$ . The *K-left-tail-vega sensitivity* of  $X$  at stress level  $K < \Omega$  and deviation level  $s^- > 0$  for the pdf  $f_\lambda$  is:

$$V(X, f_\lambda, K, s^-) = \frac{\partial \xi}{\partial s^-}(K, s^-) = \left( \int_{-\infty}^{\Omega} (\Omega - x) \frac{\partial f_\lambda}{\partial \lambda} dx \right) \left( \frac{ds^-}{d\lambda} \right)^{-1} \quad (4)$$

As in the many practical instances where threshold effects are involved, it may occur that  $\xi$  does not depend smoothly on  $s^-$ . We therefore also define a *finite difference* version of the *vega-sensitivity* as follows:

$$V(X, f_\lambda, K, s^-) = \frac{1}{2\Delta s} (\xi(K, s^- + \Delta s) - \xi(K, s^- - \Delta s)) = \int_{-\infty}^K (\Omega - x) \frac{f_\lambda(s^- + \Delta s)(x) - f_\lambda(s^- - \Delta s)(x)}{2\Delta s} dx \quad (5)$$

Hence omitting the input  $\Delta s$  implicitly assumes that  $\Delta s \rightarrow 0$ .

Note that  $\xi(K, s^-) = -\mathbb{E}(X|X < K) \mathbb{P}_{f_\lambda}(X < K)$ . It can

be decomposed into two parts:

$$\xi(K, s^-(\lambda)) = (\Omega - K)F_\lambda(K) + P_\lambda(K) \quad (6)$$

$$P_\lambda(K) = \int_{-\infty}^K (K - x)f_\lambda(x) dx \quad (7)$$

Where the first part  $(\Omega - K)F_\lambda(K)$  is proportional to the probability of the variable being below the stress level  $K$  and the second part  $P_\lambda(K)$  is the expectation of the amount by which  $X$  is below  $K$  (counting 0 when it is not). Making a parallel with financial options, while  $s^-(\lambda)$  is a ‘‘put at-the-money’’,  $\xi(K, s^-)$  is the sum of a put struck at  $K$  and a digital put also struck at  $K$  with amount  $\Omega - K$ ; it can equivalently be seen as a put struck at  $\Omega$  with a down-and-in European barrier at  $K$ .

Letting  $\lambda = \lambda(s^-)$  and integrating by part yields

$$\xi(K, s^-(\lambda)) = (\Omega - K)F_\lambda(K) + \int_{-\infty}^K F_\lambda(x)dx = \int_{-\infty}^\Omega F_\lambda^K(x) dx \quad (8)$$

Where  $F_\lambda^K(x) = F_\lambda(\min(x, K)) = \min(F_\lambda(x), F_\lambda(K))$ , so that

$$V(X, f_\lambda, K, s^-) = \frac{\partial \xi}{\partial s}(K, s^-) = \frac{\int_{-\infty}^\Omega \frac{\partial F_\lambda^K}{\partial \lambda}(x) dx}{\int_{-\infty}^\Omega \frac{\partial F_\lambda}{\partial \lambda}(x) dx} \quad (9)$$

For finite differences

$$V(X, f_\lambda, K, s^-, \Delta s) = \frac{1}{2\Delta s} \int_{-\infty}^\Omega \Delta F_{\lambda, \Delta s}^K(x) dx \quad (10)$$

Where  $\lambda_s^+$  and  $\lambda_s^-$  are such that  $s(\lambda_s^+) = s^- + \Delta s$ ,  $s(\lambda_s^-) = s^- - \Delta s$  and  $\Delta F_{\lambda, \Delta s}^K(x) = F_{\lambda_s^+}^K(x) - F_{\lambda_s^-}^K(x)$ .

## B. Mathematical Expression of Fragility

In essence, fragility is the sensitivity of a given risk measure to an error in the estimation of the (possibly one-sided) deviation parameter of a distribution, especially due to the fact that the risk measure involves parts of the distribution – tails – that are away from the portion used for estimation. The risk measure then assumes certain extrapolation rules that have first order consequences. These consequences are even more amplified when the risk measure applies to a variable that is derived from that used for estimation, when the relation between the two variables is strongly nonlinear, as is often the case.

