Systemic Risk of Modelling

amlin

Continuity is...
adapting our response
to your risk management
requirements

Investing in a better understanding of risk

Simon Beale
Group Chief Underwriting Officer, Amlin plc

Over recent years we have become more and more reliant upon technology within our personal and business lives. A car driver's reading of directions from maps and street signs has become less necessary with the advent of satellite navigation devices. The technology quickly becomes taken for granted, old skills become redundant and behaviours change. Some studies have suggested that people drive faster and take less notice of their surroundings (including pedestrians) as they gain greater confidence from the dashboard device – the risk is therefore heightened.

Translated to the financial markets such reliance has proven catastrophic. In 2008 Lehman Brothers collapsed. The technology developed to control the risks both in Lehman's, rating agencies and connected banks failed – the models were flawed, the behaviours were inappropriate for the unmodelled reality and the result was severe contagion in the financial markets.

The insurance market largely escaped this contagion, and research led by the Geneva Association has firmly established that our industry is a source of stability for the financial system, rather than a contributor to its systemic risk. But we are, perhaps equally, vulnerable to such scenarios, where modelled expectations of risk do not reflect reality and / or modelled output is taken for granted. In such situations, human behaviour becomes complacent and reaction to the real world may be inappropriate or impossible due to a lack of relevant experience or know-how.

Amlin recognises this threat to its business, but also to the insurance market as a whole. Amlin has therefore commissioned the Future of Humanity Institute at Oxford University's Oxford Martin School to research the Systemic Risk of Modelling. In order to aid the understanding of how such risk may impact the insurance market, the research has a strong collaboration across the insurance industry and academia – Amlin is not just funding the research but also taking an active role, with insurance practitioners testing academic ideas. Amlin is also encouraging clients, partners, market commentators and competitors to engage in the findings to help raise standards and deliver durable benefits for the insurance market.

I am grateful to the entire team who have coordinated the production of the initial papers for this research. I believe it is an important first step in understanding better the risks we face and ensuring that our reliance on the increasing amalgamation of models does not prove catastrophic for the insurance industry.

Systemic risk in modelling

Nick Bostrom
Director of Future of Humanity Institute, The University of Oxford

Over the past few years, systemic risk has grown into a major topic of interest for a variety of fields. Financial collapses and the increasing interconnectedness of key institutions suggest that the study of systemic risk warrants serious consideration. Some risks have extremely high stakes, so failure to understand and characterise these phenomena could be catastrophic.

The source of these risks varies according to the field. In ecology, systemic risks can emerge through networks of species, whereas in finance, these networks form between banks or other institutions. The connections can take a variety of forms as well. They can form by holding similar assets or sometimes through widespread usage of similarly flawed models. The latter is of particular interest to us, as biases of human risk cognition and structural group-think can stymie our ability to make accurate judgments.

This challenge is especially notable when such judgements are outsourced to a set of prescribed models or heuristics. When the assumptions of such models are not thoroughly questioned, ubiquitous acceptance of the model can result in systemic issues. If groups designing or implementing the models have biases or conflicts of interest, these issues can be compounded further.

We are therefore interested in how modelling practices can contribute to systemic risk. Flawed models are not simply poor predictors of reality, they are creators of externalities. Modelling practices are spread and adopted through networks, typically under the assumption that others in the network are using dissimilar models. While many models attempt to quantify systemic risk, we invert the question and ask, "what systemic risks emerge from modelling?"

Our challenge lies in determining not only how these models are constructed, but how the framework of construction and the context in which they are applied relates to the broader market. We are therefore delighted to partner with Amlin in a shared effort to think about these issues and understand the risk.

Quantitative Models: Handle with Care

JB Crozet
Head of Group Underwriting Modelling, Amlin plc

Philippe Regazzoni
CEO, Amlin Re Europe

"God, Grant us the serenity to accept the things we cannot change, the courage to change the things we can, and the wisdom to know the difference."

'The Serenity Prayer', popularised by the Alcoholics Anonymous, might sound very remote from our world of insurance and risk modelling. But the prayer is fundamentally about "change", and one of the most profound changes our industry has experienced in the last two decades is the increasing influence of quantitative models. To take catastrophe models as an example:

- Twenty years ago, only a handful of avant-garde (re)insurers would be running catastrophe models, deployed on 100MHz single-core desktops;
- Nowadays, catastrophe models are widespread throughout the industry, with:
  - investments in enterprise deployments of over 100 cores, running at 2Ghz, being fairly common; and
  - cloud-based solutions around the corner.

This change is not limited to our industry. Without the aid of increasingly advanced computer-based models, the human brain would be unable to navigate in today's globalised world. From spreadsheets to statistical models they underpin power and energy supply, global goods transport, traffic, weather forecasts, climate, population and financial models. Models are everywhere. They try to simplify complex systems to support rational decision making.

These models are an abstraction of reality; but they are only as good as:

- The quality of their input;
- The completeness and relevance of the abstraction (i.e. the theoretical model);
- The strength of the algorithms and assumptions used; and
- The competence of the interpretation of the results.

These limitations create risk in the usage of models; and very often we are not or only partially aware of them.

Their failure might mean the failure of the entire system. The recent financial crisis exemplified this in an extraordinary way, but what would be the consequences of a failure of global transport, climate or power supply models?

We have learnt to use the models, but have we learnt to manage them? Do we manage the risk of being increasingly reliant on these models? The change is reality. Have we learnt to manage this change?

In insurance, models are widespread. Financial modelling, capital adequacy calculations, the calculation of premium and risk are model applications used by all industry participants. They are often based on the same applications, using the same or similar source data.

Quantitative Models: Handle with Care

This is inevitable. Going back to using 16% of premium as a risk measure or using single scenario extrapolation or 1per mille of TSI as Loss Costs on reinsurance business is not an option. These were first steps that have led us towards quantitative modelling today.

Shareholders, regulators, competitors and clients drive insurance towards model-based management. Simply because the traditional, factor based approach isn't able to capture the complexity of the insurance business. The usage of modelling has very quickly developed towards an industry standard.

The rapid surge in model usage creates the potential for a systemic risk in our industry, by exposing it to our behavioural biases.

The purpose of the research being sponsored by Amlin is to highlight how, like many good medicines, quantitative models can have side-effects as well as the intended benefits. The research team hope to gain insights into the wider topic of the systemic risk of modelling, through analysing the specific question of catastrophe modelling in (re)insurance and looking into what systemic risks emerge and what tools and methods can be applied to manage these new risks.

Amlin's interest is to identify tools and methods to better manage our risk, and to discuss these approaches within the industry. Ultimately we wish to become a more stable and sustainable counterparty for our business partners.

The Quiet Revolution

In our mind, the title of Susan Cain's book "Quiet: The Power of Introverts in a World That Can't Stop Talking" most adeptly describes the quiet but steadfast progression of modelling within our business.

From the first Life Mortality Tables in the late 17th Century to the Internal Models in the Solvency 2 regime, our industry has embraced quantitative models whenever they were beneficial to the business. To a large extent, however, adopting modelling technology has been a passive process and not considered to be core to the (re)insurance business; unlike, for instance, the fund management industry.

Intended Benefits

The benefits of quantitative models are undeniable, which explains the speed of their adoption by the market and its regulators.

By providing a consistent, informed assessment of the risks within our business, quantitative models have helped (re)insurers on several fronts:

- Risk Management: the ability to manage risk on a probabilistic basis, and compare the riskiness of very different types of exposures (e.g. life assurance vs. property catastrophe);
- Portfolio Management: the ability to measure the risk-return profile of the current portfolio, and produce alternative "what if" scenarios;
- Technical Pricing: the ability to measure the expected cost associated with a specific contract, and compare the relative value of alternative policy structures.

These benefits have helped to partially "de-risk" the business of (re)insurance, by providing a control framework; thereby lowering our "cost of capital" and enabling more affordable (re)insurance in the market.

Potential Side Effects

While the introduction of quantitative models in each individual aspect of the business should strengthen it, they could also make the system more fragile and vulnerable to error or misuse; thereby exposing the whole to a larger systemic failure.

For instance, we now operate in an environment in which quantitative models typically underpin:

- Technical Pricing;
- Exposure Management & Realistic Disaster Scenarios;
- Enterprise Risk Management;
- Regulatory Capital & ORSA Reporting; and
- Rating Agencies' Financial Strength Assessments.

This widespread institutionalisation of quantitative models across all the different layers of defence means that:

- An organisation is more exposed to model error, e.g. a "black swan" event not adequately apprehended by the models; and
- There is sometimes a lack of appreciation that "not all models are created equal", and that they are supported by varying levels of research, statistical credibility and data quality.

The multiple natural catastrophe events experienced by the (re)insurance industry in 2011 provided a vivid illustration of these weaknesses. The market had to face non-modelled exposures (e.g. Thailand floods); non-modelled perils (e.g. Japan Tsunami); and unexpected modelled events (e.g. Japan earthquake over Mw 9; New-Zealand earthquake with ultra-liquefaction), most of which were typically poorly catered for by the models, if at all. We could conclude that every time the limits of the models are tested, the models fail.

This risk is also present across our industry, where the models are surprisingly similar; coming from the same pool of talent and subject to the same regulatory approval. The same models are then calibrated with the same factors. For example, the largest flood in France is assumed to be 10bn Euro and the industry calibrates to this number.

Human, All Too Human

The existence of model error, popularised by George Box's famous quote *"all models are wrong but some are useful"*, is reasonably understood by practitioners in the market.

Our industry is, however, much less familiar with the risks arising from the behavioural aspects of the modelling process; or, in other words: how, in human hands, "all models are wrong, but even the useful ones can be misused".

Thinking, Fast and Slow

Quantitative models have the significant advantage of scaling up with technological progress. Unlike expert judgment which is limited in speed and footprint, models become faster and more advanced as technology improves.

Often, however, the gains in calculation speed are translated into a higher reporting frequency without necessarily a full appreciation for the critical, qualitative difference between, for example:

- A Chief Underwriting Officer receiving a quarterly report on the risk profile of the portfolio, supported by qualitative commentary from his Chief Pricing Actuary highlighting the limitations of the analysis; and
- The same Chief Underwriting Officer accessing the same figures daily, on a self-service basis at the press of a button.

These two types of reporting have a purpose adapted to different tasks. To draw a parallel with Daniel Kahneman's *"Thinking, Fast and Slow"*: the slower and more deliberative approach is better adapted to more strategic situations, while the faster, instinctive reporting is best suited to monitoring contexts.

Without the awareness of this distinction, the temptation is great, however, for the Chief Underwriting Officer to rely on the faster, automated reporting for strategic decisions; leading to a "dumbing down" of the decision making process as a result of technological automation.

Limited Gene Pool

Unlike expert judgment, quantitative models are based on transparent assumptions, which can be adapted in order to improve predictive power or adapt to environmental changes over time. Similarly, a model identified to not be fit-for-purpose would quickly be disregarded if it did not adapt appropriately.

This evolutionary process is a powerful force, which has helped our industry get better and better models over time. But we must recognise that the institutionalisation of quantitative models can lead to structural "groupthink" and "limit the gene pool" by reducing the potential for model diversity.

Historically, the regulatory frameworks did not interfere with the (re)insurers' freedom to select and use models as they deemed fit. The regulatory rules for setting minimum capital requirements complemented the internal risk management perspective, with an independent view and an additional layer of defence.

The Internal Models in the UK ICAS and the coming EU Solvency 2 regimes (mimicking the Basel 2 regulations for financial institutions) have, however, taken a widely different stance. In essence, the setting of minimum capital requirements is outsourced to the (re)insurer if its Internal Model is approved by the regulator:

- Internal Model Approval requires the regulator to be comfortable with the model, which could limit the range of potential approaches and possibly introduce "asymmetrical error checking" (i.e. mostly scrutinising the models which do not fit expectations or preferences);
- The Documentation Standards require sufficient details to enable the Internal Model to be justifiable to a third party, possibly restricting the reliance on expert judgment and slowing down the introduction of innovation; and
- The Use Test requires that the Internal Model be used for risk management and key decision processes, which restricts the usage of alternative models within the organisation.

The risk for our industry is that we are unconsciously dis-incentivising the emergence of alternative approaches, which are vital for a fully functional evolutionary process.

Principal-Agent Dilemma

The large investments required to build sophisticated representations of (re)insurers' risks and the scalability of quantitative models, point to significant economies of scale from centralising and outsourcing their development to third-party vendors.

For instance, many (re)insurers license proprietary Economic Scenario Generators or Catastrophe Models from third-party vendors who generate the investment in talent and R&D.

While the financial benefits of outsourcing model-building to third-party vendors are often clear, the associated "outsourcing of cognition" presents some challenges in itself:

- The divergence in principal-agent interests might lead third-party vendors to be influenced by other priorities than modelling quality (e.g. production costs, sales potential, social and political context);
- (Re)insurers have reduced incentives to invest in modelling knowledge and talent, to the point that their decision-makers could become over-reliant on the "autopilot" and unable to critique or even function without it;
- The oligopolistic nature of markets with large economies of scale, allows the few players to be more authoritative as central source of knowledge, than is justified by the quality of their models alone.

Unfortunately, the more the industry tends to rely on a single source of knowledge, the smaller the upside when it gets things right and the greater the downside when it gets things wrong (as, one day, it inevitably will).

In Search of Sustainable Modelling

Having identified how the usage of quantitative models can introduce a systemic risk within an organisation and within the industry, the key questions are:

- How do we manage this risk, which is behavioural in nature?
- How do we educate users of model results about the potential pitfalls?
- How do we develop a sustainable, robust usage of models within our business and within the industry?
- Is there a generation gap developing in the acceptability of model vs. reality?

Our standard approach of "quantifying the risk using a model" seems clearly inappropriate on its own, as it would compound the risk rather than mitigate it. Model-independent approaches for risk management can be useful, but they remain more of a complement than a supplement.

In the words of Albert Einstein, *"problems cannot be solved with the same mind-set that created them"*.

Our collaborative research project with the Future of Humanity Institute ("FHI") at Oxford Martin School looks at our industry as a "human experiment" rather than a business.

We are hoping that combining our knowledge with this behavioural perspective could bring rich insights for our industry, and for the understanding of Systemic Risk in general… so that one day, we will have the serenity to accept the things we cannot model, the courage to model the things we can, and the wisdom to know the difference.

The FHI-Amlin Research Collaboration on Systemic Risk of Modelling

This project will pursue better understanding and management of systemic risk through the strategic collaboration between the Future of Humanity Institute at Oxford University and Amlin.

Systemic risks concern the unexpected collapse of an entire market, methods of doing business or methods of modelling, and are of great importance to managing risk on the large scale. The collaboration will be enabling research into how systemic risks emerge and can be managed, disseminating this research within Amlin and outside, and educating towards a self-sustaining culture of accurate thinking about risk.

Risks that emerge from complex decision-making, including decision-making about risk itself, are the main focus of interest in this project. Distributed thinking in organisations, the increasing use of complex models, and a rapidly changing world pose many understudied challenges that FHI and Amlin are well placed to explore together in order to find better ways of managing large scale risks.

This includes the problems of systemic risk in catastrophe modelling, how to maintain necessary institutional critical thinking despite biasing incentives, how to validate complex models, and how to handle changing uncertainty.

Contents

The Team

L-R Nick Bostrom, Nick Beckstead, Seán Ó hÉigeartaigh, Anders Sandberg, Vincent Müller, Andrew Snyder-Beattie, Stuart Armstrong

Defining systemic risk

Insurance Industry Perspective

James Illingworth

Chief Risk Officer, Amlin plc

White Paper Context

This paper explores the definition of systemic risk, clarifies how it differs from systematic and system risks, and presents the circumstances through which such risks can develop or manifest.

It does not seek to distil a single definition. Instead it highlights numerous variations in meaning that depend on context, and attempts to extract the common core concepts that run through each.

A key finding is that parts of a system may function well individually, but become vulnerable to a joint risk when connected, leading to a spread of risk that potentially affects the entire system.

It provides a unique challenge as, unlike other risks, adaptation and risk mitigation (including regulation) are not separate from the system, and can actually increase the systemic risk. Additionally much of the risk comes from the structure of the system, which is often constrained, making strong changes infeasible.

Industry Relevance

Insurance plays an understated and undervalued role in the world economy, through the provision of risk transfer mechanisms to enable wealth-creating entities to manage their exposure to an array of perils. In the absence of insurance mechanisms these entities would have to carry significantly more capital, much of it unused, or be prone to a wholly unacceptable risk of ruin.

At a micro level, individuals similarly rely on insurance risk transfer to deal with elements of fortuity in their lives. Fundamental to the provision of risk transfer is the belief from buyers that insurers' "promise to pay" is both reliable and effective. Insurers have the responsibility to maintain processes so that they are able to charge an adequate price for this risk transfer product, ensure that exposures are fully understood and reserved for, and that the risk capital being carried is sufficient to meet extreme circumstances. Therefore risk management processes in insurers are of critical importance to all stakeholders in an insurance enterprise, most particularly shareholders, policy holders and regulators.

As insurers' risk management structures, tools and models have developed in recent years, practices and methodologies have converged. This is a result of the development of accepted proprietary models and is driven by regulators' understandable requirements for levels of practice to be enhanced, reviewed or even benchmarked.

Systemic risk is greater in an environment where commonality of approach and methods overrides differing human and manual activities. This is apparent in many insurers' basic systems for pricing risk, reserving for future liabilities, management and modelling of exposure and in economic capital modelling.

This development accompanies a far wider change in the structures of international financial markets and the scope of large firms. Insurers were historically single territory or even specialised in one sector. Today many have diversified to multi-territory and most are multi-line. Globalisation has led to far greater multi-national operations and interconnectedness of risk between and among insurers; another key example of the interconnectedness of financial markets in the modern global economy.

This convergence carries a significant long-term risk to the industry. In common practice lies common Weakness. Consequently, whilst insurance entities and the industry as a whole may be very well prepared for the risks covered by common risk management systems, it may be unprepared for "black swan" type risk events which may expose risk system inadequacies. These potential weaknesses in the sector could critically undermine the confidence of policyholders, shareholders and regulators and undermine a vital component of the modern economy.

