How to Prevent Other Financial Crises

Nassim Nicholas Taleb and George A. Martin

This article argues that the crisis of 2007–2008 happened because of an explosive combination of agency problems, moral hazard, and “scientism”—the illusion that ostensibly scientific techniques would manage risks and predict rare events in spite of the stark empirical and theoretical realities that suggested otherwise. The authors analyze the varied behaviors, ideas and effects that in combination created a financial meltdown, and discuss the players responsible for the consequences. In formulating a set of expectations for future financial management, they suggest that financial agents need more “skin in the game” to prevent irresponsible risk-taking from continuing.

Introduction

Let us start with our conclusion, which is also a simple policy recommendation, and one that is not just easy to implement but has been part of history until recent days. We believe that “less is more” in complex systems—that simple heuristics and protocols are necessary for complex problems as elaborate rules often lead to “multiplicative branching” of side effects that cumulatively may have first order effects. So instead of relying on thousands of meandering pages of regulation, we should enforce a basic principle of “skin in the game” when it comes to financial oversight:

“The captain goes down with the ship; every captain and every ship.”

In other words, nobody should be in a position to have the upside without sharing the downside, particularly when others may be harmed. While this principle seems simple, we have moved away from it in the finance world, particularly when it comes to financial organizations that have been deemed “too big to fail.”
The best risk-management rule was formulated nearly 4,000 years ago. Hammurabi’s code specifies:

“If a builder builds a house for a man and does not make its construction firm, and the house which he has built collapses and causes the death of the owner of the house, that builder shall be put to death.”

Clearly, the Babylonians understood that the builder will always know more about the risks than the client, and can hide fragilities and improve his profitability by cutting corners—in, say, the foundation. The builder can also fool the inspector (or the regulator). The person hiding risk has a large informational advantage over the one looking for it.

The potency of the classical rule lies in the idea that people do not consciously wish to harm themselves. We feel much safer on a plane because the pilot, and not a drone, is at the controls. This principle has been applied by all civilizations, from the Roman heuristic that engineers spend time sleeping under the bridges they have built, to the maritime rule that the captain should be last to leave the ship when there is a risk of sinking.

The Hammurabi rule marks the separation between an agent’s interests and those of the client, or principal, she is supposed to represent. This is called the agency problem in the social sciences. Often closely associated is the problem of moral hazard, wherein an actor has incentive to behave in an economically or socially suboptimal manner (e.g. overly risky) because she does not bear all of the actual and/or potential costs of her action. In banking, these two are combined most acutely in the case of large institutions that may be deemed “too big to fail,” as increased risk taking (moral hazard) may lead to greater interim compensation to management (agent) at the expense of junior claimants such as shareholders and guarantors (taxpayers, etc.) (principals). The Hammurabi rule solves the joint agency and moral hazard problem by ensuring that the agent has sufficient non-diversifiable risk to incent the agent to act in the joint interest of the agent and the principal.

Nor does it contradict capitalism: for example, Adam Smith was chary of the joint stock company form as he worried it could be gamed by managers. Nor does it require much beyond some skin in the game for economic agents.

In sum, we believe the crisis of 2007–2008 happened because of an explosive combination of agency problems, moral hazard, and “scientism”—the illusion that ostensibly scientific techniques would manage risks and predict rare events in spite of the stark empirical and theoretical realities that suggested otherwise.
Part I: Etiology

The key driver of the Financial Crisis of 2007–2008 is the interplay of the following six forces, each of which can be linked to the misperception, misunderstanding, and the active hiding of the risks of consequential but low probability events (“Black Swans”) by those that stood to benefit from the obscuring of consequential risk. Other diagnoses, for example those of the Financial Crisis Inquiry Commission, focus more on epiphenomenal aspects of the crisis such as excessive borrowing, risky investments, opacity of markets, or failures of corporate governance.