1) *Definition of Fragility: The Intrinsic Case:* The local fragility of a random variable  $X_\lambda$  depending on parameter  $\lambda$ , at stress level  $K$  and semi-deviation level  $s^-(\lambda)$  with pdf  $f_\lambda$  is its  $K$ -left-tailed semi-vega sensitivity  $V(X, f_\lambda, K, s^-)$ .

The finite-difference fragility of  $X_\lambda$  at stress level  $K$  and semi-deviation level  $s^-(\lambda) \pm \Delta s$  with pdf  $f_\lambda$  is its  $K$ -left-tailed finite-difference semi-vega sensitivity  $V(X, f_\lambda, K, s^-, \Delta s)$ .

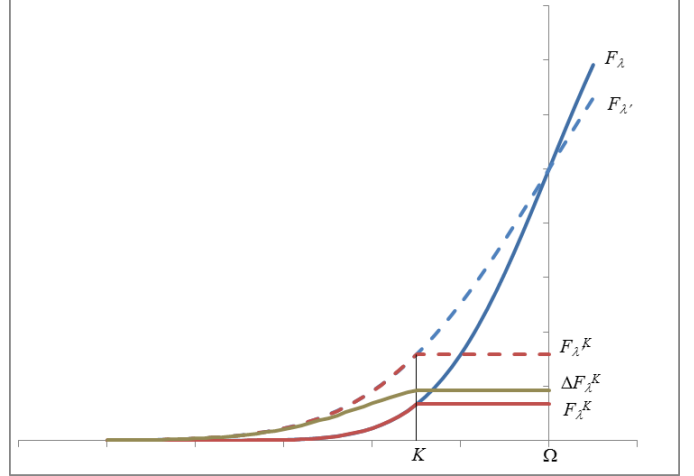


Figure 5: The different curves of  $F_\lambda(K)$  and  $F_{\lambda'}(K)$  showing the difference in sensitivity to changes at different levels of  $K$ .

In this definition, the *fragility* relies in the unsaid assumptions made when extrapolating the distribution of  $X_\lambda$  from areas used to estimate the semi-absolute deviation  $s^-(\lambda)$ , around  $\Omega$ , to areas around  $K$  on which the risk measure  $\xi$  depends.

2) *Definition of Fragility: The Inherited Case:* Next we consider the particular case where a random variable  $Y = \varphi(X)$  depends on another source of risk  $X$ , itself subject to a parameter  $\lambda$ . Let us keep the above notations for  $X$ , while we denote by  $g_\lambda$  the pdf of  $Y$ ,  $\Omega_Y = \varphi(\Omega)$  and  $u^-(\lambda)$  the left-semi-deviation of  $Y$ . Given a ‘‘strike’’ level

$L = \varphi(K)$ , let us define, as in the case of  $X$  :

$$\zeta(L, u^-(\lambda)) = \int_{-\infty}^K (\Omega_Y - y)g_\lambda(y) dy \quad (11)$$

The inherited fragility of  $Y$  with respect to  $X$  at stress level  $L = \varphi(K)$  and left-semi-deviation level  $s^-(\lambda)$  of  $X$  is the partial derivative:

$$V_X(Y, g_\lambda, L, s^-(\lambda)) = \frac{\partial \zeta}{\partial s}(L, u^-(\lambda)) = \left( \int_{-\infty}^K (\Omega_Y - Y) \frac{\partial g_\lambda}{\partial \lambda}(y) dy \right) \left( \frac{ds^-}{d\lambda} \right)^{-1} \quad (12)$$

Note that the stress level and the pdf are defined for the variable  $Y$ , but the parameter which is used for differentiation is the left-semi-absolute deviation of  $X$ ,  $s^-(\lambda)$ . Indeed, in this process, one first measures the distribution of  $X$  and its left-semi-absolute deviation, then the function  $\varphi$  is applied, using some mathematical model of  $Y$  with respect to  $X$  and the risk measure  $\zeta$  is estimated. If an error is made when measuring  $s^-(\lambda)$ , its impact on the risk measure of  $Y$  is amplified by the ratio given by the ‘‘inherited fragility’’.

Once again, one may use finite differences and define the *finite-difference inherited fragility* of  $Y$  with respect to  $X$ , by replacing, in the above equation, differentiation by finite differences between values  $\lambda^+$  and  $\lambda^-$ , where  $s^-(\lambda^+) = s^- + \Delta s$  and  $s^-(\lambda^-) = s^- - \Delta s$ .