Next Steps

This white paper sets the scene for further research, in particular:

- Attempting to model and understand institutional group-think, through modelling of small markets and modelling itself (creating a "metamodel").
- Developing qualitative tools to assist with analyses of systemic model risk and catastrophe modelling.

Defining systemic risk

Anders Sandberg
Senior Research Fellow,
FHI-Amlin Collaboration

Nick Beckstead
Research Fellow,
FHI-Amlin Collaboration

Stuart Armstrong
Research Fellow,
FHI-Amlin Collaboration

Introduction

Systemic risk is a term that is widely used, yet ill defined. This paper will explore some of the meanings of systemic risk and related concepts such as resilience. The aim is not to distill a single, better definition, but rather to analyse core concepts that are important for reasoning about, predicting and mitigating systemic risks.

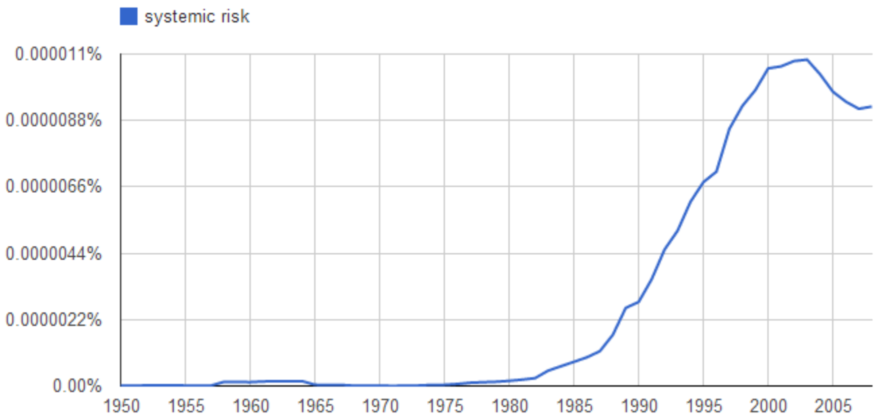


Figure 1: Google n-gram data for use of "systemic risk" in their corpus of scanned English books.

The use of "Systemic risk" as an expression grew tremendously across the 1990s, likely as part of the general move towards a "risk society" that began in the post-Cold War era[1]. It might have peaked a few years back, although uncertainties in measuring methods make this somewhat tentative. Examining Google trends (the number of searches for a term) show that the number is still increasing – and one of the most common searches is for "definition systemic risk".

Like many evocative terms such as "complexity" and "resiliency", "systemic risk" might seem somewhat intuitive when used. Unfortunately this evocativeness also makes different people and groups use it in divergent ways that actually mean different things.

The contrasting term "resiliency" has similar issues. It was originally proposed as a scientific term in ecology by Holling in 1973[2] (a measure of a system's ability to maintain its operational integrity). Increasingly the term is used frequently in economics, psychiatry, robotics, medicine, military strategy, governance and other disciplines.[3] One of the reasons is the intuition that underlying patterns of resilient systems or methods of increasing resiliency might be applicable across disciplines.

When the usage was actually studied in detail[4] it turned out that there were more than a dozen definitions and understandings local to different disciplines.

[1] In sociology the concept of a "risk society" is a society that is increasingly preoccupied with the future, safety and observing itself and its processes, and hence focuses on risk and risk management. (Beck 1992)

[2]. (Holling 1973)

[3] http://chronicle.com/article/That-Elastic-Term-/138917

[4] (Downes et al. 2013)

Defining systemic risk

Because of these multiple definitions one should not therefore expect statements about systemic risk (or resiliency) to directly carry between different groups. Conversely, even when overarching insights might be possible, they might not be recognized because different groups misunderstand each other. Hence confusion about definitions can impair mitigation efforts[5].

To make matters worse there might even be a degree of observer dependence. Alan Greenspan observed that[6] although:

> "[i]t is generally agreed that systemic risk represents a propensity for some sort of financial system disruption[,] one observer might use the term 'market failure' to describe what another would deem to have been a market outcome that was natural and healthy, even if harsh."

Even defining the edges of the system is fraught. One common feature of "black swans"[7] is that they represent an unexpected influence via links to markets or phenomena not normally considered part of the system under study. Since little or no effort had gone into analysing them, the impact has more disruptive effects than it would otherwise have had. In other words, a black swan that has been predicted and analysed, is not a black swan.

However, all this does not mean the word "systemic" is irrelevant: as we will see what matters is more *how* things are connected than where boundaries are drawn.

Financial systemic risk

The original use of systemic risk comes from finance, where the concept was spread in the mid-1990s. Since then usage has spread outside finance, for example into safety engineering[8], and generalised. We collected a sampling of definitions of systemic risk (Appendix), using them to study how the concept is used.

Contagion affecting the whole system

George G. Kaufman described systemic risk in a review[9] as

> "the risk or probability of breakdowns (losses) in an entire system as opposed to breakdowns in individual parts or components and is evidenced by comovements (correlation) among most or all the parts."

The effect on the entire system due to correlated action of the parts is a common implicit assumption, but relatively often unstated .

Kaufman looks at three types of definition that covers much of the financial usage. The first is the role of causation. It is common to denote a simultaneous shock that affects the whole market or system adversely as a systemic risk. In current usage this is a *systematic* risk rather than *systemic* risk (see below). Here the cause is external to the system rather than internal.

Second, many definitions focus on contagion, spillover or chain reaction effects. Here causation flows from one firm or institution to another, for example if one default leads to the next. This type of behaviour requires strong linkages, a hallmark of systemic risks that many authors have focused on[10].

For example, Steven L. Schwarcz[11]:

> "We can reach a working definition of systemic risk: the risk that (i) an economic shock such as market or institutional failure triggers (through a panic or otherwise) either (X) the failure of a chain of markets or institutions or (Y) a chain of significant losses to financial institutions, (ii) resulting in increases in the cost of capital or decreases in its availability, often evidenced by substantial financial-market price volatility"

[5] (Liedtke 2010)
[6] (Greenspan 1995)
[7] (Taleb 2007)

[8] (Reniers, Sörensen & Dullaert 2012)
[9] (Kaufmann 2000)
[10] For example, (Markose et al 2009, May & Arinaminpathy 2010, Cont et al 2011, Haldane &
May 2011, Battison et al 2012)
[11] (Schwarcz 2008)

Finally, spillover can occur without strong linkages that directly cause subsequent collapses. This can for example happen when the parties have similar risk exposure, and adapt to each other. An adverse shock to a firm causes other firms to react to this uncertainty, for example by examining other units they have interests in for similar exposure and withdrawing as soon as possible. This causes liquidity problems or other adverse events, making it even more rational for other firms to withdraw. In this kind of "common shock" situation there is weak causation: no single agent is the cause of the trouble of any other agent, yet there is a correlated pattern leading to an adverse outcome.

Some authors distinguish between rational/information-based systemic risk and irrational/random/non-information based systemic risk. In practice the distinction is blurred and might depend on having different time-horizons, preferences and prior probabilities. There might well be an element of "just world" bias in attempting to separate rational systemic risks (where there would be guilty parties if innocents were harmed) from the more nihilistic random systemic risks. Since markets typically contain agents of varying levels of rationality, assuming only rational systemic risk is also somewhat optimistic.

Contagion channels and connectivity

Federal Reserve Governor Daniel Tarullo described four common channels of risk proliferation[12]:

- Domino effects where the failure of one company causes its creditors to fail, causing their creditors to fail in turn.
- Self-reinforcing fire sales when a product serves as a financial collateral or in markets where participants must post risk-based margin. If a failure to pay causes lenders to seize the collateral and sell it at a distressed price this causes further losses on other holders of the asset, making them fail to post margin or default on their loans.
- Signalling contagion, where the failure of one firm signals to investors that other firms in the same industry or holding similar assets are likely to be in financial trouble. This can be either a rational signal, or a panic reaction (bank runs are an obvious example).
- If a firm provides a unique service with no close substitutes, its failure might block critical functions. This could be a clearinghouse with a monopoly on a settlement services in a particular market, or a supplier of a minor but key ingredient in an industrial process[13].

These channels are commonly recognized, although mitigating them is of course another matter.

The first two channels are also linked to firm or institution size, leading to the too-big-to-fail (TBTF) syndrome (the last channel might also be described as "too-connected-to-fail"). TBTF is also relevant not just as a description of the current state of the system but an explanation of why this kind of risk emerges. Since a firm regarding itself as TBTF expects bailouts if it is in trouble, it hence has less incentive to reduce its risks, underprices them, and may gain economic advantages that further helps establish its dominance[14]. Even without this moral hazard dynamics, TBTF can of course be produced through economics of scale and scope, path dependence and oligopolies etc.

Note that each of the channels may in themselves not be enough to cause a system failure, but a combination of them can have multiplicative effects to worsen the situation.

Clearly high connectivity, either through direct financial links or through signalling, enables risk proliferation through these channels. Indeed, without connectivity the risk is systematic rather than systemic.

The link between system connectivity, contagion and systemic risk is an overarching theme that links much of the work in the financial literature to the non-financial systemic risk literature. Concepts from network theory, percolation, and epidemiology may be useful for analysing the risks and in particular under what conditions they can spread across the system.

[12] (Tarullo 2011)

[13] An interesting example is the 1993 fire at Sumitomo Chemical Company which destroyed half of the world's production of a resin used in chip casing manufacturing, causing a worldwide spike in memory prices.

[14] This has been studied both theoretically and empirically. See for example (Warburton et al. 2013), (Dam & Koetter 2011)

Defining systemic risk

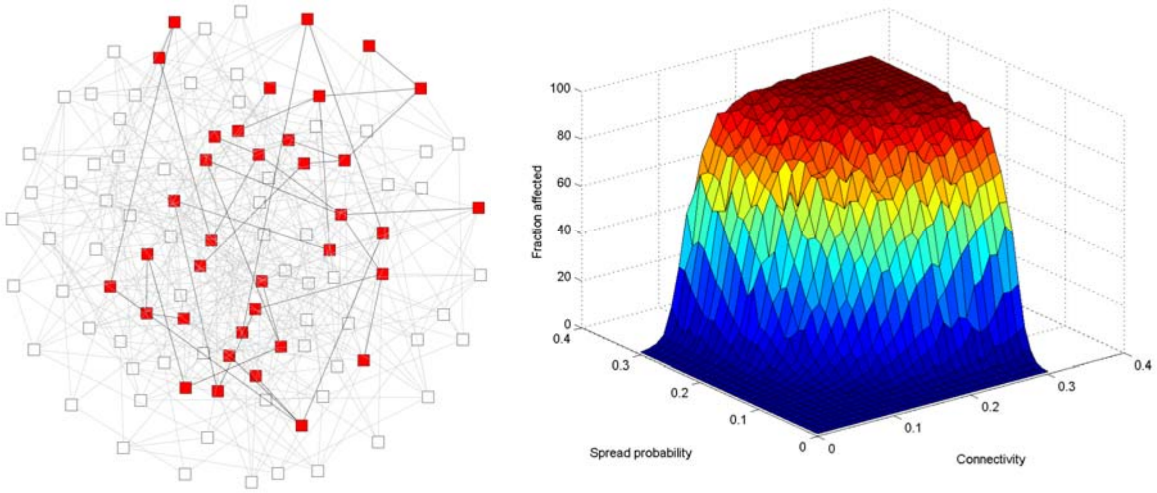


Figure 2: Simple model of cascades in a financial system. Nodes (companies) are randomly connected to each other with a fixed connection probability. One node fails, and connected nodes of a failed node have a probability of failing too. The surface shows the mean fraction of nodes in the failure cascade. For low probabilities cascades rarely get beyond the initial failure; as the product of connectivity and spread increases cascades grow to encompass most of the system. A densely connected system needs to have a very low contagion probability to remain stable.

Densely connected systems easily develop collective behaviours, and some of these may pose risk. Figure 2 shows a simple model of a financial system with cascading failures. As the connectedness and contagiousness increase it has a fairly sharp transition to a vulnerable state where an initial failure is likely to cover most of the system. In this simple case the system can be analysed thoroughly using percolation theory. More complex models may lack the same theoretical clarity but often exhibit the same universal behaviour.

For example, Arinaminpathy, Kapadia and May studied systemic risks in bank lending networks[15]: each node in the network would correspond to a bank, and each edge interbank lending. If a bank goes insolvent its creditors may fail too if the shock is larger than their capital reserves. The contagiousness of failures depends on the proportion of assets in interbank loans, but also market confidence and fire sales of assets. They found that when the system was stressed by external events cascades naturally emerged, with thresholds not dissimilar from the above simple model. But more importantly, the role of large and small banks in the market could be analysed: failures of large banks tended to cause systemic collapse because of their high connectivity and large effect on the market, but as long as they were adequately capitalized they helped protect the system from failures of smaller banks. Simulations may help understand some of the nontrivial interactions between the different contagion channels.

Scope and externalities

Another approach to defining financial systemic risk is the pragmatic one taken by the Financial Stability Oversight Council[16]:

> "There is no single, commonly accepted definition of the term systemic risk among financial professionals. The FSOC annual reports address the definition of systemic risk as follows: "Although there is no one way to define systemic risk, all definitions attempt to capture risks to the stability of the financial system as a whole, as opposed to the risk facing individual financial institutions or market participants." Possible features of systemic risks include externalities and the fallacy of composition. With externalities, there are costs or benefits of actions by financial market participants that are not borne by those participants. With fallacies of composition, what is true for each individual firm in isolation may not be true when all firms follow similar strategies—just as while one person standing in a crowded stadium sees

[15] (Arinaminpathy, Kapadia & May 2012)

[16] (Murphy 2013)

> better, that strategy will fail if everyone stands at the same time."

In this case externalities go far outside the common usage of systemic risks. They can be viewed as risks caused by the market to outsiders, a form of external system risk (see below), although of course externalities do occur inside markets too. The fallacy of composition points at the important emergent aspect of systemic risks that make them both scientifically interesting and hard to regulate (see the generalised systemic risk section).

Some pragmatic definitions stretch the concept in the direction of scope instead. The G10 Report on Consolidation in the Financial Sector 2001[17] suggested a working definition that has been widely cited:

> "*Systemic financial risk* is the risk that an event will trigger a loss of economic value or confidence in, and attendant increases in uncertainly about, a substantial portion of the financial system that is serious enough to quite probably have significant adverse effects on the real economy."

The report argues that the negative externalities must deal with the real economy: a pure financial disaster is not a systemic risk by this definition. It also suggests distinguishing between the direct impact and transmission/contagion effects and their "width" (the fraction of firms or markets simultaneously affected by the impact) and "depth" (the fraction of firms or markets subsequently affected by the shock during the transmission phase):

> "Thus, a systemic financial risk event can be viewed as a shock whose impact and transmission effects are wide and deep enough to severely impair, with high probability, the allocation of resources and risks throughout the financial and real economic systems."

Although the definition aims more at defining systemic risks as a practical risk for policymakers to handle (and why they need to handle it), it again points at the key role played by contagion/transmission effects and how they overspill from the original recipients of the shock – including overspill into other systems. Narrowing it to only risks that affect the real economy might be relevant for the purposes of G10, but appears unnecessarily narrow in practice: to a financial company the risk of a purely inter-finance collapse is still a systemic risk.

Socio-technical systems

A variant of this can be seen in the definitions following from the 2003 OECD Emerging Systemic Risks report:

> "A systemic risk, in the terminology of this report, is one that affects the systems on which society depends – health, transport, environment, telecommunications, etc."

Here the focus is completely on the target of the risk, not the dynamics or causes of the problem. Systemic risks are simply those affecting important systems. Another very wide definition is found in[18] where the key aspect is stated to be that the risks are embedded within complex social, cognitive and ethical systems. These definitions can apply to nearly all important risks, diluting the concept of systemic risk far beyond usability[19].

[17] https://www.imf.org/external/np/g10/2001/01/Eng/pdf/file3.pdf

[18] (Renn & Klinke 2004)

[19] However, (Renn & Klinke 2004) does tackle the epistemic problems of uncertain and unpredictable risks affecting large-scale systems: terminological loseness is not necessarily a sign of irrelevance.

Defining systemic risk

Systemic, systematic and system risk

Systematic risk: when the environment is at fault

To maximize confusion, *systematic* risk in finance and economics is a separate concept from *systemic* risk. Systematic risk (aggregate risk, market risk or undiversifiable risk) represents risk that is shared across the market due to events that affect overall market returns, income or other factors. It is linked to correlated forcings, whether due to disasters, lack of market completeness or merely correlated inputs. For example, a war, recession or a change in interest rate affects most of the market. Systematic risk is largely an exogenous risk: it is not so much the markets "fault" (although different markets may of course be differently vulnerable to systematic risks because of their structure).

This has an ecological analogy in the form of the Moran Effect, which states that separate populations of the same species will fluctuate in a correlated fashion because of correlated environmental fluctuations (for example weather). That the same bad winter causes the collapse of several populations does not mean the collapses caused each other: there was an exogenous effect. A systemic ecological risk would be more like the spread of a disease or the Allee effects, which makes populations close to extinction more vulnerable because of their low number.