The immediate precursor to the Crisis was the collapse of the securitized residential subprime market, which, along with other forms of collateral, was responsible for losses to financial institutions of more than $500 billion in 2007. By the end of 2009, the estimate for the total value of write-downs to credit instruments held by global banking and other financial organizations was close to $3.4 trillion. This was the result of:

1) Increases in hidden risks associated with low probability, large-consequence events (also known as “tail risks”) across all aspects of economic life, not just in banking. Tail risks could not (and cannot) be reliably priced, either mathematically or practically, as uncertainty about key aspects of tail risk has typically been on the order of, or greater than, understanding of the actuarial price of such risks. Nonlinearity in pricing risk is also exacerbated by an increase in debt, operational leverage and complexity, and the use of complex derivatives.

   a. The first author has shown that it is impossible to directly measure the risks in the tails of the probability distributions typical of financial markets. The relative errors swell in proportion to the remoteness of the event. Small variations in input, smaller than any uncertainty we have in the estimation of parameters, assuming generously that one has the right model of underlying probability distribution, can underestimate the probability of rare but recurring events (e.g., “12 sigma,” that is, twelve standard deviations) by close to a trillion times—a fact that has been so far routinely ignored by the bulk of the finance and economics establishment.

   b. Financial institutions amassed exposures in the “Fourth Quadrant,” which is where errors are both consequential and impossible to price, and the institutional vulnerability to these errors is large. For example, a key source of losses to financial institutions was “super senior” tranches of collateralized debt obligations (CDOs) which received credit ratings of “AAA” or above and therefore were assumed to have effectively zero probability of default (and therefore required no capital set aside to insulate against default risk). These products were retained or purchased by banks and other financial institutions...
engaged in producing or facilitating the production of CDOs. Such tranches typically generated positive carry on the order of ten basis points (bps), and were held in bank trading books so as to avoid capital charges “by the yard” (billion).

c. Fragility in the Fourth Quadrant can be re-expressed as “concavity to errors,” where losses from uncertain events vastly exceed possible profits from it over short horizons. One class of investment strategies that typically have this property are so-called “short volatility” trading strategies or positions (e.g. such as naked put-writing), which are often manifest in “carry trades.” These exposures increased significantly with concentration of positions in and across banks and other financial institutions as these organizations tended to find themselves attracted to similar strategies that had an apparent “history” of profitability without much realized risk/volatility. As more and more organizations found themselves with greater and similar exposures, the interconnection of exposure across asset classes and financial institutions contributed to virulent effects on asset prices: there were pressures on these institutions to unwind exposure as well as downward pressure on the value of assets as future cash flows associated with such assets were marked down due to increasing risk premia.

2) Asymmetric and flawed incentives that favor risk hiding in the tails. There were three primary flaws in the compensation methods that led to artificial earnings and not adequately and appropriately risk-adjusted compensation. They are a) asymmetric payoff: upside, limited, or no economic downside (a free or underpriced option); b) flawed frequency: annual compensation for risks that blow-up every few years, with absence of claw-back provisions; and c) misattribution: compensation for returns that are an attribute of the market (e.g. incentive fees that do not control for market beta or baseline risk premia), or an attribute of the organization (compensation for revenues derived from the use of balance sheet to generate “carry”).

3) Misunderstanding of elementary notions of probabilistic payoffs across economic life. The general public fails to notice that a manager “paid on profits” is not really “paid on profits” in the way it is presented and not compensated in the same way as the owner of a business, given the absence of negative payment on losses (this is known as “the fooled by randomness” argument). States of the world yielding financial failure are effectively ignored—this is “probabilistic blindness.” This asymmetry is called the “manager option,” or the “free option,” as it behaves exactly like a call option on the company granted by the shareholders, for free or close to little compensation. Thanks to the bailout of 2008–2009 (“TARP”), banks could use public funds to generate profits, compensating themselves generously in the process, having the undue assent of, or otherwise having convinced, the public and government that this compensation was somehow justified since they brought profits to the public purse. They hid the fact that the public would have been the sole payer in the event of losses from this gamble.

a. Mismatch of bonus frequency. Less misunderstood by policymakers is the idea that a manager paid on an annual frequency does not have an incentive to maximize profits; his incentive is to extend the time
to losses so he can accumulate bonuses before an eventual “blowup” for which he does not have to repay previous compensation. This provides the incentive to make a series of asymmetric bets (with a high probability of small profits and small probability of large losses) below probabilistic fair value. Thus, contrary to arguments of financial economists as well as management, asymmetric incentive structures may be worse for shareholders and/or investors than fixed payments to management. In addition, there is no inherent reason to believe that managerial risk tolerance is more or less than shareholder risk tolerance, so it is not clear ex ante that capital should be afforded an asymmetric payoff structure rather than management.