### C. Effect of Nonlinearity on Intrinsic Fragility

Let us study the case of a random variable  $Y = \varphi(X)$ ; the pdf  $g_\lambda$  of which also depends on parameter  $\lambda$ , related to a variable  $X$  by the nonlinear function  $\varphi$ . We are now interested in comparing their *intrinsic fragilities*. We shall say, for instance, that  $Y$  is *more fragile* at the stress level  $L$  and left-semi-deviation level  $u^-(\lambda)$  than the random variable  $X$ , at stress level  $K$  and left-semi-deviation level  $s^-(\lambda)$  if the  $L$ -left-tailed semi-vega sensitivity of  $Y_\lambda$  is higher than the  $K$ -left-tailed semi-vega sensitivity of  $X_\lambda$ :

$$V(Y, g_\lambda, L, \mu^-) > V(X, f_\lambda, K, s^-) \quad (13)$$

One may use finite differences to compare the fragility of two random variables:  $V(Y, g_\lambda, L, \Delta\mu) > V(X, f_\lambda, K, \Delta s)$ . In this case, finite variations must be comparable in size, namely  $\Delta u/u^- = \Delta s/s^-$ .

Let us assume, to start, that  $\varphi$  is differentiable, strictly increasing and scaled so that  $\Omega_Y = \varphi(\Omega) = \Omega$ . We also assume that, for any given  $x < \Omega$ ,  $\frac{\partial F_\lambda}{\partial \lambda}(x) > 0$ .

In this case, as observed above,  $\lambda \rightarrow s^-(\lambda)$  is also increasing.

Let us denote  $G_y(y) = \mathbb{P}_{g_\lambda}(Y < y)$ . We have:

$$G_\lambda(\phi(x)) = \mathbb{P}_{g_\lambda}(Y < \phi(y)) = \mathbb{P}_{f_\lambda}(X < x) = F_\lambda(x). \quad (14)$$

Hence, if  $\zeta(L, u^-)$  denotes the equivalent of  $\xi(K), s^-$  with variable  $(Y, g_\lambda)$  instead of  $(X, f_\lambda)$ , we have:

$$\zeta(L, u^-(\lambda)) = \int_{-\infty}^{\Omega} F_\lambda^K(x) \frac{d\phi}{dx}(x) dx \quad (15)$$

Because  $\varphi$  is increasing and  $\min(\varphi(x), \varphi(K)) = \varphi(\min(x, K))$ . In particular

$$\mu^-(\lambda) = \zeta(\Omega, \mu^-(\lambda)) = \int_{-\infty}^{\Omega} F_\lambda^K(x) \frac{d\phi}{dx}(x) dx \quad (16)$$

The  $L$ -left-tail-vega sensitivity of  $Y$  is therefore:

$$V(Y, g_\lambda, L, u^-(\lambda)) = \frac{\int_{-\infty}^{\Omega} \frac{\partial F_\lambda^K}{\partial \lambda}(x) \frac{d\phi}{dx}(x) dx}{\int_{-\infty}^{\Omega} \frac{\partial F_\lambda}{\partial \lambda}(x) \frac{d\phi}{dx}(x) dx} \quad (17)$$

For finite variations:

$$V(Y, g_\lambda, L, u^-(\lambda), \Delta u) = \frac{1}{2\Delta u} \int_{-\infty}^{\Omega} \Delta F_{\lambda, \Delta u}^K(x) \frac{d\phi}{dx}(x) dx \quad (18)$$

Where  $\lambda_{u^-}^+$  and  $\lambda_{u^-}^-$  are such that  $u(\lambda_{u^-}^+) = u^- + \Delta u$ ,  $u(\lambda_{u^-}^-) = u^- - \Delta u$  and  $F_{\lambda, \Delta u}^K(x) = F_{\lambda_{u^-}^+}^K(x) - F_{\lambda_{u^-}^-}^K(x)$ .

Next, Theorem 1 proves how a concave transformation  $\varphi(x)$  of a random variable  $x$  produces fragility.