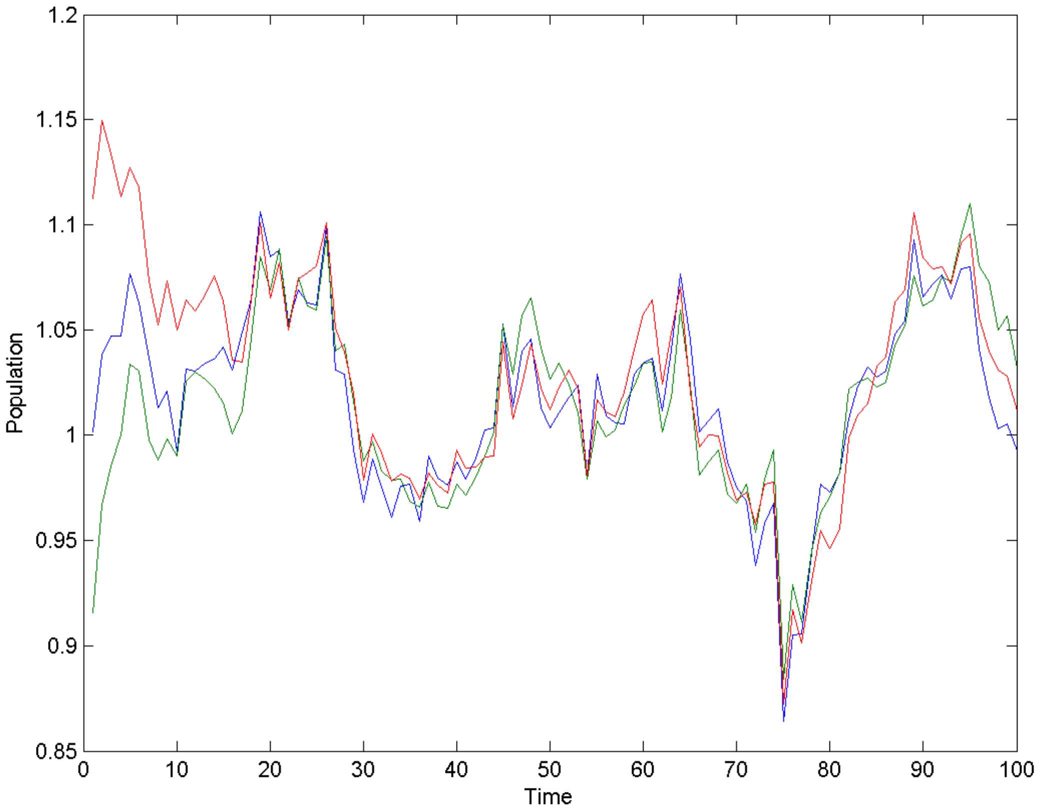


Figure 3: Example of the Moran effect: three separate populations of animals are simulated subject to an environment where the carrying capacity randomly varies in a correlated fashion. Despite each population being separate they end up strongly correlated because of the exogenous input.

Clearly exogenous forcings can make endogenous systemic risks worse: a vulnerable market or population has less ability to weather an adverse period. From a theoretical perspective the most interesting systemic risk case is when collapse can occur without any external forcing, emerging only from internal dynamics (e.g. market crashes triggered by the market generating its own news). From a practical perspective understanding how large the zone of resilience is matters more: exogenous forcings are always present besides the internal dynamics. The key question is what properties make a system vulnerable to collapse in response to a forcing that normally (or in the past) would have been manageable, and whether these properties can be detected and corrected before disaster.

System risk: when the structure is at fault

One possibly valuable conceptual distinction is between *systemic*, *systematic* and *system* risk.

A risk may be due to the structure of a system, in which case it makes sense to call it a system risk.

A classic form of system is risk is single point of failure: one step or function in a larger system is necessary for its overall function (e.g. the heart in the body, or a router in a computer network). Much reliability engineering deals with detecting and removing potential single point failures. Unfortunately in complex evolving organisations and technical systems it can be hard to notice all dependencies. This can evolve into full systemic risk, but it can merely remain a bottleneck. Bottlenecks that impair function visibly tend to get corrected, but true system risks introduce hard-to-notice problems (unknown increase in overall risk level, reduction of information quality, subtle inefficiencies) that only become apparent in retrospect.

Many system risks do not reach the level of threat implied by a systemic risk. Systemic risks are a subset of system risks that have additional properties – the ability to break the entire system, contagion, or other severe hazards. Still, mere system risks can be important in that they cause costs, friction or lost opportunities.

Some systems exhibit "robust-yet-fragile" dynamics: they are highly robust against random disturbances but are vulnerable to attacks on certain key components[20]. This makes systems particularly vulnerable to malicious intent, but may also hide their inherent fragility in unusual situations.

Error correction can make the difference between a system risk and a systemic risk. Lloyd's solution to the LMX spiral is an example: they contained the system risk and rebooted the system. Any method that allows the system to recover (e.g. higher capital requirements, a resetting function) could fill that role. Conversely, when error correcting abilities are removed, system risk can become immediately systemic.

One ironic effect of increasing central coordination in order to reduce systemic risk is that it introduces a single node (the regulator) connected to every node in the system. A failure of the regulatory node can introduce correlated risk across the system.

To sum up, systematic risks are exogenously generated risks, systemic risks are endogenously generated risks or properties of a system that can make a small exogenous event snowball, and system risks are misfeatures of the system that even if they do not cause a crisis, reduce the functionality of the system.

Still, given the potential for confusion we do not recommend using the term without clearly explaining it.

[20] (Albert, Jeong & Barabási 2000),(Doyle et al. 2005)

Defining systemic risk

Complex systems theory and systemic risk

The generalised use of systemic risk borrows from the vision of complex systems theory (which also influenced resiliency studies) that many complex adaptive systems have fundamental similarities despite their very different appearances. This in turn was stimulated by profound 20th century mathematical results about renormalization, structural stability, and universality in chaotic systems: surprisingly large classes of apparently unpredictable phenomena are governed by the same simple laws despite their exact functional form being vastly different[21].

The complex system view emphasizes the endogenous aspects of systemic risks. Complex systems typically have many parts interacting in nonlinear ways that can include positive feedback loops and internal adaptation. This produces potentially paradoxical behavior such as random or chaotic activity (limiting predictability), multiple attractor states, and sudden jumps and/or resistance to external control[22]. The complex system view of systemic risk emphasizes the internal dynamics of the system emerging from the structure of its internal workings rather than what kind of entities these are.

Ian Goldin and Tiffany Vogel approached systemic risk and risk governance from a complex systems angle[23]:

> "While systemic risk has been seen as a threat caused by unpredictable, highly improbable, exogenous stochastic events (Albeverio et al., 2006; Taleb, 2007), we see systemic risk as reflecting endogenous structural weakness"
> "While these networks involve the transmission of materials, capital, information and knowledge, recent decades of intense global integration mean that these highly interconnected networks also have the potential to originate and propagate risk. This central property of interconnectedness in networks (Jervis, 1997) can be paradoxical in both its structure and impacts. Increasingly connected networks facilitated by globalization can lead to both greater robustness and more fragility"

They use "robustness" in a somewhat idiosyncratic way to denote spreading risks across a more interconnected network. This way the overall ability to diffuse risks go up, but the global correlations also increase, making the whole system more brittle.

This definition points out one of the more counterintuitive and relevant problems with systemic risk management: rational actions in managing risks can increase overall risk. A similar observation of the paradoxical nature of handling systemic risk was made by Doyne Farmer[24]:

> "Systemic risk occurs when individual actors unknowingly create risk through their systemic interactions with each other (which they are often unable to model). It often occurs precisely due to their attempts to lower their own risk."

A typical example is how reliability of power-grids has increased in recent years. However, the size distributions of blackouts are power-law distributed and have shifted towards fewer, but much larger, blackouts. Reducing the number of forest fires leads to build-up of undergrowth that can sustain worse fires. Regulations reducing risk in a domain may itself lead to risk. Increasing agent coordination might allow them to model their interactions better, but could produce stronger linkages for other systemic risks. And so on.

Attractor states

Darryll Hendricks[25] suggests a more theoretical definition from the sciences (besides a near-copy of the G10 definition)

> "A *systemic risk* is the risk of a phase transition from one equilibrium to another, much less optimal equilibrium, characterized by multiple self-reinforcing feedback mechanisms making it difficult to reverse."

While equilibrium is a common term in ecology and economics, strictly speaking we are talking about

[21] (Cvitanović 1989), (Sornette 2003)
[22] (Helbing 2009), (Cleeland 2011)
[23] (Goldin & Vogel 2010)

[24] http://ineteconomics.org/sites/inet.civicactions.net/files/INETSession2-Lunch-Farmer_0.pdf
[25] http://www.pewtrusts.org/uploadedFiles/wwwpewtrustsorg/Reports/Economic_Mobility/PTF-Note-1-Defining-Systemic-Risk.pdf?n=3489

attractor states: states the system tends to return to when disturbed. This can include time-varying states, noisy states or complex "strange attractors".

This definition leaves out the causality of what caused the transition, but points out that once it has begun the structure of the system makes it hard to reverse. There is an asymmetry (hysteresis) between the states. A classic example is the collapse of the Grand Banks cod fisheries, where overfishing caused a transition to a different ecological state that persists even when fishing stops.

Another ecological example is the impact of removing otters from the North Pacific kelp ecosystem. The otters were hunted because they were seen as competing with humans for abalone, but were also a major predator of sea urchins. Without them the urchins multiplied, killing the kelp that was the basis for the rich ecosystem. This reduced productivity and diversity precipitously. Sea stars took the role as the main predator of urchins and a new stable but relatively barren ecosystem emerged.

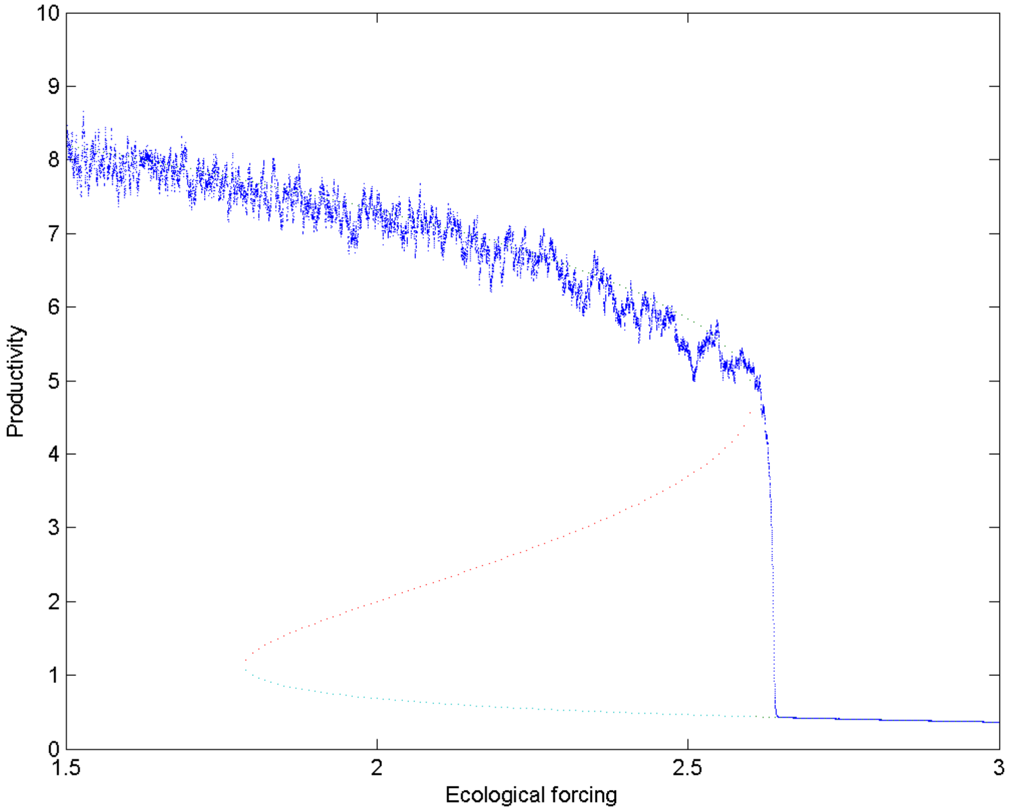


Figure 4: Switch of attractor state in a simple ecological model subject to random noise and a gradually worsening external forcing. The blue curve represents the productivity of the ecosystem, the dotted green and red curves represent the equilibria without noise. As the forcing increases the ecosystem suddenly falls into the low productivity state. Note the slowing and coarsening of fluctuations around the equilibrium as the transition is approaches.

Whether the new equilibrium is less optimal may be a matter of judgement (sea urchins presumably prefer an otter-free environment) but the mere fact that the equilibrium is hard to escape might be disadvantageous. Complex ecosystems, minds and financial markets have many potential modes of function and can shift adaptively between them; being stuck in a single state is a sign of inflexibility – there are no alternative if future even larger shocks occur.

In the ecological resiliency literature resiliency is sometimes defined as either the amount of disturbance that a system could withstand without changing to an alternative stable state, or the speed with which the system returns to a stable state after disturbance[26]. These are often correlated

[26] Beside these two definitions, resiliency can be defined in terms of adapting to the disturbances. See (Gunderson 2000)

to each other, but the first represents the size of the basin of attraction of a stable state, the second its "depth". Systemic risk in this case can then be seen as a system or parameters causing small, easily escaped attractor states[27]. As the system is pushed close to the edge of the basin recovery dynamics tends to slow down ("critical slowing down") and the variance and autocorrelation of fluctuations tend to increase. These signs are potential indicators of the approach to a tipping point[28], although they might not necessarily occur for all kinds of systems.

While ecosystems and environmental science might be the starting point for a large part of the natural science discourse on systemic risk, many results do look generalizable to other domains. In particular the celebrated result by Robert May that complex ecosystems in general are unstable[29] does not make use of particular properties of them being ecosystems: the result is general for many forms of interacting systems, and has recently been applied to banking[30].

Statistical mechanics of critical systems

Complex adaptive systems exhibit events on all size scales. One reason is that they are often hierarchical structures (e.g. an economy contains markets containing companies owning assets) but there are also phenomena that naturally tend to scale-free statistical distributions such as power-laws[31]. The canonical (and sometimes over-used) example is the "sandpile models" of self-organized criticality[32]: the system consists of locations where tension (the height of a pile of objects, seismic strain, financial risk) can build up. If it reaches a critical threshold in a location it is released, resetting the location (the pile collapses, the fault moves, a company goes bankrupt) but distributing some or all of the tension to the neighbouring locations. The result is cascades of release. Over time, if tension is continually added to the system, it reaches a state where on average as much tension is released as is added per unit of time. However, the cascades releasing the tension come in all sizes from the smallest to cascades affecting the whole system. Their frequency is typically a power law of their size: very large adjustments may occur rarely, but there is no size cut-off besides the size of the entire system. These cascades are also triggered by micro-events rather than strong external disturbances.

[27] Compare to engineering, where designs typically have safety factors so that even unusually strong disturbances are within design limits and will be dampened out by the construction. More comprehensive systems safety engineering will also seek to design systems so that the dynamics avoids hazards and possible end-states are acceptable.

[28] (Scheffer et al. 2009)

[29] (May 1972)

[30] (Haldane & May 2011)

[31] This includes the Gutenberg-Richter law of earthquake frequency as function of magnitude, frequencies of wildfires (Malamud, Millington & Perry 2005), asteroid impacts (Chapman & Morrison 1994), landslides (Malamud et al. 2004), city fires (Song et al. 2003) and many other systems.

[32] (Bak, Tang, Wiesenfeldt 1987)

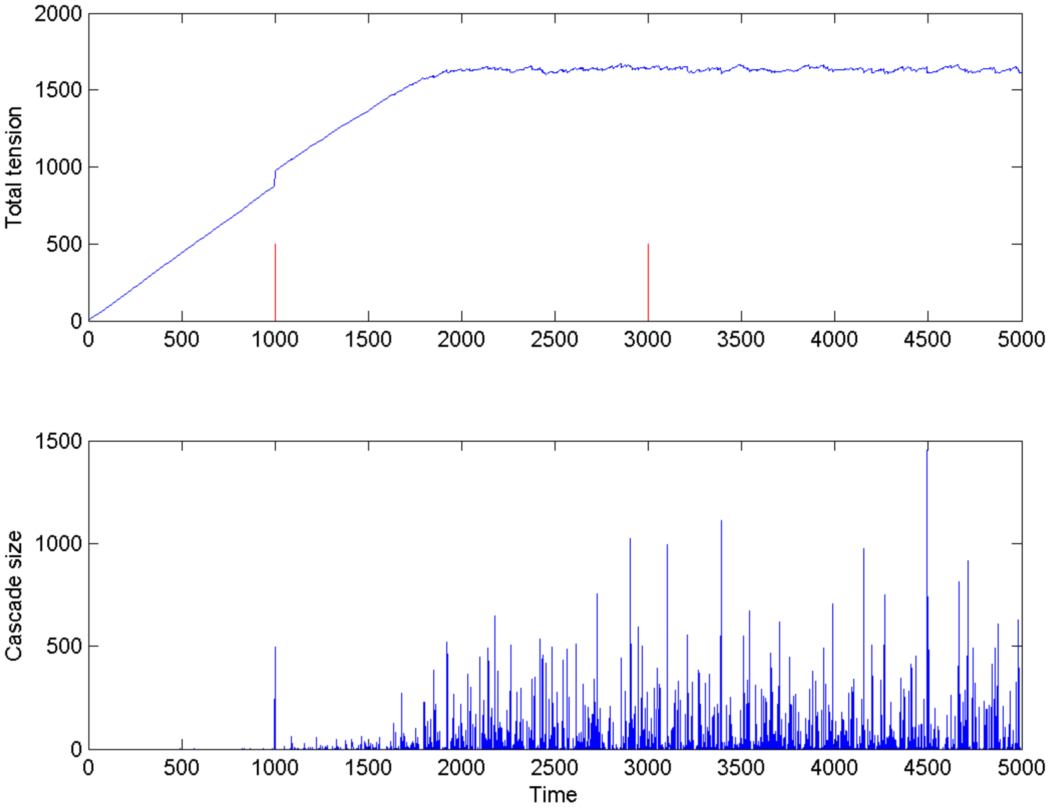


Figure 5: Plot of the dynamics of the "sandpile model" over time. Top: starting from an empty state with no tension it gradually fills out reaching a dynamically stable state where the influx of tension is equal to the release. Bottom: size of the cascades. As the system reaches equilibrium they form a scale free distribution. At time 1000 and 3000 an exogenous "kick" to the system occurs, adding extra tension in a random location. In the early case much of the tension remains in the system (the jump in the top curve). In the late case the tension is successfully diffused – but the price is ongoing endogenously generated cascades.

Critical systems of this kind have been extensively studied in the sciences and may provide insights into systemic risks. In a sense systemic risks in this kind of system are natural: they serve to release tension just as much as the smaller scale cascades. Just like in the wildfire example (which can indeed be modelled this way) removing small events increases the probability of large events[33]. A human might wish to change the system to reduce the incidence of large events, but must then work against the implications of the underlying microdynamics[34].