b. Misattribution of performance. The most pernicious form of this is the compensation of individuals based on perceived economic profitability (e.g. arising from a lower cost of finance) that essentially derives from implicit or explicit government guarantee. This produces a cost of capital savings for the de facto “too big to fail” firms, as the implicit government backing serves to lower the perceived probability that lenders will be paid back, and in turn, the interest rate that these banks must pay on their debt.

c. The agency problem is far more vicious in the tails, as it can explain the growing “left-skewness” (fragility) of corporations as they grow larger.9 The contrary argument—that size is necessary for economies of scale and scope in financial services firms, especially when adjusted for product mix—is largely unsupported by empirical evidence. Alan Greenspan, in his 2010 mea culpa, “The Crisis,” notes:

For years the Federal Reserve had been concerned about the ever larger size of our financial institutions. Federal Reserve research had been unable to find economies of scale in banking beyond a modest-sized institution. A decade ago, citing such evidence, I noted that ‘megabanks being formed by growth and consolidation are increasingly complex entities that create the potential for unusually large systemic risks in the national and international economy should they fail.’ Regrettably, we did little to address the problem.10

More recent evidence, which does not directly consider tail risk, has suggested that while there may be some economies of scale in U.S. banking11 beyond those directly attributable to the presumption of “too big to fail” policies, cost savings nevertheless may be achieved by breaking up banks and altering product mix.

4) Increased promotion of methods helping to hide tail risks. Value at Risk and similar methods have also promoted tail risks.12

a. As we note above, knowledge degrades very quickly in the tails of the distributions, making tail risks non-measurable (or, rather, impossible to estimate—“measure” conveys the wrong impression). Yet vendors have been promoting a method of risk management called “Value at Risk” (VaR) that just measures the risk at a particular point in the tail! It is supposed to project the expected extreme loss in an institution’s portfolio that can occur over a specific time frame at a specified level of confidence.13 For example: a standard daily VaR of
$1 million at a 1% probability tells you that you have less than a 1% chance of losing $1 million or more on a given day. There are many modifications around VaR, such as “conditional VaR,” equally exposed to errors in the tails.

Although such a definition of VaR is often presented as a “maximum” loss, it is technically not so in an open-ended exposure, since, conditional on losing more than $1 million, you may lose a lot more, say $5 million. So simply put, VaR encourages risk-taking in the tails and the appearance of “low volatility.” Moreover, application of Gaussian VaR models systematically hide the expected consequence of VaR violations, as larger and larger losses become exponentially smaller in probability.

Note here that regulators required banks to shift from hard heuristics or protocols—which are more robust to model error—to such “scientific” measurements. Academics and consultants have supported industry reliance on VaR, and have promoted risk management practices based on numerate, often backward-looking policies at the expense of policies based on prospective, qualitative risk management protocols. Such tools are particularly dangerous for managing risk in “relative value” or “netted” positions since actual risk becomes very sensitive to estimates of correlation—or dependence more generally—between positions, and not just tail risk.

Moreover, most of these tools focus on the risk associated with individual institutions, and underestimate the contagion that arises across institutions, or the concentration of risk in larger institutions that function as intermediaries to smaller institutions. Efforts by regulators to develop “predictive” models of the probability of systemic risk have largely been unsuccessful—in part because of the difficulty in forecasting systemic risk events, and in part because of the problem of false positives (also known as “Type I errors”) that, because of the multiple testing problem, occur at frequencies greater than the actual population value.

Criticism has been countered with the argument that “we have nothing better,” ignoring iatrogenic effects and mere phrenetic common sense.

b. Iatrogenics of measurements (harm done by the healer). These estimations presented as “measures” are known to increase risk taking. Numerous experiments provide evidence that professionals are significantly influenced by numbers that they know to be irrelevant to a professional decision, like writing down the last four digits of one’s social security number before making a numerical estimate of potential market moves. German judges who rolled dice before sentencing showed an increase of 50% in the length of the sentence when the dice showed a high number, without being conscious of this subconscious influence.

c. Linguistic conflation and reification. Calling these risk estimation “measures” creates confusion, making people think that something in current existence (not yet existing in the future) is being measured. Usually, these metrics are not appropriately presented as mere pre-
dictions with an abnormally huge error (as we saw during the most recent financial crisis, several orders of magnitude).