#### Fragility Transfer Theorem

**Theorem 1.** *Let, with the above notations,  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  be a twice differentiable function such that  $\varphi(\Omega) = \Omega$  and for any  $x < \Omega$ ,  $\frac{d\varphi}{dx}(x) > 0$ . The random variable  $Y = \varphi(X)$  is more*

*fragile at level  $L = \varphi(K)$  and pdf  $g_\lambda$  than  $X$  at level  $K$  and pdf  $f_\lambda$  if, and only if, one has:*

$$\int_{-\infty}^{\Omega} H_\lambda^K(x) \frac{d^2\varphi}{dx^2}(x) dx < 0$$

Where

$$H_\lambda^K(x) = \frac{\partial P_\lambda^K}{\partial \lambda}(x) \Big/ \frac{\partial P_\lambda^K}{\partial \lambda}(\Omega) - \frac{\partial P_\lambda}{\partial \lambda}(x) \Big/ \frac{\partial P_\lambda}{\partial \lambda}(\Omega) \quad (19)$$

and where

$$P_\lambda(x) = \int_{-\infty}^x F_\lambda(t) dt \quad (20)$$

is the price of the "put option" on  $X_\lambda$  with "strike"  $x$  and

$$P_\lambda^K(x) = \int_{-\infty}^x F_\lambda^K(t) dt$$

is that of a "put option" with "strike"  $x$  and "European down-and-in barrier" at  $K$ .

$H$  can be seen as a *transfer function*, expressed as the difference between two ratios. For a given level  $x$  of the random variable on the left hand side of  $\Omega$ , the second one is the ratio of the vega of a put struck at  $x$  normalized by that of a put "at the money" (i.e. struck at  $\Omega$ ), while the first one is the same ratio, but where puts struck at  $x$  and  $\Omega$  are "European down-and-in options" with triggering barrier at the level  $K$ .

The proof is detailed in [?] and [?].

#### Fragility Exacerbation Theorem

**Theorem 2.** *With the above notations, there exists a threshold  $\Theta_\lambda < \Omega$  such that, if  $K \leq \Theta_\lambda$  then  $H_\lambda^K(x) > 0$  for  $x \in (\infty, \kappa_\lambda]$  with  $K < \kappa_\lambda < \Omega$ . As a consequence, if the change of variable  $\varphi$  is concave on  $(-\infty, \kappa_\lambda]$  and linear on  $[\kappa_\lambda, \Omega]$ , then  $Y$  is more fragile at  $L = \varphi(K)$  than  $X$  at  $K$ .*

One can prove that, for a monomodal distribution,  $\Theta_\lambda < \kappa_\lambda < \Omega$  (see discussion below), so whatever the stress level  $K$  below the threshold  $\Theta_\lambda$ , it suffices that the change of variable  $\varphi$  be concave on the interval  $(-\infty, \Theta_\lambda]$  and linear on  $[\Theta_\lambda, \Omega]$  for  $Y$  to become more fragile at  $L$  than  $X$  at  $K$ . In practice, as long as the change of variable is concave around the stress level  $K$  and has limited convexity/concavity away from  $K$ , the fragility of  $Y$  is greater than that of  $X$ .

Figure 6 shows the shape of  $H_\lambda^K(x)$  in the case of a Gaussian distribution where  $\lambda$  is a simple scaling parameter ( $\lambda$  is the standard deviation  $\sigma$ ) and  $\Omega = 0$ . We represented  $K = -2\lambda$  while in this Gaussian case,  $\Theta_\lambda = -1.585\lambda$ .

#### DISCUSSION

##### Monomodal case

We say that the family of distributions  $(f_\lambda)$  is *left-monomodal* if there exists  $K_\lambda < \Omega$  such that  $\frac{\partial f_\lambda}{\partial \lambda} \geq 0$  on  $(-\infty, \kappa_\lambda]$  and  $\frac{\partial f_\lambda}{\partial \lambda} \leq 0$  on  $[\mu_\lambda, \Omega]$ . In this case  $\frac{\partial P_\lambda}{\partial \lambda}$  is a convex function on the left half-line  $(-\infty, \mu_\lambda]$ , then concave after the inflexion point  $\mu_\lambda$ . For  $K \leq \mu_\lambda$ , the function  $\frac{\partial P_\lambda^K}{\partial \lambda}$  coincides with  $\frac{\partial P_\lambda}{\partial \lambda}$  on  $(-\infty, K]$ , then is a linear extension, following the tangent to the graph of  $\frac{\partial P_\lambda}{\partial \lambda}$  in  $K$  (see graph