Didier Sornette has analysed the difference between endogenous and exogenous shocks in critical systems[35]. He found that at least in some systems (for example, book sales, financial volatility, financial crashes) there are clear differences in dynamics: externally induced shocks occur suddenly and then exhibit a power-law decay as the system returns to a stable state. Endogenously generated shocks on the other hand exhibit a power-law lead up as more and more microevents avalanche. Once the peak is reached the system shows a symmetric return to stability as the avalanche peters out.

[33] One might even make an analogy with Schumpeter's concept of "creative destruction": by continually weeding out weakly functioning entities the overall system maintains its vitality.

[34] There might also exist optimal levels of interconnectivity or other parameters that make the system as efficient as possible without too disruptive cascades. See for example (Brummitt, D'Souza & Leicht 2012)

[35] http://www.er.ethz.ch/essays/origins

Defining systemic risk

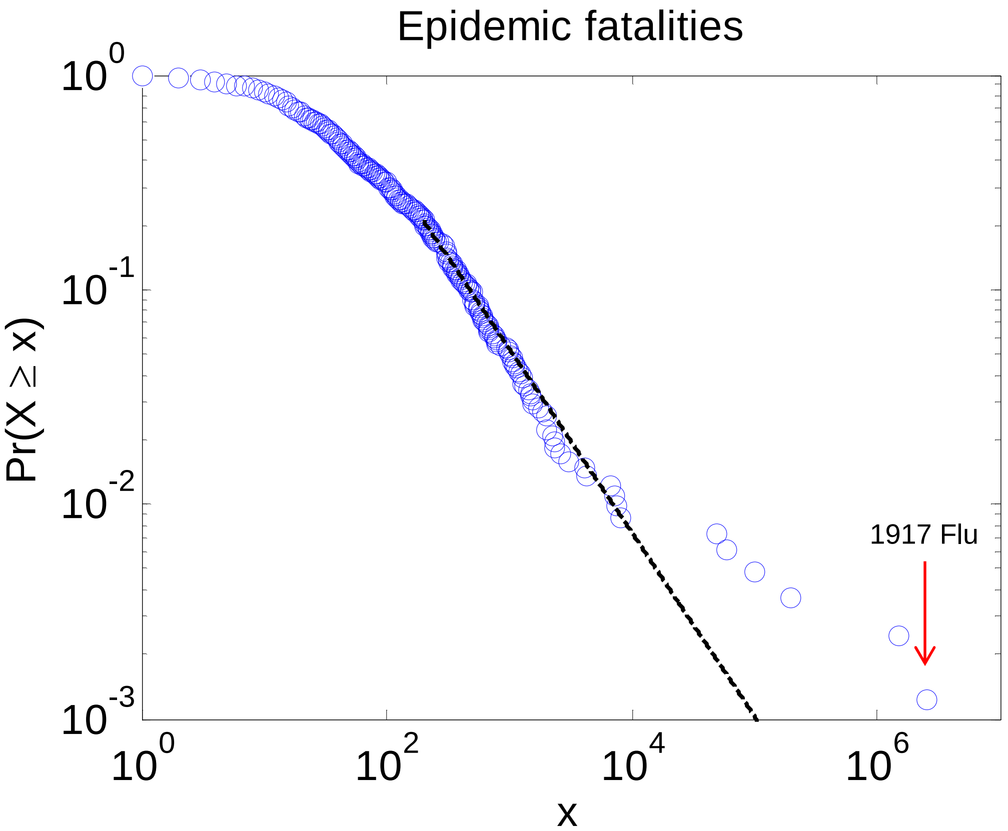


Figure 6: Cumulative probability distribution of epidemic fatalities. The distribution fits a power-law reasonably well in the mid-region. For major pandemics the distribution exhibits a break towards a different dynamics, possibly a sign of the "dragon-king" phenomenon.

Another property of certain critical systems is the existence of "dragon-kings" (a term coined by Sornette): rare extreme events that are much larger than expected even based on a power-law or extreme value distribution[36]. They typically reveal a different form of underlying dynamics than the run-of-the-mill events, for example a phase transition or the introduction of a previously non-existent dynamics. For example, acoustic events in a strained metal bar have a power-law distribution, together with a dragon-king outlier for the final rupture. Similarly extreme financial losses may increase correlations in ways that make further losses much more likely than expected from the standard dynamics. Since their occurrence is due to a different kind of dynamic than normal events they may be surprising "black swans" when they first occur.

This critical systems approach to systemic risk is to a large degree descriptive rather than prescriptive, although it might be useful for predicting some behaviours.

Homer-Dixon's hybrid model

An interesting complement to this statistical view is the systems view expressed by Thomas Homer-Dixon[37] in regards to large societal risks. He argued that there are three main risk channels: cascades, multiplicative effects and increased correlations in strained systems. In his model different domains (energy, food, policing, finance, etc.) normally functions fairly independently. They are subject to systemic risks on their own, that can lead to cascades causing problems. However, the actual damage of a crisis depends on how well other domains function: if healthcare functions well, the impact of a terrorist attack is somewhat reduced; a good economy can manage a food shortage by increased imports. Conversely, if other domains fail at the same time the damage increases. There is a multiplicative effect of simultaneous failures.

If failures are uncorrelated this is not a major issue, but he argued that in many recent crises correlations have increased before they struck.

[36] (Sornette 2009)

[37] (Homer-Dixon 2011)

The reason is that domains that are under strain due to exogenous factors or ongoing micro-crises require resources or support from other domains, increasing the correlation between them. In an energy- or climate-constrained world biofuels become more important, strengthening the correlation between fluctuations in food and energy prices (especially since energy is a major factor in food production costs). This means that a major failure in one domain can now trigger failure in the other (in this example an energy crisis causing a food production crisis) and this leads to a lateral cascade to further domains (food scarcity causing political unrest and economic trouble) – while the damage is multiplicative.

This form of hybrid systemic risk model shows the importance of examining not just in-system risk channels, but also weak connections to other systems that may become stronger – especially if their strength can change surprisingly fast.

Conclusions

In describing the diverse views of systemic risk among participants at a conference cosponsored by the Federal Reserve Bank of New York and the National Academy of Sciences, the authors of the conference report[38] noted that:

> "An adage among traders is that, in times of crisis, everything is correlated. Though conference participants did not share a consensus on the definition of systemic risk, the descriptions of systemic events by risk managers at the conference reflected this view"
> "Under such regime shifts, the normal assumptions culled from historical experience that guide fay-to-day trading break down."
> "In the tentative vocabulary of systemic risk suggested above, the self-reinforcing uncertainty and market panic that can characterize a systemic episode are a clear example of contagion. The jump in correlations appearing at the onset of a systemic event can in turn be seen as an example of self-reinforcing feedback and synchrony. Furthermore, the transition from a normally functioning market to one in which prices are generated by the internal market microstructure is accompanied by widespread and simultaneous liquidations. Financing constraints and the loss of liquidity make a return to the pre-crisis state very difficult - an asymmetrical transition and example of hysteresis."

This neatly fits in the various components of the systems view of systemic risk: as a transition is approached, critical slowing or increased linkage occurs, increasing correlations. As the dynamics changes, old experience becomes inapplicable. Failures begin to spread in the institutional network, both along direct links and due to information cascades. The end result is a transition to another, possibly worse state.

[38] (Board on Mathematical Sciences and Their Applications, Division on Engineering and Physical Sciences, National Research Council 2007)

Defining systemic risk

From the perspective of finance the relevant aspect is that an unpredictable external disturbance or single failure initiates a snowballing cascade of disruption affecting all actors within the market. The end result is a situation where often outside intervention is needed to stop the process or prevent the new state from harming other markets or institutions.

- The term “systemic risk” is used in several meanings, sometimes with fairly deep differences.
- There is however a common core idea that parts that individually may function well when connected to a system become vulnerable to a joint risk that can spread from part to part, potentially affecting the entire system and possibly related outside systems.
- The originating causes of the eventual disaster may be exogenous or endogenous.
- Systematic risks are separate from systemic risks, but can trigger them.
- System risks come from the structure of the system, but are risks that merely cause emergent adverse outcomes, not the contagion effects seen in systemic risks proper.
- Connections between the parts of systems vulnerable to systemic risk are often strong, but the connection does not have to be obvious – in many cases shared correlations and exposures suffice to make risks systemic.
- The key problem is that much of the risk comes from the structure of the system, and this is often constrained historically, legally, economically or practically: changing the structure strongly is often not feasible.
- Adaptation and risk mitigation (including regulation) are not separate from the system, and can increase systemic risks.

When discussing systemic risks the specific meaning of the term used should be stressed, to reduce risk of misunderstanding.

References

Julian Adams. Remarks on Globalization and Systemic Risk: Nonbank Financial Intermediaries. In Globalization And Systemic Risk World Scientific Studies in International Economics 2009

Réka Albert, Hawoong Jeong & Albert-László Barabási, Error and attack tolerance of complex networks, Nature 406, 378-382 (27 July 2000)

Nimalan Arinaminpathy, Sujit Kapadia, and Robert M. May. Size and complexity in model financial systems. PNAS 2012 109 (45) 18338-18343

Bak, P., Tang, C. and Wiesenfeld, K. (1987). "Self-organized criticality: an explanation of 1/f noise". Physical Review Letters 59 (4): 381–384. Bibcode:1987PhRvL..59..381B. doi:10.1103/PhysRevLett.59.381

Olivier de bandt & Philipp Hartmann. Systemic risk: a survey. ECB Working paper no. 35 2000 https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp035.pdf

Battiston, Stefano, et al. "Liaisons dangereuses: Increasing connectivity, risk sharing, and systemic risk." *Journal of Economic Dynamics and Control*36.8 (2012): 1121-1141.

Ulrich Beck (1992). Risk Society: Towards a New Modernity. London: Sage Publications

Monica Billio, Mila Getmansky, Andrew W. Lo, Loriana Pelizzon, Econometric measures of connectedness and systemic risk in the finance and insurance sectors, Journal of Financial Economics, Volume 104, Issue 3, June 2012, Pages 535-559, ISSN 0304-405X, http://dx.doi.org/10.1016/j.jfineco.2011.12.010

Board on Mathematical Sciences and Their Applications, Division on Engineering and Physical Sciences, National Research Council. Eds. John Kambhu, Scott Weidman, Neel Krishnan. New Directions for Understanding Systemic Risk: A Report on a Conference Cosponsored by the Federal Reserve Bank of New York and the National Academy of Sciences. National Academies Press, 2007

Charles D. Brummitt, Raissa M. D'Souza, and E. A. Leicht. Suppressing cascades of load in interdependent networks PNAS 2012 109 (12) E680-E689

James Bullard, Christopher J. Neely, and David C. Wheelock. Systemic Risk and the Financial Crisis: A Primer. FEDERAL RESERVE BANK OF ST. LOUI S REVI EW SEPTEMBER/ OCTOBER, PART 1 2009 403

Myriam Dunn Cavelty. Systems at Risk as Risk to the System. Limn No. 1 http://limn.it/systems-at-risk-as-risk-to-the-system/

Chapman, C. R. & Morrison, D. (1994) Nature 367 , 33-40.

Cont, Rama, Amal Moussa, and Edson Santos. "Network structure and systemic risk in banking systems." *Edson Bastos e, Network Structure and Systemic Risk in Banking Systems (December 1, 2010)* (2011).

Belinda Cleeland, Contributing Factors to the Emergence of Systemic Risks, Technikfolgenabschätzung – Theorie und Praxis 20. Jg., Heft 3, Dezember 2011

J. David Cummins and Mary A. Weiss, Systemic Risk and Regulation of the U.S. Insurance Industry, Networks Financial Institute Policy Brief, 2013-PB-02 March 9, 2013 http://indstate.edu/business/nfi/leadership/briefs/2013-PB-02_Cummins-Weiss.pdf

Predrag Cvitanović, Ed., Universality in Chaos, Taylor & Francis; 2nd ed 1989.

Lammertjan Dam, Michael Koetter, Bank bailouts, interventions, and moral hazard. Deutsche Bundesbank Research Centre, Discussion Paper Series 2: Banking and Financial Studies No 10/2011 http://www.econstor.eu/bitstream/10419/50000/1/667625526.pdf

Barbara J Downes, Fiona Miller, Jon Barnett, Alena Glaister and Heidi Ellemor. How do we know about resilience? An analysis of empirical research on resilience, and implications for interdisciplinary praxis. Environ. Res. Lett. 8 014041 doi:10.1088/1748-9326/8/1/014041

John C. Doyle, David L. Alderson, Lun Li, Steven Low, Matthew Roughan, Stanislav Shalunov, Reiko Tanaka, and Walter Willinger The “robust yet fragile” nature of the Internet. PNAS 2005 102 (41) 14497-14502

Defining systemic risk

European Central bank (2004) Annual Report 2004, ECB, Frankfurt, Germany

Michael Faure, T. Hartlief Insurance and Expanding Systemic Risks Policy Issues in Insurance, No. 5 2003

FSB, IMF, BIS (2009), Report to G20 Finance Ministers and Governors “Guidance to Assess the Systemic Importance of Financial Institutions, Markets and Instruments: Initial Considerations”

Ian Goldin and Tiffany Vogel, "Global Governance and Systemic Risk in the 21 st Century: Lessons from the financial crisis", Global Policy, Vol. 1, No. 1., 4-15, January 2010

Alan Greenspan, Remarks at a Conference on Risk Measurement and Systemic Risk,

Board of Governors of the Federal Reserve System (Nov. 16, 1995)), via Schwarcz

Gunderson, L.H. (2000). "Ecological Resilience — In Theory and Application". Annual Review of Ecology & Systematics 31: 425

Group of Ten. Report on Consolidation in the Financial Sector. Chapter III: Effects of consolidation on financial risk January 25, 2001 https://www.imf.org/external/np/g10/2001/01/Eng/pdf/file3.pdf

Group of Thirty, Reinsurance and International Financial Markets, 2006 http://www.group30.org/images/PDF/Reinsurance%20and%20International%20Financial%20Markets.pdf

Andrew G. Haldane & Robert M. May, Systemic risk in banking ecosystems, Nature 469, 351–355 (20 January 2011)

Dirk Helbing, Systemic Risks in Society and Economics. SFI WORKING PAPER: 2009-12-04.

Hendricks, Darryll. 2009. "Defining Systemic Risk." The Pew Financial Reform Project.

Holling, C.S. (1973). "Resilience and stability of ecological systems". Annual Review of Ecology and Systematics 4: 1–23

Homer-Dixon, Thomas. "Complexity Science." *Oxford Leadership Journal* 2.1 (2011): 15.

International Monetary Fund,Global Financial Stability Report, April 2009, p. 113.

Andreas A. Jobst, 2012, "Systemic Risk in the Insurance Sector-General Issues and a First Assessment of Large Commercial (Re-)Insurers in Bermuda," Working paper (March 14).

George G. Kaufman. Banking and currency crises and systemic risk: Lessons from recent events. Economic perspectives. 2000 Q III p. 9-28. Federal Reserve Bank of Chicago

George G. Kaufman & Kenneth E. Scott. What Is Systemic Risk, and Do Bank Regulators Retard or Contribute to It? The Independent Review, v. VII, n. 3, Winter 2003, ISSN 1086-1653, 2003, pp. 371– 391. http://www.independent.org/pdf/tir/tir_07_3_scott.pdf

Patrick M. Liedtke, The lack of an appropriate definition of systemic risk, Insurance and Finance No. 6, July 2010 https://www.genevaassociation.org/media/78823/ga2010-if06-liedtke.pdf

John Lipsky, Through the Looking Glass: The Links between Financial Globalization and Systemic Risk.In Globalization And Systemic Risk World Scientific Studies in International Economics 2009

Bruce D. Malamud, James D. A. Millington, and George L. W. Perry. Characterizing wildfire regimes in the United States. PNAS 2005 102 (13) 4694-4699; published ahead of print March 21, 2005, doi:10.1073/pnas.0500880102

Malamud, B. D., Turcotte, D. L., Guzzetti, F. & Reichenbach, P. (2004) Earth Surf. Processes Landforms 29 , 687-711.

Markose, Sheri, et al. "Too interconnected to fail: Financial contagion and systemic risk in network model of cds and other credit enhancement obligations of us banks." *analysis* (2009).

Robert May, Will a Large Complex System be Stable? Nature 238, 413 - 414 (18 August 1972)

May, Robert M., & Nimalan Arinaminpathy. "Systemic risk: the dynamics of model banking systems." *Journal of the Royal Society Interface* 7.46 (2010): 823-838.

Edward V. Murphy. Financial Stability Oversight Council: A Framework to Mitigate Systemic Risk. Congressional Research Service 5-5700 www.crs.gov R42083 May 21, 2013 http://www.fas.org/sgp/crs/misc/R42083.pdf

OECD (2003) Emerging Systemic Risks. Final Report to the OECD Futures Project. Paris, France: OECD.

Property Casualty Insurers Association of America, Systemic Risks Defined. 2009

G.L.L. Reniers, K. Sörensen, W. Dullaert, A multi-attribute Systemic Risk Index for comparing and prioritizing chemical industrial areas, Reliability Engineering & System Safety, Volume 98, Issue 1, February 2012, Pages 35-42

Renn O, Klinke A (2004) Systemic risks: a new challenge for risk management. EMBO Rep 5(S1):S41–S46 http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1299208/

María Rodríguez-Moreno, Juan Ignacio Peña, Systemic risk measures: The simpler the better?, Journal of Banking & Finance, Volume 37, Issue 6, June 2013, Pages 1817-1831, ISSN 0378-4266, http://dx.doi.org/10.1016/j.jbankfin.2012.07.010.