5) «Increased role of tail events in economic life thanks to “complexification” by global telecommunications and global economic integration, in addition to optimization of the systems.

a. The logic of winner-take-all effects. The Black Swan provides a review of “fat tail effects” coming from the organization of systems. Consider the “island effect,” which shows how a continent will have more acute concentration effects as species concentration drops in larger areas. The increase in winner-take-all effects—which includes blowups—is evident across economic variables.

b. Optimization. Specifically, the elimination of slack or redundancy makes systems left-skewed and more prone to extreme losses—this can be seen in concavity effects under the perturbation of parameters.

6) Growing misunderstanding of tail risks. Ironically, while tail risks have increased, financial and economic theories that discount tail risks have been more vigorously promoted; historically, operators understood risks more heuristically. This was particularly the case after the stock market crash of 1987 and after the Nobel was handed out to the creators of modern “portfolio theory.” Note the outrageous fact that the financial economics establishment missed the rise in these risks up to and including 2008, without incurring significant and extensive problems in credibility—for example, how well have business school curriculums been updated to incorporate the lessons of 2007–2008?

Indeed, principal errors by the financial economics establishment contribute to financial fragility by means of:

a. Ignorance of “true” fat tail effects. Also, misunderstanding that fat tails lead to massive imprecision in the measurement of low probability events (such as the use of Poisson jumps by Merton, or the more general versions of subordinated processes—these models fit the past with precision on paper but are impossible to calibrate out of sample in practice and induce a false sense of confidence). Misunderstanding those true-fat-tails cancels the core of academic financial theory and econometric methods used in practice.

b. Lack of awareness of the effect of parameter estimation on a model. Some models—actually almost all models—take parameters for granted when the process of parameter discovery in real-life leads to massive degradation of their results (due to “negative convexity effects” from layers of uncertainty).

c. Interpolation vs. extrapolation. Misunderstanding of the “atypicality of events” leads to looking for past disturbances for guidance when we have obvious evidence of lack of precedence. For instance, Rogoff and Reinhart look at past data without realizing that in fat-tailed domains one should extrapolate some properties from history, instead of interpolating or looking for naive similarities.
d. Optimization. It can be shown that optimization causes fragility when the payoff is concave under perturbation errors (e.g., in most cases, particularly when loss functions are based on convex functions).

e. Economies of scale. There are fragilities coming from size—as efficiencies typically rely on non-redundancies—for financial institutions themselves, as well as failure in size being more likely to cause externalities. 

Part II: The Responsible Parties

- Government officials: They promoted blindness to tail risks and nonlinearities (e.g., Bernanke’s pronouncement of a “great moderation” in macroeconomic volatility) and flawed tools in the hands of policymakers not making the distinction between different classes of randomness.
- Bankers/company executives: These individuals had an incentive to hide tail risks as a strategy to collect bonuses.
- Risk vendors and professional associations: The Chartered Financial Analyst (CFA) designation and International Association of Financial Engineers promoted Gaussian or quasi-Gaussian portfolio theory and Value-at-Risk methods without sufficient critical context.
- Business schools and the financial economics establishment: Specifically, promotion and teaching of Gaussian portfolio theory and inadequate risk measurement methods on grounds that “we need to give students something” (such arguments were used by medieval medicine). They still teach this.
- It would be rare to find an airplane pilot who would accept using a map of Saudi Arabia when flying over the Himalayas on grounds that “there is nothing else”—human intuitions know better. Yet once framed in financial terms, the reasoning reverses thanks to the agency problem: it is others that are harmed.
- Regulators: They promoted quantitative risk methods (VaR) over heuristics, the use of flawed risk metrics (credit ratings, such as AAA), and encouraged a certain class of risk taking.
- Bank of Sweden Prize, a.k.a. the Nobel Prize in Economics: This gave the Nobel stamp of scientific validity to empirically, mathematically, and scientifically invalid financial theories, such as Gaussian portfolio theory (Markowitz and Sharpe), option pricing (Scholes and Merton), Engle’s GARCH, Modigliani-Miller, and many more. In general their scientific invalidity comes from the use of wrong models of uncertainty that provide exactly the opposite results to what an empirically and mathematically more rigorous model of uncertainty would do.