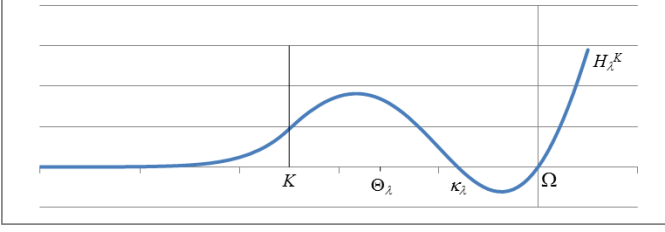


Figure 6: The Transfer function  $H$  for different portions of the distribution: its sign flips in the region slightly below  $\Omega$

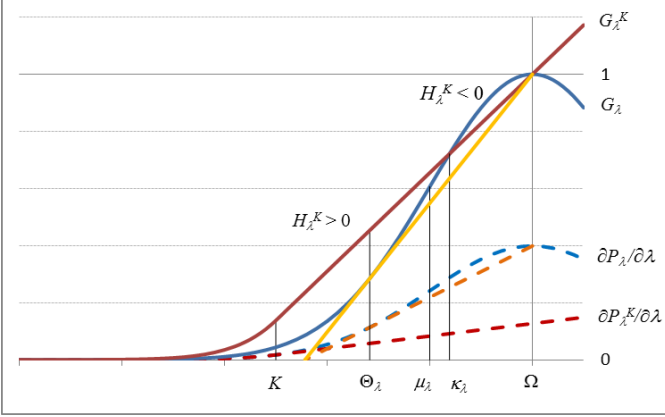


Figure 7: The distribution of  $G_\lambda$  and the various derivatives of the unconditional shortfalls

below). The value of  $\frac{\partial P_\lambda^K}{\partial \lambda}(\Omega)$  corresponds to the intersection point of this tangent with the vertical axis. It increases with  $K$ , from 0 when  $K \rightarrow -\infty$  to a value above  $\frac{\partial P_\lambda}{\partial \lambda}(\Omega)$  when  $K = \mu_\lambda$ . The threshold  $\Theta_\lambda$  corresponds to the unique value of  $K$  such that  $\frac{\partial P_\lambda^K}{\partial \lambda}(\Omega) = \frac{\partial P_\lambda}{\partial \lambda}(\Omega)$ . When  $K < \Theta_\lambda$  then  $G_\lambda(x) = \frac{\partial P_\lambda}{\partial \lambda}(x) / \frac{\partial P_\lambda}{\partial \lambda}(\Omega)$  and  $G_\lambda^K(x) = \frac{\partial P_\lambda^K}{\partial \lambda}(x) / \frac{\partial P_\lambda^K}{\partial \lambda}(\Omega)$  are functions such that  $G_\lambda(\Omega) = G_\lambda^K(\Omega) = 1$  and which are proportional for  $x \leq K$ , the latter being linear on  $[K, \Omega]$ . On the other hand, if  $K > \Theta_\lambda$  then  $\frac{\partial P_\lambda^K}{\partial \lambda}(\Omega) < \frac{\partial P_\lambda}{\partial \lambda}(\Omega)$  and  $G_\lambda(K) < G_\lambda^K(K)$ , which implies that  $G_\lambda(x) < G_\lambda^K(x)$  for  $x \leq K$ . An elementary convexity analysis shows that, in this case, the equation  $G_\lambda(x) = G_\lambda^K(x)$  has a unique solution  $\kappa_\lambda$  with  $\mu_\lambda < \kappa_\lambda < \Omega$ . The “transfer” function  $H_\lambda^K(x)$  is positive for  $x < \kappa_\lambda$ , in particular when  $x \leq \mu_\lambda$  and negative for  $\kappa_\lambda < x < \Omega$ .