Marten Scheffer, Jordi Bascompte, William A. Brock, Victor Brovkin, Stephen R. Carpenter, Vasilis Dakos, Hermann Held, Egbert H. van Nes, Max Rietkerk & George Sugihara. Early-warning signals for critical transitions. Nature 461, 53-59 (3 September 2009

Schwarcz, Steven L., Systemic Risk. Duke Law School Legal Studies Paper No. 163; Georgetown Law Journal, Vol. 97, No. 1, 2008. Available at SSRN: http://ssrn.com/abstract=1008326

W.G. Song, H.P. Zhang, T. Chen, W.C. Fan. Power-law distribution of city fires. Fire Safety Journal 38 (2003) 453–465

Didier Sornette, Critical Phenomena in Natural Science. Springer 2003

Didier Sornette. 2009. Dragon-Kings, Black Swans and the Prediction of Crises. International Journal of Terraspace Science and Engineering 2(1): 1-18 http://arxiv.org/abs/0907.4290

Martin Summer. Quantitative Modeling of Systemic Risk in a Globalized Banking System. In Globalization And Systemic Risk World Scientific Studies in International Economics 2009

Nassim Taleb, *The Black Swan*, Allen Lane 2007

Daniel K. Tarullo, “Regulating Systemic Risk,” Speech, 2011 Credit Markets Symposium, North Carolina, Charlotte, March 31, 2011, Board of Governors of the Federal Reserve System. http://www.federalreserve.gov/newsevents/speech/tarullo20110331a.htm

Trainar, P. (2010) “Are (re)insurance operations source of systemic risk?”, Insurance and Finance Newsletter of The Geneva Association, No 6. July 2010

Warburton, A. Joseph and Anginer, Deniz and Acharya, Viral V., The End of Market Discipline? Investor Expectations of Implicit State Guarantees (January 1, 2013).

http://ssrn.com/abstract=1961656 or http://dx.doi.org/10.2139/ssrn.1961656

Appendix: a sampling of definitions of systemic risk

These definitions represent a small set of definitions used in the literature. They can roughly be grouped into (1) a set based on risks affecting the entire market/system and often having spillover effects outside the system (this is the pragmatic G10 report view, mainly focused on finance and how to regulate it), (2) definitions stressing contagion, (3) definitions stressing the systemic aspect of the risk, (4) definitions stressing that the risks occur in important system. Whether systematic risk is included or separated varies.

Definition	Outside effects	Affects whole market/system	Contagion	Correlations	System failure	Systematic risk	Suddenness	Size	Endogenousness	Affects important systems	Uncertain/uncontrollable	Other	Source
"Systemic risk" refers to the likelihood and degree of negative consequences to the larger body. With respect to federal financial regulation, the systemic risk of a financial institution is the likelihood and the degree that the institution's activities will negatively affect the larger economy such that unusual and extreme federal intervention would be required to ameliorate the effects."	X	X											(Property Casualty Insurers Association of America 2009)
"We define systemic risk to be the risk of a failure in a transaction or series of transactions extending beyond the parties directly involved, impacting many or most participants in the marketplace. And the public gains awareness of these systemic effects on the larger group only after the breakdown has occurred. "	X	X											American Academy of Actuaries, Concepts for Successful Regulation of Systemic Risk
"Systemic risks are developments that threaten the stability of the financial system as a whole and consequently the broader economy, not just that of one or two institutions. "	X	X											b. Bernanke, in his letter to U.S. Senator Bob Corker, 30 october 2009, as reported in The Wall Street Journal
‘The risk of disruption to financial services that is *(i)* caused by an impairment of all or parts of the financial system and *(ii)* has the potential to have serious negative consequences for the real economy ”	X	X											Financial Stability Board. 2011. Policy measures to address systemically important financial institutions.
"This report adopts the Group of Ten's 2001 definition of systemic risk: "Systemic financial risk is the risk that an event will trigger a loss of economic value or confidence in, and attendant increases in uncertainty about, a substantial portion of the financial system that is serious enough to quite probably have significant adverse effects on the real economy.""	X	X											(Group of Thirty 2006)
"Systemic financial risk is the risk that an event will trigger a loss of economic value or confidence in, and attendant increases in uncertainly about, a substantial portion of the financial system that is serious enough to quite probably have significant adverse effects on the real economy."	X	X											(Group of ten, 2001)
"Systemic risk can be defined as the risk that an event will trigger a loss of economic value or confidence in a substantial segment of the financial system that is serious enough to have significant adverse effects on the real economy with a high probability."	X	X											(Cummins & Weiss 2013
"Systemic risk refers to the possibility that a triggering event, such as the failure of an individual firm, will seriously impair other firms or markets and harm the broader economy. "	X	X							X				(Bullard, Neely & Wheelock 2009)

“Returning to the basic question of defining systemic risk in securitized globalized markets, experience suggests that systemic risk is created by unexpected events that heighten uncertainty sharply and impair market liquidity. Illiquidity leads to “price gaps” in individual markets and in the pricing of specific assets. The associated stress subsequently extends to the funding liquidity of financial institutions across the globe that are supporting those individual markets. Market illiquidity in turn can lead to potentially significant real economic effects, thus justifying policy action, especially by central banks.”	X	X					X				X		(Lipsky 2009)
“There are various definitions of what constitutes “systemic risk”, but virtually all of them have the following common features: the first is a notion of contagion — risk spreading from one firm or sector to another — and the second is that, regardless of its point of origin, a systemic risk should be capable of having a negative impact on the wider economy.”	X		X										(Adams 2009)
“A systemic event is defined as a financial crisis that causes a substantial reduction in aggregate economic activity, such variables as housing starts, home sales, consumption, output and employment. Systemic risk is the possibility that a systemic event may occur. “	X												http://www.fhfa.gov/webfiles/1145/sysrisk.pdf
“The notion of systemic risk is perhaps one of the most popular terms used in connection with the discussion of crises in the banking system, both by regulators and in the academic literature. It is used as a description of many different phenomena as has been pointed out by Dow (2000) and by DeBandt and Hartmann (2000). It is used to describe crises related to the payment system, to bank runs and banking panics, to spillover effects between financial markets up to a very broadly understood notion of financially-driven macroeconomic crises. Despite the lack of a precise definition, when the term, “systemic risk”, is used in connection with the banking system, it seems that most authors have in mind the problem of simultaneous failures of many institutions with significant consequences for the real economy.”	X						X						(Summer 2009)
“Systemic risk appears when generalized malfunctioning in the financial system threatens economic growth and welfare. “ “The causes of this malfunction are multiple and therefore a single measure of systemic risk may neither be appropriate nor desirable. “	X												(Rodríguez-Moreno & Peña 2013)
“The risk that the inability of one institution to meet its obligations when due will cause other intuitions to be unable to meet their obligations when due. Such a failure may cause significant liquidity or credit problems and, as a result, could threaten the stability of or confidence in markets.”		X	X										(European Central bank 2004)
"Systemic risk refers to the risk or probability of breakdowns in an entire system, as opposed to breakdowns in individual parts or components, and is evidenced by comovements (correlation) among most or all the parts."		X	X	X									(Kaufman & Scott 2003)
“the risk or probability of breakdowns (losses) in an entire system as opposed to breakdowns in individual parts or components and is evidenced by comovements (correlation) among most or all the parts.”		X	X	X									(Kaufman 2000)
"the risk that the failure of one financial institution (as a bank) could cause other interconnected institutions to fail and harm the economy as a whole "		X	X										Merriam-Webster Dictionary http://www.merriam-webster.com/dictionary/systemic%20risk

“Systemic risk to financial markets is often defined as the risk of a major and rapid disruption in one or more of the core functions of the financial system caused by the initial failure of one or more financial firms or a segment of the financial system ([3], p. 3.) “		X	X				X		X				Modeling systemic risks in financial markets http://arxiv.org/abs/1311.3764v1
“We define a systemic event in the narrow sense as an event, where the release of "bad news" about a financial institution, or even its failure, or the crash of a financial market leads in a sequential fashion to considerable adverse effects on one or several other financial institutions or markets, e.g. their failure or crash. ... Essential is the "domino effect" from one institution to the other or from one market to the other emanating from a limited ("idiosyncratic") shock. Systemic events in the broad sense ... include not only the events described above but also simultaneous adverse effects on a large number of institutions or markets as a consequence of severe and widespread ("systematic") shocks. “ “Based on this terminology a systemic crisis (in the narrow and broad sense) can be defined as a systemic event that affects a considerable number of financial institutions or markets in a strong sense, thereby severely impairing the general well-functioning (of an important part) of the financial system.” “Systemic risk (in the narrow and broad sense) can then be de fined as the risk of experiencing systemic events in the strong sense.”		X	X			X	X		X				(de Bandt & Hartmann 2000)
“We can reach a working definition of systemic risk: the risk that (i) an economic shock such as market or institutional failure triggers (through a panic or otherwise) either (X) the failure of a chain of markets or institutions or (Y) a chain of significant losses to financial institutions, (ii) resulting in increases in the cost of capital or decreases in its availability, often evidenced by substantial financial-market price volatility”		X	X										(Schwarcz 2008)
“There are two main strands of model development, which resonate with different policy objectives and corresponding risk indicators of systemic risk: (i) a particular activity causes a firm to fail, whose importance to the system imposes marginal distress on other firms (or markets) (“contribution approach”),14 or (ii) a firm experiences losses from a single (or multiple) large shock(s) due a significant exposure to the commonly affected sector, country and/or currency (“concentration of activity”) and either amplifies systemic risk due to own distress (“participation-contribution approach”) or demonstrates sufficient resilience to absorb common shocks (“participation approach”).”		X				X							(Jobst 2012)
"1. General: Probability of loss or failure common to all members of a class or group or to an entire system. Erroneously also called systematic risk. 2. Investing and trading: Probability of loss common to all businesses and investment opportunities, and inherent in all dealings in a market. Also called market risk, it cannot be circumvented or eliminated by portfolio diversification but may be reduced by hedging. In stock markets, systemic risk is measured by beta-coefficient."		X				X							Business Dictionary http://www.businessdictionary.com/definition/systemic-risk.html
“Although there is no one way to define systemic risk, all definitions attempt to capture risks to the stability of the financial system as a whole, as opposed to the risk facing individual financial institutions or market participants.”		X											FSCO annual report http://www.scribd.com/doc/61877765/2011-FSOC-Annual-Report
“Systemic risk “is often viewed as a phenomenon that is there “when we see it,” reflecting a sense of a broad-based breakdown in the functioning of the financial system, which is normally realized, ex-post, by a large number of failures of financial institutions (usually banks).”		X						X					International Monetary Fund 2009)

“Like Justice Potter Stewart's description of pornography, systemic risk seems to be hard to define but we think we know it when we see it. “ “A more formal definition is any set of circumstances that threatens the stability of or public confidence in the financial system.”		X											(Billio et al. 2012)
"In finance, systemic risk is the risk of collapse of an entire financial system or entire market, as opposed to risk associated with any one individual entity, group or component of a system."		X											Wikipedia https://en.wikipedia.org/wiki/Systemic_risk
“Establishing what constitutes systemic importance has proved difficult, and most G+20 members do not have a formal definition.” “in practice G+20 members consider an institution, market or instrument as systemic if its failure or malfunction causes widespread distress, either as a direct impact or as a trigger for broader contagion.”			X					X					(FSB, IMF, BIS 2009)
“While systemic risk has been seen as a threat caused by unpredictable, highly improbable, exogenous stochastic events (Albeverio et al., 2006; Taleb, 2007), we see systemic risk as reflecting endogenous structural weakness” “While these networks involve the transmission of materials, capital, information and knowledge, recent decades of intense global integration mean that these highly interconnected networks also have the potential to originate and propagate risk. This central property of interconnectedness in networks (Jervis, 1997) can be paradoxical in both its structure and impacts. Increasingly connected networks facilitated by globalization can lead to both greater robustness and more fragility”			X						X			Structural weakness	(Goldin & Vogel 2010)
“An adage among traders is that, in times of crisis, everything is correlated. Though conference participants did not share a consensus on the definition of systemic risk, the descriptions of systemic events by risk managers at the conference reflected this view”.				X									(Board on Mathematical Sciences and Their Applications, Division on Engineering and Physical Sciences, National Research Council 2007)
Use the FSB and IMF definition, endorsed G20													Geneva_Association_Systemic_Risk_in_Insurance_Report_March2010
"A systemic risk is the risk of a phase transition from one equilibrium to another, much less optimal equilibrium, characterized by multiple self-reinforcing feedback mechanisms making it difficult to reverse."												Phase transition Self-reinforcing feedbacks	(Hendricks 2009)
“Systemic risk occurs when individual actors unknowingly create risk through their systemic interactions with each other (which they are often unable to model).					X								http://ineteconomics.org/sites/inet.civicactions.net/files/INETSession2-Lunch-Farmer_0.pdf
“In this study we will mainly address the insurance issues related to the emergence of systemic risks, as defined in the OECD study on emerging systemic risks. In this study, these are defined as risks that occur as a result of current or future functioning of *major systems*. These major systems lead to complex interactions, which therefore lead to increasing risks. “										X			(Faure & Hartlief 2003)
“The potential for large-scale disasters or catastrophes characterized by both extreme uncertainty and a potential for extensive and perhaps irreversible harm.”								X			X		(Cavelty)
"A systemic risk, in the terminology of this report, is one that affects the systems on which society depends – health, transport, environment, telecommunications, etc."										X			(OECD 2003)

Defining systemic risk

"This term denotes the fact that risk to human health and the environment is embedded in a larger context of social, financial and economic risks and opportunities. " "A holistic and systemic concept of risk must expand the scope of risk assessment beyond its two classic components: extent of damage and probability of occurrence."												Risk inside various systems	(Renn & Klinke 2004)
“Risks that can trigger unexpected large-scale changes of a system or imply uncontrollable large-scale threats to it”							X	X			X		(Helbing 2009)
“Rigorously speaking, systemic risk refers to systemic failure, i.e. to the failure common to an entire system, is it the financial system or the market or the whole economic system including the government.”					X								(Trainar 2010)

The autopilot problem

Insurance Industry Perspective

Christian Bieri

Market Unit Head,
Amlin Re Europe

White Paper Context

Automation and models have been introduced in many industries, with human performance supplemented or replaced by computers and algorithms. In many cases, performance does not improve as much as was expected, and in a few cases it actually worsens. One reason for this is that often the humans involved do not perform at the same level of competence as they did before automation was introduced. The consequences of this can be drastic, involving plane and car crashes, medical errors, and even financial crashes.

This phenomenon can be analysed from a general, cross-industry perspective. The same features reoccur in a variety of different domains, with the same human and computer errors repeated from insurance to aviation. Since one of the most evocative and well-studied examples of this issue involves errors from airline pilots' over-reliance on autopilot systems, we call this "the autopilot problem".

In this paper the four root causes of the problem are explained and potential ways to reduce or mitigate their effects are suggested.

Industry Relevance

In (re)insurance, modelling is becoming more and more important, with catastrophe models immediately springing to mind as an important part of (re)insurer risk assessment. But it does not stop there, with capital models, for example, being used as the basis for important strategic decisions.

People in our industry often believe that the more we can model, the better our assessment of a risk must be. And despite the claim to be "not focused only on model output", more often than not, decisions are taken based on model results.

Applying the autopilot problem to (re)insurance raises a number of questions such as;

- Can underwriters still get a handle on the catastrophe exposure in a larger portfolio, without simply following the model results?
- Increased usage of models can lead to a degradation of skills amongst the user. Are we aware of where and in which functions within the organisation the impact of skills degradation will have the most effect?
- Are decision makers aware of the shortcomings of the models involved in the decision making process?

A potential outcome could be the need to retrain underwriters or change their role, at least in parts of the underwriting process.

Next Steps

This white paper sets the scene for further research, in particular:

- Testing to what extent the autopilot problem actually applies to underwriting (and potentially capital management) by assessing whether the four root causes of the problem are valid in our context;
- Elaborating which mitigation strategies can work in (re)insurance and to what extent the problem has to be accepted.

The autopilot problem

Stuart Armstrong
Research Fellow,
FHI-Amlin Collaboration

Heather Bradshaw
Visiting Fellow at the Uehiro
Centre for Practical Ethics,
University of Oxford

Nick Beckstead
Research Fellow,
FHI-Amlin Collaboration

Anders Sandberg
Senior Research Fellow,
FHI-Amlin Collaboration

Introduction

On the 1st of June 2009, at 02:10:05 UTC, the autopilot of Air France Flight 447 disengaged suddenly while the plane was on the way from Rio to Paris. The pilots, who'd been overseeing the flight in a very general way, were suddenly dropped into the middle of an emergency. This was triggered by a technical failure, but the pilots couldn't know this, as they had not been actively flying the plane to that point. They were trying desperately to simultaneously figure out what was going wrong and correct it without the necessary comparison data. But the pilots never were to know what the problem was, before the plane hit the waters of the Atlantic, killing everyone on board.

At 1:45AM on September 15, 2008, Lehman brothers filed for bankruptcy. Over the next few weeks, traders and investors tried to cope with the ongoing contagion of collapse and bad debt. Their usual tools for measuring risk – such as the financial models derived from the Black-Scholes equations (Black & Scholes, 1973), and the credit rating agencies – turned out to be flawed. The collapse in confidence, credit and liquidity was far greater than the models had ever imagined, while the credit agencies gave AAA ratings to investments that turned out to be worthless[1]. Bankers and investors had to come up with some other way of assessing the true risk of investments and corporations, least credit dry up entirely and the crisis become self-sustaining. They failed, and the world was set for many years of recession.

Both of these are illustrations of a similar underlying problem, the "autopilot problem"[2]. This emerges when any position which requires human skills (pilot, or financial risk assessor) is replaced with an automated system, with the human role transformed into that of an overseer. For pilots, the automated system was the actual autopilot; for financial risk assessors and traders, it was the models of financial mathematics. The actual problem is that the human overseer will not be as capable as the original human 'pilot' was, either as an overseer or as replacement for the automation in cases where it no longer functions. Systems built without this realisation will lead to errors and underperform expectations.

This need not result in huge disasters – though, for example, loss of life due to drivers blindly following GPS to their deaths has been reported[3] – but the problem is surprisingly common[4]. Variants of the autopilot problem are apparent in piloting, air-traffic control, vehicle driving, navigation (on land and on sea), cancer detection and other medical diagnostics, industrial processes, military strike decisions, political predictions, financial trading, insurance and even building security[5].