... [T]he financial economics establishment should have been aware of the use of the wrong tools and the complete fiasco in implementing these theories, but they kept pushing the warnings under the rug, or hiding their responses.
Note that the entities listed above do not have meaningful “skin in the game”—they have an asymmetric version of it. That is, they get the upside benefits when things work out well without direct downside.

**Ethical considerations**

Surprisingly, the financial economics establishment should have been aware of the use of the wrong tools and the complete fiasco in implementing these theories, but they kept pushing the warnings under the rug, or hiding their responses. There has also been some diffusion of responsibility that is at the core of the system moving forward. The first author has debated: Robert Engle, Myron Scholes, Robert Merton, and Stephen Ross, among others, without any hint of their willingness to accept the risks they were creating with their “Procrustean Bed” methods of approximation—forcing reality to fit theory, at the expense of accurate practice, and in a manner that has induced greater net risk taking.

**Part III: A Suggested Remedy:**

**Symmetry in Rewards and Skin in the Game**

As we saw with banks, Toyota’s problem, the BP oil spill, and other similar cases of blowups from underinsured small risks, an economic system with a severe agency problem builds a natural tendency to push and hide risks in the tails, even without help from the academic economics establishment. When risks keep growing where they can be seen the least, there is a need to eliminate moral hazard by making everyone involved in the accumulation of material risks accountable both chronologically and statistically.

Hence the principle of symmetry in rewards—or “skin in the game”—that we presented in the introduction. It mandates that captains go down with ships. Accountability for everyone involved in risk-bearing for others, especially systemically; no exception, not a single one—contractually, morally, legally, or by means of whatever can be done to evade responsibility.

Accountability for everyone involved in risk-bearing for others, especially systemically; no exception, not a single one—contractually, morally, legally, or by means of whatever can be done to evade responsibility. This includes the academic finance establishment (including those that legitimate bad practice, such as rating agencies, forecasters, bank managers, etc.). The academic establishment, including professional organizations, has only engaged in limited introspection and self-criticism, and has largely been exempt from criticism from outside the academy (for an example of this, see the report of the Financial Crisis Inquiry Commission). It is time to realize that capitalism is not about providing free options.
What Should We Do?

Preaching is largely ineffective in the background of such natural resistance and absence of accountability. Legal recourse and regulatory change is needed to enforce skin in the game.

International regulatory and supervisory authorities have recognized that “compensation practices at large and systemically important financial institutions were a key contributing factor to the global financial crisis” and these bodies have in turn proposed principles and policies designed to align compensation and risk-taking at large financial institutions. While it is laudatory that compensation policies for large and systemically important financial institutions are now viewed as within the purview of regulatory and supervisory authorities by those authorities, and that there have been some efforts at larger institutions to institute compensation policies that, at least on the surface, are designed to mitigate compensation for risk-taking, nevertheless, these policies 1) have seen limited enforcement, 2) have suffered from substantial time-inconsistency, with the implementation of actual claw-backs being rare, and 3) have not addressed the link between short-term actions and the problem of long-term risk bearing (e.g. compensation in shares is essentially compensation in a random cash flow that is largely unaffected by individual actions).

In many circumstances, particularly for financial institutions subject to implicit government guaranty, it is particularly important that compensation policies not be allowed to devolve into subsidization of risk takers, nor allow compensation to deplete capital cushions. In many cases where systemic risks are high and implicit subsidies are present, we believe that meaningful asymmetry in compensation should be prohibited. We believe that shareholders should be similarly concerned.

But it is not just material risk takers who need to be held to account; those who enable moral hazard by breaking or obscuring the link between risk taking and asymmetry in rewards should be held responsible too.

Of course, academics can claim exemption as the generators of speculative thought. It would be hard to, say, make postmodern theorists accountable for their ideas on health and medicine. However, to the extent that finance academics offer themselves as sources of professional education on financial risk (as so many do through business schools), they should be subject to criticism. So it is those helping to implement misguided ideas on risk in practice who need to be made accountable. A theoretical biologist does not bear the same responsibility for harm as, say, the American Medical Association or a private doctor.