### Scaling Parameter

We assume here that  $\lambda$  is a scaling parameter, i.e.  $X_\lambda = \Omega + \lambda(X_1 - \Omega)$ . In this case, as we saw above, we have

$$f_\lambda(x) = \frac{1}{\lambda} f_1 \left( \Omega + \frac{x - \Omega}{\lambda} \right), F_\lambda(x) = F_1 \left( \Omega + \frac{x - \Omega}{\lambda} \right)$$

$$P_\lambda(x) = \lambda P_1 \left( \Omega + \frac{x - \Omega}{\lambda} \right) \text{ and } s^-(\lambda) = \lambda s^-(1).$$

Hence

$$\xi(K, s^-(\lambda)) = (\Omega - K) F_1 \left( \Omega + \frac{K - \Omega}{\lambda} \right) + \lambda P_1 \left( \Omega + \frac{K - \Omega}{\lambda} \right) \quad (21)$$

$$\begin{aligned} \frac{\partial \xi}{\partial s^-}(K, s^-) &= \frac{1}{s^-(1)} \frac{\partial \xi}{\partial \lambda}(K, \lambda) \\ &= \frac{1}{s^-(\lambda)} \left( P_\lambda(K) + (\Omega - K) F_\lambda(K) + (\Omega - K)^2 f_\lambda(K) \right) \end{aligned} \quad (22)$$

When we apply a nonlinear transformation  $\varphi$ , the action of the parameter  $\lambda$  is no longer a scaling: when small negative values of  $X$  are multiplied by a scalar  $\lambda$ , so are large negative values of  $X$ . The scaling  $\lambda$  applies to small negative values of the transformed variable  $Y$  with a coefficient  $\frac{d\varphi}{dx}(0)$ , but large negative values are subject to a different coefficient  $\frac{d\varphi}{dx}(K)$ , which can potentially be very different.

### D. Fragility Drift

*Fragility* is defined as the sensitivity – i.e. the first partial derivative – of the tail estimate  $\xi$  with respect to the left semi-deviation  $s^-$ . Let us now define the *fragility drift*:

$$V'_K(X, f_\lambda, K, s^-) = \frac{\partial^2 \xi}{\partial K \partial s^-}(K, s^-) \quad (23)$$

In practice, fragility always occurs as the result of *fragility*, indeed, by definition, we know that  $\xi(\Omega, s^-) = s^-$ , hence  $V(X, f_\lambda, \Omega, s^-) = 1$ . The *fragility drift* measures the speed at which fragility departs from its original value 1 when  $K$  departs from the center  $\Omega$ .

### Second-order Fragility

The *second-order fragility* is the second order derivative of the tail estimate  $\xi$  with respect to the semi-absolute deviation  $s^-$ :

$$V''_{s^-}(X, f_\lambda, K, s^-) = \frac{\partial^2 \xi}{(\partial s^-)^2}(K, s^-)$$

As we shall see later, the *second-order fragility* drives the bias in the estimation of stress tests when the value of  $s^-$  is subject to uncertainty, through Jensen’s inequality.

### E. Definitions of Robustness and Antifragility

Antifragility is not the simple opposite of fragility, as we saw in Table 1. Measuring antifragility, on the one hand, consists of the flipside of fragility on the right-hand side, but on the other hand requires a control on the *robustness* of the probability distribution on the left-hand side. From that aspect, unlike fragility, antifragility cannot be summarized in one single figure but necessitates at least two of them.

When a random variable depends on another source of randomness:  $Y_\lambda = \varphi(X_\lambda)$ , we shall study the antifragility of  $Y_\lambda$  with respect to that of  $X_\lambda$  and to the properties of the function  $\varphi$ .

## DEFINITION OF ROBUSTNESS

Let  $(X_\lambda)$  be a one-parameter family of random variables with pdf  $f_\lambda$ . Robustness is an upper control on the *fragility* of  $X$ , which resides on the left hand side of the distribution.

We say that  $f_\lambda$  is *b-robust beyond stress level*  $K < \Omega$  if  $V(X_\lambda, f_\lambda, K', s(\lambda)) \leq b$  for any  $K' \leq K$ . In other words, the robustness of  $f_\lambda$  on the half-line  $(-\infty, K]$  is

$$R_{(-\infty, K]}(X_\lambda, f_\lambda, K, s^-(\lambda)) = \max_{K' \leq K} V(X_\lambda, f_\lambda, K', s^-(\lambda)), \quad (24)$$

so that *b-robustness simply means*

$$R_{(-\infty, K]}(X_\lambda, f_\lambda, K, s^-(\lambda)) \leq b$$

We also define *b-robustness over a given interval*  $[K_1, K_2]$  by the same inequality being valid for any  $K' \in [K_1, K_2]$ . In this case we use

$$R_{[K_1, K_2]}(X_\lambda, f_\lambda, K, s^-(\lambda)) = \max_{K_1 \leq K' \leq K_2} V(X_\lambda, f_\lambda, K', s^-(\lambda)). \quad (25)$$

Note that the *lower*  $R$ , the tighter the control and the *more robust* the distribution  $f_\lambda$ .