[1] Famously, this was the case with the "synthetic" Collateralized Debt Obligation (CDO), with continual (and successful) commercial pressure being brought on the rating agencies to rate them safer than their true risk http://www.bloomberg.com/apps/news?pid=newsarchive&sid=ax3vfya_Vtdo .

[2] Similar problems exist in many fields where automation replaces or extends human capabilities, and are often called different names. The term "automation bias" is a common one, though it is often used to refer to a specific subproblem of the autopilot problem (Parasuraman & Manzey, 2010).

[3] See for instance http://www.npr.org/2011/07/26/137646147/the-gps-a-fatally-misleading-travel-companion .

[4] Indeed it is starting to enter the press's consciousness, e.g. http://www.economist.com/blogs/babbage/2013/08/cockpit-automation .

[5] See Bruce Shneier's example https://www.schneier.com/blog/archives/2013/09/humanmachine_tr.html .

The autopilot problem

On a more personal level, people are experiencing versions of the autopilot problem as their smart phones take over organising their information and their daily tasks, and future increases in automation or models – such as prediction markets[6] – will cause new versions of the problem in the years to come.

This paper will analyse the autopilot problem by first decomposing it into four separate contributory factors: loss of situational awareness, skills degradation, human error causing misplaced trust and complacency, and unreliable modal estimates from the automation. Each of these factors will be analysed with reference to the literature and specific examples.

It will then draw on the extensive literature existing across the different fields that have experienced the autopilot problem, and the successful and unsuccessful attempts that have been made to solve it. The problems can be mitigated to some extent by retraining the human overseer or reprogramming the automation to better interface with them. But there seems to be limits to what can be achieved in this way. More radical solutions involve dramatically changing the role of the overseer, thus transforming the joint system into something radically different (and hopefully better) than the pre-automation system. Finally, if the problem is too egregious and intractable, there may be no alternative but accepting the problem without being able to solve it – or reducing automation and returning to the older way of doing things.

Defining the autopilot problem

Why does the 'autopilot overseer' not perform as well as the 'pilot' did before? To both define the problem and generalise it, we need to isolate the significant features of the autopilot problem, stretching from pilots in airplanes to modellers in finance[7]. At its core, the autopilot problem has four main features that contribute to degraded performance:

1. Loss of situational awareness.
2. Skills degradation.
3. Human error: misplaced trust and complacency.
4. Unreliable modal estimates.

The first three problems are well known in the research (Cummings, 2004) and have been apparent, in one form or another, since the early days of automation (Bainbridge, 1983), while the fourth is under-analysed to date.

Loss of situational awareness

Situational awareness is defined as "the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" (Endsley, 1995). It is essentially the continual perception of the values of important variables, changes in their values, and the meaning of these changes. These variables might be height and angle of the plane, or the volatility or risk appetite of the market. Automation tends to reduce situational awareness, by distancing the 'pilot' from these variables and making them less-decision relevant.

The crash of Air France 447[8] perfectly illustrates the loss of situational awareness. As they had not been manually flying the plane, the pilots had neither immediate historical knowledge of the key flight parameters such as airspeed, nor did they have the human intuitive sense of the changes in the machine's performance derived from physical control and contact. Thus they were unable to notice the discrepancy between the instrument information and actual performance. Not only did they not realise their instruments had failed, but they also did not have enough intuitive data to recognise the most basic flying emergency – a stall.

[6] These are speculative markets created for the purpose of making predictions. The current market prices can then be interpreted as predictions of the probability of the event or the expected value of the parameter (Arrow, et al., 2008). If decision makers come to rely on them instead of making their own decisions, the autopilot problem can result.

[7] From now on we will use the term 'pilot', in quotation marks, to refer to anyone potentially suffering from the autopilot problem, not just to actual airline pilots. The term 'autopilot' will be used similarly.

[8] The flight was from Rio de Janeiro to Paris, crashing on the 1st of June 2009. See http://en.wikipedia.org/wiki/Air_France_Flight_447 .

On a lesser scale, adaptive cruise control causes loss of situational awareness in driving, delaying driver reactions in critical situations such as narrow curves and fog banks (Vollratha, Schleicherb, & Gelau, 2011). Drivers using ACC were consistently 5 seconds slower to react.

A similar situation developed in the financial markets with collateralized debt obligations (CDO) based on sub-prime mortgages, or on other CDOs. Models and rating agencies (using models) gave their estimation of the risks involved. The headline numbers allowed traders to value and sell these instruments, while distancing them from the real underlying assets. The loss of situational awareness was so complete that when the subprime crisis hit, banks and investment organisations were unaware of the extent of their exposure to these complex financial instrument: they didn't understand what risks they carried on their books, as the actual risk was buried under many levels of financial transformations. This contributed to a severe liquidity crisis[9].

Industry relevant points:

- *There are different kinds of situational awareness, and the autopilot problem does not apply to all.*
- *Insurers probably have great situational awareness about market conditions, about the reliability of brokers, or about the strengths, weaknesses and strategies of other insurers.*
- *Insurers try to develop situational awareness about the physical conditions of the things they insure, for instance by improving reporting, getting more details, plotting their locations, and so on.*
- *Insurers probably lack good situational awareness about the underlying physical processes behind disasters, and the relevant conditions on the ground.*
- *Insurers probably only have moderate situational awareness of the model-maker's biases and errors, as they probably have only partial knowledge of the details of the modelling companies internal decisions.*
- *Good situational awareness would allow insurers to detect when changing conditions could make the Cat models unreliable (e.g. changes in solar radiation or shifts in oceanic currents).*

Skill degradation

In order to maintain expertise, experts must be put into situations where they receive frequent feedback about their actions (Kahneman & Klein, 2009) (Shanteau, 1992). Lacking this feedback, their expertise and skills degrade, and they are no longer capable of performing as well. The older generation of 'pilots' can sometimes "ride on their skills" (perform at a reduced, but still acceptable, level by using the skills and techniques acquired before the introduction of the 'autopilot' (Bainbridge, 1983)), while subsequent generations cannot develop these skills in the first place.

This is the case in aviation, where the typical pilot spends more time managing the flight or the autopilot, but much less time actually flying the plane than they used to[10]. Adaptive cruise control has a similar effect: drivers are spending a smaller proportion of their trip actually in direct control of their vehicle.

Similarly, traders used to have to try to gain an understanding of the securities they invested in, to estimate the risk properly, and establish a price that reflected their understanding and intuitions without relying on sophisticated financial models – the models didn't exist, nor did the computers with power to run them. But the rise of mathematical models changed the profession entirely: mathematicians and physicists[11] flooded in, equipped with model-manipulation skills, not with deeper understanding of the markets[12]. For instance traders spent a lot of their time controlling the 'Greeks' on their option portfolios (Haug, 2007). These Greeks were numbers, *derived from the Black-Scholes model*, that purported to measure the sensitivity of the value of the portfolio to small changes in underlying parameters. A few investors (such as Warren Buffet, famously) maintained a deeper analysis of company performance[13], but in general these skills were less sought-after and were less used, leading to a degradation of expertise.

[9] Starting on the 9th of August 2007, when many banks stopped lending to each other, as they could not assess each other's exposure and risk http://www.theguardian.com/business/economics-blog/2012/aug/05/economic-crisis-myths-sustain .

[10] See the description of a typical flight in http://flyingforeveryone.blogspot.co.uk/2012/01/autopilot-myth-what-your-pilot-really.html .

[11] Called 'quants' (quantitative analysts) in trading.

[12] See for instance this quote from Emanuel Derman in 2003: "There are always implicit assumptions behind a model and its solution method. But human beings have limited foresight and great imagination, so that, inevitably, a model will be used in ways its creator never intended. This is especially true in trading environments, where not enough time can be spent on making interfaces fail-safe…" (Derman, 2003)

[13] Investing in this way is known as "value trading".

The autopilot problem

Industry relevant points:

- *Maintaining a longer institutional memory can help against some types of skill degradation.*
- *But this institutional memory can only go so far: there is a limit to what can be imparted by teaching or by experience, if the skills are no longer currently used.*
- *Those most closely connected with the data may be able to preserve their skills if given the right tasks.*
- *Those at the top of the organisations have a different set of skills (networking, management, planning) that would be less likely to be affected by degradation.*
- *Thus an insurer's skill degradation is likely most acute at the mid-level, where people deal mostly with models and information that has been summarised or collated.*
- *These mid-level people may develop other skills, or may be trained to develop other skills – see the solution section.*

Human error: misplaced trust and complacency.

Human reasoning has well established flaws and weaknesses, and the worst designed 'autopilots' exacerbate rather than mitigate these effects. Humans are often cognitive misers, using as little information as they can and placing heavy reliance on cognitive shortcuts (Fiske & Taylor, 1991). Kahneman distinguished between the slow "system two" reasoning and an automatic and quick "system one", operating with little effort and outside of voluntary control (Kahneman, 2011): human brains have a preference for using system one whenever possible.

By taking over part of the decision process, the 'autopilot' facilitates the overuse of system one. Computerised warning systems can replace long and thorough (and tiring) processes of manual risk-checking. If the 'autopilot' is highly accurate, and makes few mistakes (or even makes no mistakes, in the career of the human overseeing it), this will inevitably breed complacency and bias in the 'pilot' (Parasuraman & Manzey, 2010). This can be understood mainly as attention allocation errors: if the humans had all the time they needed, and thoroughly followed all the procedures they were supposed to follow, including checking for errors carefully, then they would be less prone to falling victim to complacency or excess trust in automation.

But that is not a realistic scenario in humans, and cognitive pressure or task load make it worse (Biros, Daly, & Gunsch, 2004). Humans fall victim to automation complacency in many variants of the autopilot problem, including in the medical profession, where erroneous advice was more likely to be followed when given by a an automated clinical decision support system (CDSS) than by a non-CDSS control (Goddard, Roudsari, & Wyatt, 2012). The automated advice was trusted more, without good reason for this trust, while non-automated advice was more likely to be questioned.

To a large extent, misplaced trust flows from the 'autopilot's' seeming success: the 'pilot' has a lot of experience of the 'autopilot's' successes, and little of its errors. Traders who used the Black-Scholes models (Black & Scholes, 1973) of finance had long experience of successful pricing, punctuated only by rare 'crises'[14], which were easy to dismiss.

This example could do with some further analysis, because the errors made are very illuminating. To vastly oversimplify, the Black-Scholes model rests on two foundational assumptions: that arbitrage opportunities are impossible[15] and that the volatility of stocks are log-normally distributed. The first assumption is often justified, the second far less so – indeed Benoit Mandelbrot criticised it as far back as 1963, advocating using instead a "Levy-Stable distribution with an α of 1.7"[16] (Mandelbrot, 1963). This seems to have been ignored, possibly because the difference would have only been visible in large rare events[17]. Traders tend to ignore such tail risks, possibly because they experienced them so rarely, or possibly because of a "narrative fallacy" which causes traders to seek and believe narrative explanations for large swings (Taleb, 2007). With such an explanation in hand, the large swings could plausibly be discounted as exceptions, and the standard model preserved.

But there were more egregious errors than this. The standard Black-Scholes model predicted that the "implied volatility" of an option would be constant as a function of its strike price; experience showed that this was false, and that it followed instead a "volatility smile" (Hull, 2010). At that point, traders should have accepted that the underpinning of their model was wrong, and that they could only use it in empirical fashion, checking it against the actual behaviour of the market. In particular, the model could not be used to predict rare extreme events, as it was not calibrated on these.

[14] Such as the 1987 "Black Monday" collapse where 22% of the Dow Jones Industrial Average's value was lost, and the 1998 Russian financial crisis, which sunk the Long-Term Capital Management (on the board of which was Myron Scholes, co-author of Black–Scholes model).

[15] Because traders will instantly take advantage of any pricing situations that allow for risk-free income, thus removing the opportunity of others to further profit from that arbitrage opportunity.

[16] Rather than an ⍰ of 2 for the normal distribution.

[17] This is often referred to as "tail risk", since these events happen out in the tails of the probability distribution of likely events.

Instead, what seems to have happened is that the volatility smiles were added arbitrarily to the Black-Scholes model, and may have paradoxically increased confidence in the combined model[18]. There is a common human tendency (the "affect heuristic" (Finucane, Alhakami, Slovic, & Johnson, 2000)) to judge the quality of something based on general feelings of goodness or badness rather than a decomposition of advantages and disadvantages. It is plausible that having used their model to price options successfully day after day (and having plausible narrative reasons for any failure), traders came to believe that it was of high quality even outside of this domain of validity, and could even account for large-scale tail risks (a position for which there was no evidence).

Industry relevant points:

- *These kinds of errors are most likely to be problems under time pressure or stress.*
- *Business culture and attitude can help against alleviate these problems to some extent, as can good business procedures and workloads.*
- *But some problems will remain, even in the best of cases: some of these flaws are intrinsic to the human condition, and cannot be removed (see the section on retraining the pilot for examples of biases that cannot be overcome through training).*
- *The problem is especially acute when a successful model is used beyond its domain of validity (a historical example being the use of early Californian quake 'models' in areas such as Australia, without reassessing for the local hazard). In these cases, the model users may have a very erroneous impression of the model's validity.*

Unreliable **modal** estimates

Often the most likely estimate the model gives – the "modal" estimate – is known to be inaccurate, but is still used. For instance the Black-Scholes model of finance allowed traders to put a price on very exotic and unusual derivatives – financial instruments that could not be priced before[19]. The model is, of course, flawed, and the price is certainly incorrect, to some degree at least. But nevertheless it was a price for that financial instrument where none was previously known, allowing it to be compared with other instruments and traded in large quantities.

This is a common feature of the autopilot problem in insurance and finance. The 'autopilot' comes up with an estimate, which is known to be flawed, but where the true value is either impossible or very difficult to evaluate correctly. And in many circumstances, this flawed estimate starts being taken as accurate and used as a basis for decisions, without the necessary scepticism.

There are several reasons for this. People may feel more psychologically comfortable trusting a precise numerical estimate, even if the precision is spurious (Egger, Schneider, & Smith, 1998). There is a tendency to anchor on an initial value, by taking it as a starting point and not moving far enough from it when further evidence comes in (Tversky & Kahneman, 1974). The complacency mentioned in the previous section can also play a role, of course.

But there are also systemic reasons for sticking close to the 'autopilot's' estimate. It can provide 'cover' for those seeking to justify their decisions to superiors[20]. Following the modal estimate and getting things very wrong is unlikely to be punished as severely as deviating from the standard wisdom and getting things very wrong. In the second case, the 'pilot' must be willing to accept a greater accountability for their actions than in the first, something not all humans are comfortable with, even if it leads to improved performance (Skitka, Mosier, & Burdick, 2000).

The 'pilot' may not even be aware of the flaws. This happens often in models, where different measures of risk or uncertainty get combined together. Often, only the headline numbers are reported, with the 'pilots' not cognisant of the assumptions going into them. They are therefore left with no choice except to take the numbers as given.

In finance, if the model is in common use throughout the industry, then it makes sense to stick to the price given[21]. Even if the price is wrong, the traders know that they will be able to buy (or sell) their financial instrument at that

[18] Adapting a model to new empirical data is a praiseworthy task in general. The problem here is that the confidence levels were not adjusted to take on board the fact the initial model had failed: the uncertainty was underestimated as a result. Also, the model may "degrade" over time if extra "patches" are added without understanding the fundamental assumptions underlying the model, especially if the patch *seems* to extend the model's domain of validity. Model feedback is essential but needs to be carefully handled.

[19] For instance, consider the so-called 'barrier' options, which are standard options that are brought into existence, or extinguished, when the underlying asset breaches a pre-set *barrier level*. These cannot be priced by classical Black–Scholes, more complex methods building on this can be used. Other examples are interest rate and currency swaps, which used to be considered exotic derivatives, but became to be considered standard financial instruments once models allowed them to be priced – models that were known to be incoherent (Boenkost & Schmidt, 2005) but these models were still used (at least until the 2007 financial crisis)!

[20] See Tetlock's 'Social Contingency Model' (Tetlock, 1992) for a more detailed analysis of the factors surrounding accountability is social situations.

[21] Also regulators and quasi-regulators (such as Lloyds of London) can make it costlier to deviate from the mainstream, as any deviation has to be explained and possibly justified, sometimes at length.

price, so the model's estimate takes on a life of its own as the market price.

There is no advantage then to deviating from the model price, unless the trader can detect in which direction the model is biased, and thus figure out if it is under-pricing or over-pricing[22]. In other words, the flawed model's estimate will be accepted as accurate, unless there is a better source of information around[23]. And though some traders can profit from flawed models by taking the long view and betting on extreme events[24], most traders were incentivised for short term returns.

Industry relevant points:

- *This is a general problem, but particularly relevant to insurance and trading models.*
- *Reporting the uncertainty doesn't solve the problem: the 'modal estimate' would then be the mean plus uncertainty (or even the whole distribution). It can help to report 'autopilot' uncertainties on specific problems (rather than global uncertainties of the 'autopilot'), but as long as there are reasons to suspect the distribution is wrong (but no easy way of correcting it), part of the problem will remain.*
- *Anchoring happens without people being aware of it, and is very hard to combat – all obvious solutions fail. Smart, well informed, and well incentivised people still anchor, even when aware of the anchoring effect (see later section on retraining the pilot for citations).*
- *Some of the problems are systemic, and can't be resolved by one insurer in isolation, as insurers will have to follow the market to some extent in some areas in order to remain solvent. Though it may be possible for some to perform better than before, even in the presence of unresolved systemic problems.*
- *For Cat models, if the underlying dynamics shift – e.g. due to a major earthquake that shifts the stresses in the surrounding fault zones – then it is likely the shift will be detected before a new model comes out. So there will be a period when the current model is known to be unreliable, but the directions of the errors are unknown.*
- *Using models beyond their domain of validity (e.g. using standard models for estimating liquefaction risks in New Zealand) is a very common source of unreliable modal estimates.*

[22] Another drawback to being "outside the market" is that the market has auto-corrective features that a single model – even a better one – may lack. There is a difference between "if everyone used model X instead of model Y, it would be better for everyone" and "I should myself change from model X to model Y".
[23] It is a well-known truism that "all models are wrong, but some are useful" (Box & Draper, 1987), but without ability to act on that maxim, mere knowledge of it is insufficient.
[24] Nassim Taleb and Mark Spitznagel made a large profit off the 2008 financial crisis http://online.wsj.com/news/articles/SB122567265138591705 .