Two final parties that need to be brought to critical account are: 1) those associations and vendors that put the above flawed risk techniques

it is particularly important that compensation policies not be allowed to devolve into subsidization of risk takers, nor allow compensation to deplete capital cushions.
into the hands of practitioners and cause unwarranted increased confidence; and 2) the Bank of Sweden Prize that has given (and still gives) the Nobel stamp of scientific legitimacy to these techniques. For the Nobel stamp has given these methods the credibility to propagate and partially displace time-tested risk management heuristics and protocols.

Notes

1 This paper is a synthesis and expansion of the deposition of the first author in front of the Financial Crisis Inquiry Commission, 2010.

2 By epiphenomenal we mean that they were present in situations without crisis, and in and of themselves were not causal. For instance “greed” is always observed when there is a crisis, but we can observe situations of greed without crises. The primary reason behind the 2008 crisis was the accumulation of tail risks, as global financial markets had never previously become so sensitive to such risks.


8 The “option” terminology is used in order to connote the actual financial products that are called “options” in order to emphasize that these structural asymmetries in contracts or other forms of private ordering have concrete and material economic value to the agent.

9 Left-skewness is shown in F. Degeorge et al., and re-discussed in the argument on convexity. See F. Degeorge et al., “Earnings Management to Exceed Thresholds,” Journal of Business 72, no. 1 (1999), 1–33.


12 Even Alan Greenspan makes a limited version of this observation post hoc: “It is in such circumstances that we depend on our highly sophisticated global system of financial risk management to contain market breakdowns. How could it have failed on so broad a scale? The paradigm that spawned Nobel Prize winners in economics was so thoroughly embraced by academia, central banks, and regulators that by 2006 it became the core of global regulatory standards (Basel II). Many quantitative firms whose number crunching sought to expose profitable market trading principles were successful so long as risk aversion moved incrementally (which it did much of the time). But crunching data that covered only the last 2 or 3 decades prior to the current crisis did not yield a model that could anticipate a crisis.” See A. Greenspan, “The Crisis,” Brookings Papers in Economic Activity, 2010, 11.


14 Data show that with methods meant to improve the standard VaR, like “expected shortfall” or “conditional VaR” with economic variables, past losses (or lack thereof) do not predict future losses in the extreme tails. Stress testing may also be suspicious because of the subjective nature of a “reasonable stress” number—we tend to underestimate the magnitude of outliers due to the dynamic nature of financial markets (we do not have a static history from which to form actuarial estimates of events). The frequency and severity of “jumps” (negative or positive) in financial asset values is not predictable from past jumps.
J. Nocera writes in a 2009 *New York Times* article: “In the late 1990s, as the use of derivatives was exploding, the Securities and Exchange Commission ruled that firms had to include a quantitative disclosure of market risks in their financial statements for the convenience of investors, and VaR became the main tool for doing so. Around the same time, an important international rule-making body, the Basel Committee on Banking Supervision, went even further to validate VaR by saying that firms and banks could rely on their own internal VaR calculations to set their capital requirements.” J. Nocera, “Risk Management,” *New York Times*, January 4, 2009.


The key problem with finance theory has been supplanting embedded and time-derived heuristics, such as the interdicts against debt and forecasting, with models akin to “replacing a real hand with an artificial one.”


In early 2009 a *Forbes* journalist in the process of writing a profile of the first author spoke to NYU’s Robert Engle who got the Bank of Sweden Prize (“Nobel”) for methods that patently have never worked outside of papers. The reporter noted to the first author that Engle’s response was that academia was not responsible for tail risks, since it is the government’s job to cover the losses beyond a certain point. This is the worst type of moral hazard argument which played into the hands of the too-big-to-fail problem.

Author conversations with the King of Sweden and members of the Swedish Academy resulted in the following astonishing observation: they do not feel concerned, nor act as if they are in any way responsible for the destruction since, for them, “this is not the Nobel”, just a Bank of Sweden prize.

Speculators using their own funds have been reviled, but unlike professors, *New York Times* journalists, and others, speculators (particularly those without the free option of society’s bailout) directly bear the costs of their mistakes.