Once again, the definition of *b-robustness* can be transposed, using finite differences  $V(X_\lambda, f_\lambda, K', s^-(\lambda), \Delta s)$ .

In practical situations, setting a material upper bound  $b$  to the fragility is particularly important: one need to be able to come with actual estimates of the impact of the error on the estimate of the left-semi-deviation. However, when dealing with certain class of models, such as Gaussian, exponential of stable distributions, we may be lead to consider asymptotic definitions of robustness, related to certain classes.

For instance, for a given decay exponent  $a > 0$ , assuming that  $f_\lambda(x) = O(e^{ax})$  when  $x \rightarrow -\infty$ , the *a-exponential asymptotic robustness* of  $X_\lambda$  below the level  $K$  is:

$$R_{\text{exp}}(X_\lambda, f_\lambda, K, s^-(\lambda), a) = \max_{K' \leq K} \left( e^{a(\Omega - K')} V(X_\lambda, f_\lambda, K', s^-(\lambda)) \right) \quad (26)$$

If one of the two quantities  $e^{a(\Omega - K')} f_\lambda(K')$  or  $e^{a(\Omega - K')} V(X_\lambda, f_\lambda, K', s^-(\lambda))$  is not bounded from above when  $K \rightarrow -\infty$ , then  $R_{\text{exp}} = +\infty$  and  $X_\lambda$  is considered as not *a-exponentially robust*.

Similarly, for a given power  $\alpha > 0$ , and assuming that  $f_\lambda(x) = O(x^{-\alpha})$  when  $x \rightarrow -\infty$ , the  *$\alpha$ -power asymptotic robustness* of  $X_\lambda$  below the level  $K$  is:

$$R_{\text{pow}}(X_\lambda, f_\lambda, K, s^-(\lambda), a) = \max_{K' \leq K} \left( (\Omega - K')^{\alpha-2} V(X_\lambda, f_\lambda, K', s^-(\lambda)) \right)$$

If one of the two quantities

$$(\Omega - K')^\alpha f_\lambda(K') \\ (\Omega - K')^{\alpha-2} V(X_\lambda, f_\lambda, K', s^-(\lambda))$$

is not bounded from above when  $K' \rightarrow -\infty$ , then  $R_{\text{pow}} = +\infty$  and  $X_\lambda$  is considered as not  *$\alpha$ -power robust*. Note the exponent  $\alpha - 2$  used with the fragility, for homogeneity

reasons, e.g. in the case of stable distributions, when a random variable  $Y_\lambda = \varphi(X_\lambda)$  depends on another source of risk  $X_\lambda$ .

**Definition 1.** *Left-Robustness (monomodal distribution).* A payoff  $y = \varphi(x)$  is said *(a, b)-robust below*  $L = \varphi(K)$  for a source of randomness  $X$  with pdf  $f_\lambda$  assumed monomodal if, letting  $g_\lambda$  be the pdf of  $Y = \varphi(X)$ , one has, for any  $K' \leq K$  and  $L = \varphi(K)$ :

$$V_X(Y, g_\lambda, L', s^-(\lambda)) \leq aV(X, f_\lambda, K', s^-(\lambda)) + b \quad (27)$$

The quantity  $b$  is of order deemed of “negligible utility” (subjectively), that is, does not exceed some tolerance level in relation with the context, while  $a$  is a scaling parameter between variables  $X$  and  $Y$ .

Note that robustness is in effect impervious to changes of probability distributions. Also note that this measure robustness ignores first order variations since owing to their higher frequency, these are detected (and remedied) very early on.