Solving the autopilot problem

The autopilot problem can be very severe, causing plane and economic crashes. Thus, methods to reduce or solve the problem would be of great value. Given the analysis of the previous section, how can the problem best be addressed?

There have been several attempts to reduce the autopilot problem in a variety of fields, which can be decomposed into five categories:

1. Retrain the pilot.
2. Change the autopilot.
3. Change the pilot's role.
4. Accept the cost of the autopilot problem.
5. Reduce or remove the autopilot.

This section will analyse the five approaches in detail, illustrating them with practical examples from literature. The actual airplane autopilot problem, and the excessive dependence on mathematical models (in the financial or insurance worlds, for instance) will be taken as two canonical examples of the autopilot problem, and improvements relevant to these two will be presented. Note that there may be unusual systemic effects: though it is clear that individual actors could benefit from these improvements, their impact may be different on the industry as a whole. Considerations of systemic issues will be addressed in other papers, as it is beyond the scope of the autopilot problem.

Retrain the pilot

Retraining the 'pilot' is the first and most obvious solution. Since the 'pilot' performed at a higher level without the 'autopilot', it seems reasonable to assume that the previous performance[25] can be recaptured in some way, with the correct training, retaining the old skills, re-developing situational awareness, and so on.

But some biases cannot be overcome by training[26] (similarly, some are completely independent of cognitive ability (Stanovich & West, 2008)). Training cannot overcome the anchoring bias, for instance (a key part of the "Unreliable modal estimates" problem). Explaining the anchoring effect and warning against its consequences was insufficient to correct it (Wilson, Houston, Etling, & Brekke, 1996). Even when financial rewards were offered for correct responses, the anchoring effect persisted (Simmons, LeBoeuf, & Nelson, 2010). Parasuraman and Manzey found that many putative interventions failed to reduce the problems of complacency and excess trust in the 'autopilot': more training in the task at hand, more experience (and many alternative ideas, such adding more than one 'pilot') failed to improve the situation (Parasuraman & Manzey, 2010). Thus, many obvious interventions do not achieve the expected results.

Some improvements are possible, however. One option is to increase the accountability of the 'pilot', making them – and not the 'autopilot' – responsible for the outcome. In this situation, errors of omission (failing to respond to system irregularities which the 'autopilot' did not flag) and of commission (following the advice of the 'autopilot' despite other information indicating it was in error) were both reduced (Skitka, Mosier, & Burdick, 2000). Internalised senses of accountability were more effective at reducing the bias than externally manipulated accountability demands (Mosier, Skitka, Heers, & Burdick, 1998). These results hold as long as the 'pilot' was accountable for the outcome: if the 'pilot' was accountable for the time taken, then performance was worsened, not improved. This suggests that efforts aimed at reducing time pressure and cognitive load could also help improve performance.

The above approach is best used to combat issues of complacency and, to some extent, misplaced trust. It does not help with skill degradation, loss of situational awareness or unreliable modal estimates. To combat skill degradation, the 'pilot' needs to be put in a situation where they use their skills and get correct feedback. Real pilots currently spend little time flying planes: this could be combatted by making them train in simulators, or by requiring them to fly the plane by hand in situations when this can be done with acceptable risk levels (such as when planes have to fly without passengers on board). Traders or insurance underwriters could be required to make a proportion of their decisions based on their skills alone, without using models[27]. This would also be useful for comparing and contrasting the skills of humans versus that of models, and for designing further interventions and improved approaches.

It should also be possible to make airplane pilots perform tasks that maintain their situational awareness of the plane. The various features of the plane's situation (height,

[25] This assumes the 'pilot's' new role is similar to their old one, or where they need to be able to reprise their old role in some situations. See later sections for cases where the role is quite different.

[26] Though some are more susceptible to training, such as confirmation bias: "In some but not all studies, basic education about specific cognitive biases (e.g., brief and nontechnical tutorials on confirmation bias) also decreases participants' tendency to fall prey to certain errors, including confirmation bias." (Lilienfeld, Ammirati, & Landfield, 2009)

[27] There could be "back-to-basics" days where the traders are denied use of models for a day and performing analysis from first principles, checking their assumptions and the limitations of these as they go.

heading, angle of attack, etc.) could be gamified[28], for instance, rewarding the pilot in a competitive way that maintains their vigilance and awareness. Or pilots could be required to perform tasks can only be completed successfully if they keep a successful track of the plane's properties.

The same approach could be used in finance, but it is less clear what the critical variables are, and how relevant they are to issues of tail risk, where a lot of the financial risk lies. The ideal would be to make sure that human skills remain valuable in day-to-day decisions, but this may not be achievable. Modern trading algorithms trade millions of times a second; there is no way for humans to develop a reasonable situational awareness on these scales. Large errors are likely to be "Black Swans" (such as the 2007-2012 financial crisis). Since these events are rare and hard to predict, humans would be unlikely to have relevant awareness or skills, and it is very unclear how they could be trained to acquire them (Shanteau, 1992).

Informing the 'pilot' of the weaknesses and errors of the 'autopilot' gave mixed results[29]. Typically, 'pilots' overestimate the reliability of the 'autopilot' until they had experience of it making errors; after that, they mistrusted even reliable 'autopilots'. Upon being informed of the features of the 'autopilot' and why it could err, they started trusting the 'autopilot' again. But this trust increase happened in both situations where it was justified (high reliability) and when it was not (low reliability) (Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003). So simply being aware of the strengths and weaknesses of the 'autopilot' seems insufficient.

Industry relevant points:

- *It should be possible to give underwriters better understanding of the weaknesses of their models. Some of the training given to underwriters could be akin to that given to model-makers.*
- *Such training could make use of feedback and exercises and hypothetical situations: the training should be tailored to take advantage of human abilities and limitations, rather than fighting against them.*
- *If the retraining prevents skill degradation, it will help to maintain performance during periods where the Cat models are suspected to be or have become unreliable, as human expertise will be able to compensate for this to some extent.*
- *However, a lot of retraining ideas may seem like good ideas, but won't achieve anything: many problems are subconscious "system one" failings rather than failed rationality. It is not enough for the retraining idea to sound good, it must be backed up by evidence (and its implementation and effects assessed).*
- *There are limits to what can be done with retraining only. Though it seems the most obvious and easy answer, other types of solutions may prove to be easier and more effective.*

Change the autopilot

Rather than re-training the 'pilot', the 'autopilot' itself can be changed to reduce the problem. Decreasing the level of automation can help: there is evidence that the best 'autopilots' are those that support processes of information integration and analysis on the part of the 'pilots'. Those that instead provide specific recommendation to action worsen the problem (Parasuraman & Manzey, 2010) (Crocoll, 1990).

Varying the reliability of the 'autopilot' can also improve performance (Goddard, Roudsari, & Wyatt, 2012): if the 'pilot' has reasons to suspect that the 'autopilot' is not perfect, they are less likely to trust it blindly. There seems to be a valley in which the autopilot problem comes into sharp relief: for low reliability, the 'pilots' will perform as before, and for high reliability, the 'autopilot' will work better that the 'pilot' did beforehand. It's in between, with a flawed 'autopilot' and underperforming 'pilot', that the danger lies[30]. Varying the reliability of automation is an attempt to move out of this valley, as far as the 'pilot' is concerned.

One useful intervention is to have the 'autopilot' display its confidence levels[31], and have these be updated (McGuirl & Sarter, 2006) (Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003). Many models already do this to a large extent, including standard deviations and other uncertainty information in their outputs. The main weakness of this approach is that it requires the model to have accurate calibration of its own uncertainty, including model uncertainty – the probability that the model itself is wrong, which cannot be calculated from within the model. The design of the interface has some effect on the autopilot problem as well. But all these are mainly efforts to reduce complacency and excess trust in the 'autopilot'. They may reduce skill degradation, if they make the 'pilot'

[28] Using game-like elements to promote user engagement and learning (Zichermann & Cunningham, 2011).

[29] "One study found that making users aware of the DSS reasoning process increased appropriate reliance, thus reducing Automation Bias." (Goddard, Roudsari, & Wyatt, 2012)

[30] And much of our important automation systems fall into this area and will do for the foreseeable future.

[31] For Watson, the IBM computer system that ended up winning the 'Jeopardy' quiz show, displaying and taking account of its own uncertainty levels was absolutely crucial http://www.aaai.org/Magazine/Watson/watson.php .

more likely to use their own expertise to overrule the 'autopilot'. But they do not address the loss of situational awareness or the unreliable modal estimate problem.

A very intriguing example of successful retraining-through-model-change happened after Kahneman redesigned the Israeli army interview system. After establishing that the previous subjective interview system was worthless for the task it was attempting to do, he designed a set of rigid criteria which interviewers were to follow (an autopilot, in other words). The interviewers insisted that they also be able to give their subjective opinions – opinions which were surprisingly accurate. In fact the new combined trained-subjective plus 'autopilot' method was more effective than either the 'autopilot' or the old purely subjective method. Thanks to the 'autopilot', the 'pilots' gained increased skill at their jobs. Replicating this achievement seems challenging, but it is an intriguing example, hinting at the possibility of further improvements – the right 'autopilot' can increase the 'pilot's' skills, rather than degrading them.

Industry relevant points:

- *New platforms like Oasis will give insurers opportunities to tweak their models and select more suitable ones (though model-shopping may become a problem).*
- *This selection process must consider the human-model interface, and not just the abstract virtues of the model.*
- *Over the short term, this is most easily done by redesigning interface without changing the models at all.*
- *Over the long term, models that increase the understanding of the model-user should be better than models that don't, even if the second type are abstractly "better".*
- *Just as the retraining in the previous section, this sort of intervention is likely to only result in limited improvements. Both are more patches to the autopilot problem than fundamental solutions to it.*

Change the pilot's role

The previous two strategies were attempts to undo part of the autopilot problem, to return the situation to the status quo ante for the 'pilot', while preserving the use of the 'autopilot'. This may not be feasible in many problems: the previous approaches are good at correcting human error, but much weaker at addressing the other causes of the problem.

Another alternative is to abandon the idea of re-creating the previous status quo, and instead change the role of the 'pilot' to better suit the new situation. There are many examples of professions transforming themselves completely when automation entered their field. The word processor abolished the job of the old secretary, causing redundancies for some, and moving the remaining into the new role of personal assistant. The web and smartphone caused mini-autopilot problems as human factual memory, being less useful, ended up degrading (Sparrow, Liu, & Wegner, 2011). Instead, people compensated by becoming adept Googlers, developing new types of expertise from new types of feedback – just as calculators in their day moved people away from mental arithmetic to more conceptual tasks.

Similar restructuring of roles are common whenever automation becomes viable. Airline pilots have started along this path, becoming managers of the plane[32] and this will no doubt continue as the autopilots improve, and take on more roles such as taking off and landing. The pilots of unmanned aerial vehicles have already made this transition, fitting into a completely new role and learning new skills, as well as encountering new types of problems[33]. This role restructuring may not be enough to fully solve the autopilot problem, but may alleviate it to a great extent, and would generally be a better use of human skills.

After the flash-crash[34] automatic shut-down mechanisms were introduced to stop things so that they could be reviewed by humans. Thus the automation process was changed so that humans could step in, in a new role, that of reviewer rather than trader.

This has also happened to some extent in the insurance industry. Underwriters and other insurance staff are starting to not just make use of their models, but analyse them and benchmark them[35] as well[36]. The Oasis project[37]

[32] "In modern airliners, the pilot's main responsibilities are to monitor the automatic systems to make sure the plane is flying correctly and to alter the course as needed." http://science.howstuffworks.com/transport/flight/modern/airline-crew1.htm . "The climb continues with your pilots "hand flying" the plane....all the way up to 29,000 feet where, federal regulations dictate they must turn on the autoflight system." http://flyingforeveryone.blogspot.co.uk/2012/01/autopilot-myth-what-your-pilot-really.html .

[33] Such as the stress of watching hours of close-up videos of the people killed in drone strikes: "After a strike, operators assess the damage, and unlike fighter pilots who fly thousands of feet above their targets, drone operators can see in vivid detail what they have destroyed." http://www.nytimes.com/2011/12/19/world/asia/air-force-drone-operators-show-high-levels-of-stress.html?_r=0 .

[34] A crash which occurred on the 6th of May 2010, in which the Dow Jones Industrial average lost 600 points in 5 minutes, before regaining most of the value in the next few minutes. High-frequency traders contributed strongly to this event (U.S. Securities and Exchange Commission and the Commodity Futures Trading Commission, 2010).

[35] At the most basic level, this can involve using a more simple approach (such as a certain % of Total Sum Insured) as a sanity check on the complex autopilot.

is increasing this effect, by making the model components modular and changeable. This makes the model-user implicitly responsible for these choices: whereas before there were only a few standard model choices, now there are many, and choosing among them is an exercise that must ultimately be justified.

This can increase accountability for the model user, which has already been shown to decrease human error and complacency. Ideally this approach could help to maintain both skill and situational awareness, as the 'pilot' is continually trying to understand the model/'autopilot' and compare it with reality (though at the risk of introducing human biases into the system (Tversky & Kahneman, 1974)). This has the added benefit of making the model into less of black box, so that more people (the 'pilots', but also their managers) can actually understand where decisions are coming from.

It may also be possible to reduce the impact of unreliable modal estimates. If the 'pilot' is aware of this problem, and is given an explicit mandate, they may be able to address the problem directly. They could, for instance, increase their uncertainty, take extra precautions. Traders could invest in extreme events, and insurance underwriters could pay more attention to tail risk. They could seek to find tools beyond the 'autopilot' to measure or bound the variables there is uncertainty about[38], or be tasked with benchmarking the various 'autopilots'. All in all, giving the 'pilot' the explicit role of compensating for the weaknesses of the 'autopilot' could be a sound strategy[39].

Industry relevant points:

- *This is probably the most promising route over the long term.*
- *It opens the possibility of completely solving the autopilot problem, by moving the human component into new areas where their new skills will be effective.*
- *It's hard to know what needs to be done right at the moment, though: more research and experimentation is needed.*
- *There will need to be a greater role for those tasked specifically with examining and tracking weaknesses and assumptions of different models and their associated decision-makers.*
- *This is the only approach that can allow estimates of unmodelled risks (such as some secondary perils from wind, flood or earthquake). Having people tasked with comparing models with each other and with reality, thus getting a general estimate of model error, can give an estimate of the magnitude of such missing risk.*
- *It could allow humans to focus their attention on estimating the impact of various changes on the models – for instance man-made interventions such as reclaiming land and flood diversion. These changes call some of the model assumptions into question, and it is important to figure whether the model still retains validity.*
- *The role of underwriters may shift to make them more into model-analysts than model users.*

Accept the cost of the autopilot problem

It may be worth keeping the autopilot, even if the autopilot problem can't be fixed. The 'autopilot' has caused a deskilling of 'pilots', certainly. But the 'autopilot' has many uses, as well. It provides a regularity and smoothness at the controls that no human could match over long periods of time. It isn't subject to fatigue or emotional stress and rarely behaves erratically, either in the cockpit, in the market or in clinical situations. It should be noted that flying is safer than ever[40]!

In financial markets and insurance, sophisticated mathematical models have allowed the trading of previously untradeable securities, or the insurance of previously un-insurable risks. GPS devices may cause the driver to become more of an automaton, but they are more likely to arrive at their destination at a predictable time.

Thus there are situations where the benefits of automation outweigh the drawbacks (given the particular weaknesses of the particular other half of the system – the human), even after taking the autopilot problem into account. After all efforts are made to mitigate that problem, sometimes the remaining issues just have to be accepted. There seems no sign that automation will be reversed in driving and piloting (the trends seem to strongly point in the other direction[41]), so in the assessment of many decision makers, the autopilot problem is not bad enough to go back to the old ways. There are many cases in which

[36] A change that may be needed: the Future of Humanity Institute's paper on the future of employment ranks underwriters as very susceptible to be replaced by automation (Frey & Osborne, 2013)!

[37] A new, open and modular catastrophe modelling approach, see http://www.oasislmf.org/ .

[38] For instance, insurance companies often use market-share models as a sanity check for their more complex catastrophe models.

[39] Just as the 'autopilot' is often designed to compensate for the weaknesses of 'pilot'.

[40] The International Air Transport Association (IATA) announced that the 2012 global accident rate for Western-built jets was the lowest in aviation history http://www.iata.org/pressroom/pr/pages/2013-02-28-01.aspx .

[41] The Google self-driving car being an extreme example of this trend, but even normal cars increasingly contain driver's aids.

statistical prediction rules make better predictions than leading experts (Bishop & Trout, 2005). So it is hardly surprising that there would be many cases where the combination of ‘autopilot’ and less skilled ‘pilot’ outperforms a more skilled but unassisted ‘pilot’.

Of course, when this is done, everyone should be aware of the cost that is being incurred. Modellers will have to accept the limitations of their model, and find ways of dealing with them[42] (for instance, by increasing uncertainty to account for uncertainty about the model, rather than simply within the model, or by using). Pilots will have to face up to their loss of skill and no longer believe they can fly their plane as once they could.

Industry relevant points:

- *The autopilot problem will likely never be solved completely, so some acceptance of the costs is inevitable.*
- *Accepting the costs cannot be done without a good assessment of these costs, so that needs to be a priority. It is important to find other methods for bounding the values of various uncertain variables. The approach in “Probing the Improbable” can be used to synthesise this knowledge.*
- *Properly assessing the costs will go some way towards allowing them to be mitigated – but not all the way. The two approaches can often proceed together, however.*
- *It is likely that the best approach is to accept that some loss cannot be modelled (e.g. supply chain risk, possibly), and that it should be bounded sensibly, while the rest of the loss can be treated using other methods.*
- *Properly accounting for tail risks is the main challenge to accepting the costs.*

Reduce or remove the autopilot

Finally, the autopilot problem may be so severe, that the only solution is to remove the ‘autopilot’ and go back to the old way of doing things[43]. The Warren Buffet style of investing is an example of this: cutting down on the use of derivatives and models and making greater use of value investing[44]. Mercedes did the same with its “Sensotronic Brake Control System”, which included a software model of how the brake pedal should “feel”. After customer protests and software errors, the ‘autopilot’ was removed, and the company went back to a more conventional hydraulic braking system[45]. In many situations – such as air-traffic control – there are automation “tipping points” where the ‘pilot’ accepts a certain level of automation without problem, but rejects any higher level of automation (Bekier, Molesworth, & Williamson, 2012).