## REFERENCES

- [1] VP Chistyakov. A theorem on sums of independent positive random variables and its applications to branching random processes. *Theory of Probability & Its Applications*, 9(4):640–648, 1964.
- [2] Julian Davies and Dorothy Davies. Origins and evolution of antibiotic resistance. *Microbiology and Molecular Biology Reviews*, 74(3):417–433, 2010.
- [3] Debra S Echt, Philip R Liebson, L Brent Mitchell, Robert W Peters, Dulce Obias-Manno, Allan H Barker, Daniel Arensberg, Andrea Baker, Lawrence Friedman, H Leon Greene, et al. Mortality and morbidity in patients receiving encainide, flecainide, or placebo: the cardiac arrhythmia suppression trial. *New England Journal of Medicine*, 324(12):781–788, 1991.
- [4] Paul Embrechts, Charles M Goldie, and Noël Veraverbeke. Subexponentiality and infinite divisibility. *Probability Theory and Related Fields*, 49(3):335–347, 1979.
- [5] Robert P Feldman and James T Goodrich. Psychosurgery: a historical overview. *Neurosurgery*, 48(3):647–659, 2001.
- [6] Dean Fergusson, Steve Doucette, Kathleen Cranley Glass, Stan Shapiro, David Healy, Paul Hebert, and Brian Hutton. Association between suicide attempts and selective serotonin reuptake inhibitors: systematic review of randomised controlled trials. *Bmj*, 330(7488):396, 2005.
- [7] Alfred P Fishman. Aminorex to fen/phen an epidemic foretold. *Circulation*, 99(1):156–161, 1999.
- [8] Curt D Furberg and Bertram Pitt. Withdrawal of cerivastatin from the world market. *Curr Control Trials Cardiovasc Med*, 2(5):205–7, 2001.
- [9] Arthur L Herbst, Howard Ulfelder, and David C. Poskanzer. Adenocarcinoma of the vagina. *New England Journal of Medicine*, 284(16):878–881, 1971. PMID: 5549830.
- [10] Richard Horton. Vioxx, the implosion of merck, and aftershocks at the {FDA}. *The Lancet*, 364(9450):1995 – 1996, 2004.
- [11] John T James. A new, evidence-based estimate of patient harms associated with hospital care. *Journal of patient safety*, 9(3):122–128, 2013.
- [12] Walter N Kernan, Catherine M Viscoli, Lawrence M Brass, Joseph P Broderick, Thomas Brott, Edward Feldmann, Lewis B Morgenstern, Janet Lee Wilterdink, and Ralph I Horwitz. Phenylpropanolamine and the risk of hemorrhagic stroke. *New England Journal of Medicine*, 343(25):1826–1832, 2000.
- [13] James H Kim and Anthony R Scialli. Thalidomide: the tragedy of birth defects and the effective treatment of disease. *Toxicological Sciences*, 122(1):1–6, 2011.
- [14] Anne M Larson, Julie Polson, Robert J Fontana, Timothy J Davern, Ezmina Lalani, Linda S Hynan, Joan S Reisch, Frank V Schiødt, George Ostapowicz, A Obaid Shakil, et al. Acetaminophen-induced acute liver failure: results of a united states multicenter, prospective study. *Hepatology*, 42(6):1364–1372, 2005.
- [15] Kathryn Pauly Morgan. Women and the knife: Cosmetic surgery and the colonization of women’s bodies. *Hypatia*, 6(3):25–53, 1991.

- [16] RITA Trial Participants. Coronary angioplasty versus coronary artery bypass surgery: the randomised intervention treatment of angina (rita) trial. *The Lancet*, 341(8845):573–580, 1993.
- [17] EJG Pitman. Subexponential distribution functions. *J. Austral. Math. Soc. Ser. A*, 29(3):337–347, 1980.
- [18] Woosley RL, Chen Y, Freiman JP, and Gillis RA. Mechanism of the cardiotoxic actions of terfenadine. *JAMA*, 269(12):1532–1536, 1993.
- [19] Graham AW Rook, Christopher A Lowry, and Charles L Raison. *Evolution, Medicine, and Public Health*, 2013(1):46–64, 2013.
- [20] Jozef L Teugels. The class of subexponential distributions. *The Annals of Probability*, 3(6):1000–1011, 1975.
- [21] Ina F Wallace, Nancy D Berkman, Kathleen N Lohr, Melody F Harrison, Adam J Kimple, and Michael J Steiner. Surgical treatments for otitis media with effusion: A systematic review. *Pediatrics*, pages peds–2013, 2014.
- [22] Paul B. Watkins and Randall W. Whitcomb. Hepatic dysfunction associated with troglitazone. *New England Journal of Medicine*, 338(13):916–917, 1998. PMID: 9518284.