In Tetlock’s analysis of political predictions (Tetlock, 2005), he demonstrated that “foxes” (who have a flexible, adaptive, tentative cognitive style) outperform “hedgehogs” (who are said to "know one thing and know it well" and to focus on a single, coherent theoretical framework in their analyses and predictions). This framework forms a model – an ‘autopilot – that the hedgehogs will then ‘pilot’. In contrast, the foxes are less welded to any single model and have developed the skills that allow them to reach better conclusions using a variety of different methods.

Removing one ‘autopilot’ need not mean eschewing automation entirely. For instance, models based on fundamentals have had some success at predicting election outcomes, but poll-based models have fared better[46]. It could be that the best role for the ‘pilot’ is simply to choose between ‘autopilots’.

In contrast, when statistical prediction rules make better predictions than leading experts (Bishop & Trout, 2005), adding human expertise to the mix often adds only noise: the performance of the mixed system is degraded. Thus there seem to be multiple situations where the ‘autopilot’ and the ‘pilot’ cannot mix: one of them must be removed to ensure the improvement of the system.

Industry relevant points:

- *Doing without models does not seem a feasible strategy for some insurers.*
- *Reducing the use of the model may be successful if humans can be expected to develop relevant skills. Hence model use can be reduced for intermediate risks, where human expertise may have developed, but not for tail risks.*
- *Choosing different models, or some synthesis of models, or using human judgement to choose between*

[42] For instance, by increasing uncertainty to account for uncertainty about the model, rather than simply within the model. Or they could use alternative techniques to independently bound the values of certain parameters.

[43] Though one must beware the availability heuristic (Tversky & Kahneman, 1973). The recent flaws of the autopilot will loom large, while the problems of pre-autopilot performance will not be so available. Considering that autopilots are often developed in areas of poor human performance, it may still be better to stick even with a poorly performing autopilot.

[44] See various articles on Buffet’s investment style, such as http://dealbook.nytimes.com/2011/03/14/derivatives-as-accused-by-buffett/?_r=0 and http://www.investopedia.com/articles/05/012705.asp .

[45] See for instance the report in ‘Autoweek’ at http://www.autoweek.com/apps/pbcs.dll/article?AID=/20051216/FREE/51216010&SearchID=73232069810043 .

[46] See Nate Silver’s analysis at http://fivethirtyeight.blogs.nytimes.com/2012/03/26/models-based-on-fundamentals-have-failed-at-predicting-presidential-elections/ .

models, may be a better way to successfully reduce the importance of a single model.

- *Small scale, mid-term experiments without use of models could be attempted in the company to test their efficacy. Again, this is not likely to be informative for tail risks.*

Summary and conclusion

The autopilot problem is an important and general problem across many fields that make use of automation to assist or replace human decision making. Due to loss of situational awareness, skills degradation, misplaced trust and complacency, and unreliable modal estimates, the performance of the human component of the system will be degraded from what it was before the introduction of automation.

This problem is not easy to solve. Misplaced trust and complacency stand out as the easiest to cope with, with a large class of potential solutions, but the other factors are harder to address. Since the scope of the autopilot problem is so broad, however, many different fields have experimented with many different approaches to dealing with it – and many of these solutions are at least partially transferable to other fields. These solutions come under five broad categories. The most obvious are retraining the human overseer or reprogramming the 'autopilot' to ensure a better interaction between the two, one that alleviates the problem. It is also possible to radically change the human's role to one more suited for the new situation. Doing without automation – returning the situation to the status quo ante – or severely reducing it, is another approach. Finally, if all mitigation fails, there is always the option of accepting the presence of the autopilot problem and managing the system as best can be done with that knowledge.

These solutions seem highly situationally dependent: there are no overarching solutions that apply to every variant of the autopilot problem. Instead, individual interventions must be crafted with care, with an eye on the relevant literature and an eye on the specific details of the individual setup being analysed. Continuous feedback and monitoring of the intervention is essential, to assess the degree of improvement, and what form they take. It is hoped that the analysis in the present paper will contribute to many such successful interventions, and that further research hones and improves these suggestions far beyond what was presented here. Automation has contributed and will contribute much more to most fields of human endeavour, so attempts to eliminate or reduce the autopilot problem will ensure that it reaches its full positive potential.

Bibliography

U.S. Securities and Exchange Commission and the Commodity Futures Trading Commission. (2010). *Findings Regarding the Market Events of May 6, 2010.*

Arrow, K. J., Forsythe, R., Gorham, M., Hahn, R., Hanson, R., Ledyard, J. O., et al. (2008). The promise of prediction markets. *Science, 320*(5878), 877-878.

Bainbridge, L. (1983). Ironies of automation. *Automatica, 19*(6), 775-779.

Bekier, M., Molesworth, B. R., & Williamson, A. (2012). Tipping point: The narrow path between automation acceptance and rejection in air traffic management. *Safety Science, 50*(2), 259-265.

Biros, D. P., Daly, M., & Gunsch, G. (2004). The Influence of Task Load and Automation Trust on Deception Detection. *Group Decision and Negotiation, 13*(2), 173-189.

Bishop, M. A., & Trout, J. D. (2005). *Epistemology and the psychology of human judgment.* Oxford: Oxford University Press.

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *The journal of political economy*, 637-654.

Boenkost, W., & Schmidt, W. (2005). Cross currency swap valuation. *SSRN 1375540*.

Box, G. E., & Draper, N. R. (1987). *Empirical model-building and response surfaces.* John Wiley & Sons.

Crocoll, W. M. (1990). Status or recommendation: Selecting the type of information for decision aiding. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 34*(19), 1524-1528.

Cummings, M. L. (2004). Automation bias in intelligent time critical decision support systems. *In AIAA 1st Intelligent Systems Technical Conference, 2*, 557-562.

Derman, E. (2003). Model Risk. In *The Handbook of Risk* (pp. 129-142). IMCA: John Wiley & Sons.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies, 58*(6), 697-718.

Egger, M., Schneider, M., & Smith, G. D. (1998). Spurious precision? Meta-analysis of observational studies. *Bmj, 316*(7125), 140-144.

Endsley, M. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*(1), 32-64.

Finucane, M., Alhakami, A., Slovic, P., & Johnson, S. (2000). The Affect Heuristic in Judgment of Risks and Benefits. *Journal of Behavioral Decision Making, 13*(1), 1-7.

Fiske, S., & Taylor, S. (1991). *Social cognition* (2nd ed.). New York: McGraw-Hill.

Frey, C. B., & Osborne, M. A. (2013). *The Future of Employment: How Susceptible are Jobs to Computerisation?* Oxford: Oxford Martin School.

Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: a systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association, 19*(1), 121-127.

Haug, E. G. (2007). *The Complete Guide to Option Pricing Formulas.* McGraw-Hill Professional.

Hull, J. C. (2010). *Options, Futures and Other Derivatives* (7 ed.). Pearson Education India.

Kahneman, D. (2011). *Thinking, fast and slow.* Macmillan.

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: a failure to disagree. *American Psychologist, 64*(6), 515-526.

Lilienfeld, S. O., Ammirati, R., & Landfield, K. (2009). Giving debiasing away: Can psychological research on correcting cognitive errors promote human welfare? *Perspectives on Psychological Science, 4*(4), 390-398.

The autopilot problem

Mandelbrot, B. (1963). The Variation of Certain Speculative Prices. *The Journal of Business, 36*(4), 394-419.

McGuirl, J. M., & Sarter, N. B. (2006). Supporting trust calibration and the effective use of decision aids by presenting dynamic system confidence information. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 48*(4), 656.

Mosier, K. L., Skitka, L. J., Heers, S., & Burdick, M. (1998). Automation bias: Decision making and performance in high-tech cockpits. *The International journal of aviation psychology, 8*(1), 47-63.

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factor: The Journal of the Human Factors and Ergonomics Society, 52*(3), 381-410.

Shanteau, J. (1992). Competence in experts: The role of task characteristics. *Organizational behavior and human decision processes, 53*(2), 252-266.

Simmons, J. P., LeBoeuf, R. A., & Nelson, L. D. (2010). The effect of accuracy motivation on anchoring and adjustment: Do people adjust from provided anchors? *Journal of personality and social psychology, 99*(6), 917.

Skitka, L. J., Mosier, K., & Burdick, M. D. (2000). Accountability and automation bias. *International Journal of Human-Computer Studies, 52*(4), 701-717.

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science, 333*(6043), 776-778.

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology, 94*(4), 672-695.

Taleb, N. N. (2007). *The Black Swan: The Impact of the Highly Improbable.* New York: Random House and Penguin.

Tetlock, P. E. (1992). The impact of accountability on judgment and choice: Toward a social contingency model. *Advances in experimental social psychology, 25*(3), 331-376.

Tetlock, P. E. (2005). *Expert political judgment: How good is it? How can we know?* Princeton University Press.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*(1), 207-233.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124-1131.

Vollratha, M., Schleicherb, S., & Gelau, C. (2011). The influence of Cruise Control and Adaptive Cruise Control on driving behaviour – A driving simulator study. *Accident Analysis & Prevention, 43*(3), 1134-1139.

Wilson, T. D., Houston, C. E., Etling, K. M., & Brekke, N. (1996). *Journal of Experimental Psychology: General, 125*(4), 387.

Zichermann, G., & Cunningham, C. (2011). *Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps* (1 ed.). Sebastopol, California: O'Reilly Media.

Biased error search as a risk of modelling in insurance

Insurance Industry Perspective

JB Crozet

Head of Group Underwriting Modelling, Amlin plc

White Paper Context

(Re)insurers use sophisticated but imperfect models and data sets to estimate risk.

Decisions about when to search for errors, what types of errors to look for, and when to stop looking for errors are biased in favour of vindicating pre-existing views about risk.

This pattern of biased error search is partially driven by:

- Confirmation bias: a tendency, usually subconscious, to seek, interpret, and recall evidence in a way that confirms one's current opinions; and
- Automation bias: a tendency to become overly reliant on automated decision aids, at the expense of vigilant information seeking and processing.

Under time constraints, biased error search leads to finding more expected errors and fewer unexpected errors.

The literature on confirmation and automation biases point to a number of mitigation methods. Testing some practices could shed light on the trade-offs involved in biased error search and increase detection of unexpected errors.

Industry Relevance

If we focus on our industry, we can see that (re)insurers are exposed to the risk of biased error search in two different dimensions: the validation of their models, and the review of model outputs.

The complexity of validating models means that the task often relies on:

- The movements from the previous version (e.g. Internal Model Governance under Solvency 2); and
- RAG-status ("Red, Amber, Green") against pre-set criteria.

While this management by exception has certain value in terms of efficiency, it has the drawback of taking the attention away from potential deficiencies in the figures produced or the validity of the pre-set criteria for RAG-status. For instance, the recipients of a model update report are less likely to think "model deficiency or uncertainty" if the report shows little movement from period to period and has green status.

A typical example when reviewing model outputs would be an underwriter reviewing the return-period losses for their portfolio, in order to decide how much to rely on the figures for their decision.

The use of expert judgement for complex, low probability estimation is notoriously difficult, and the underwriter is more likely to question results which differ from their experience of actual events or previous model outputs, but not otherwise.

This is especially a problem in rapidly changing risk environments (e.g. large growth in exposed values on the Florida Coast in recent decades), where a model providing stable outputs might mean that it is not reflecting the risk adequately.

Next Steps

This white paper sets the scene for further research, in particular:

- Evidencing any bias in (re)insurance modelling practices, and assessing their potential impact on the quality of modelling results;
- Assessing whether there is a systemic risk associated with this bias, and the consequences for individual (re)insurers and the industry;
- Testing the effectiveness of mitigation methods.

Biased error search as a risk of modelling in insurance

Nick Beckstead
Research Fellow,
FHI-Amlin Collaboration

Stuart Armstrong
Research Fellow,
FHI-Amlin Collaboration

Anders Sandberg
Senior Research Fellow,
FHI-Amlin Collaboration

Executive Summary

Insurance companies use sophisticated but imperfect models and data sets to estimate risk. Decisions about when to search these models and data sets for errors, what types of errors to look for, and when to stop looking for errors are biased in favor of vindicating pre-existing views about risk. This pattern of biased error search is partially driven by confirmation bias and automation bias. Under time constraints, biased error search leads to finding more expected errors and fewer unexpected errors, but more errors in total. Quantitative information about these trade-offs is unknown, but the trade-offs could be substantial.

There is not yet enough evidence to strongly recommend effective interventions. However, based on a series of interviews with employees in the insurance industry, a review of interventions aimed at reducing confirmation bias and automation bias, and a model of the consequences of biased error search, we recommend testing the following practices on a small scale, comparing the results with a control group:

1. Keep a record of all manual adjustments from initial settings on catastrophe models and exposure data. Require modellers to write a one-sentence reason whenever they make a manual adjustment to a catastrophe model or the exposure data. Inform modellers that there will be randomly spot checks of adjustments and non-adjustments.
2. Require underwriters to write at least one sentence about why they might have overestimated the loss from accepting a contract, and once sentence about why they might have underestimated the loss. Pass this on to risk review, and inform the underwriters in advance that this will be done.
3. Keep a record of model-estimated losses, both before and after making all manual adjustments, as well as the estimated losses from underwriters, and then periodically compare these with actual losses. Let modellers and underwriters know that this will be done.
4. Randomly select some model-based estimates for detailed inspection in cases where nothing seems amiss in order to find base rates of error in such cases.
5. More closely review cases where unadjusted models, adjusted models, and underwriter loss estimates differ substantially.
6. Educate modellers, underwriters, and actuaries on how to recognize confirmation bias and avoid it.
7. Design modelling software that displays an educated estimate of the model's reliability for the case in which it is being used.

A test of these practices would shed light on the trade-offs involved in biased error search and increase detection of unexpected errors.

Biased error search as a risk of modelling in insurance

Section 1: Introduction

The insurance industry makes significant use of catastrophe models and exposure data to help estimate losses from insurance contracts. These model-based estimates form a useful baseline for estimating losses, but they can be substantially mistaken for a variety of reasons. Modellers and underwriters have some experience with past losses from similar contracts, and use their experience and intuition to decide when and how to spend effort examining model-based estimates of losses. If they lean too far in favor of searching for errors primarily in cases where model-based estimates seem wrong, there is a risk of tuning models and data sets too much in the direction of pre-existing opinions. On the other hand, focusing the search for errors in places where one expects to find them can lead to finding more errors per unit time.

The objectives of this paper are to assess how modellers and underwriters make decisions about when to search for errors in these estimates, what types of errors to look for, and when to stop looking for errors, what the costs and benefits of this decision process is, and how this decision process might be improved. The points here are quite general, however, and would apply to many types of error search.

This paper draws on unstructured interviews during a week-long immersion period at a major insurance company. We spoke with modellers, underwriters, actuaries, risk reviewers, and other employees. The interviews focused on many topics, but one theme was the process for identifying possible errors in the company's use of catastrophe models and data sets and adjusting the models and data sets in light of errors or possible improvements to make. These interviews suggest that the insurance industry searches for errors in model-based loss estimates in ways that tend to vindicate pre-existing views of risk.

In cognitive science, *confirmation bias* is a tendency, usually subconscious, to seek, interpret, and recall information in a way that confirms one's current opinions (Nickerson 1998, p. 175).[1] Our interviews strongly suggested that confirmation bias played a major role in error checking and model adjustment.

In ergonomics and human factors, *automation bias* is a tendency to place too much trust in the results of automated processes for detecting risk, sometimes leading to a lower detection rate than unaided humans can achieve without automation. There is a more precise definition in section 5. Automation bias has been studied primarily in the aviation industry, though related problems have arisen with automated decision aids in clinical decision-making as well. There are some potentially significant analogies with the use of models in insurance, and many suggestions for dealing with automation bias could potentially be useful for improving model use in insurance.

The outline of the rest of the paper is as follows. Section 2 presents a simple framework for understanding biased error search, drawing on an intuitive example, an informal model, and an example from the history of science. It also delves into the cognitive science literature on confirmation bias and the ergonomics literature on automation bias. Section 3 reports on results from a series of interviews with modellers, underwriters, actuaries, and other employees in the insurance industry. These interviews suggest that the framework outlined in section 2 is realistic. Section 4 uses the framework developed to consider the positive and negative consequences of biased error search for the insurance industry, and identifies some key questions for developing our understanding of the consequences of biased error search.

Section 5 offers recommendations for experimentally measuring the positive and negative consequences of biased error search and mitigating its negative effects. The recommendations for mitigating negative effects are based on reviews of experimental studies which attempted to reduce the role of confirmation bias and automation bias.

[1] "As the term is used in this article and, I believe, generally by psychologists, confirmation bias connotes a less explicit, less consciously one-sided case-building process. It refers usually to unwitting selectivity in the acquisition and use of evidence….The assumption that people can and do engage in case-building unwittingly, without intending to treat evidence in a biased way or even being aware of doing so, is fundamental to the concept." (Nickerson 1998, pp. 175-176)

Section 2: Background on biased error search

Example: checking a dinner bill for errors

Suppose five friends go out to eat at a restaurant and they decide to split the bill. No one has added up the total, but each of them roughly expects to pay £30. Consider three possibilities:

1. The bill comes back with a total that's in line with what everyone roughly expected: £175, or £35 per person.
2. The bill comes back with a total that is surprisingly high: £300, or £60 per person.
3. The bill comes back with a total that is surprisingly low: £75, or £15 per person.

Whether they check the bill for errors, and what kind of errors they look for, depends on how the total on the bill fits with their expectations, and perhaps also on whether correcting the bill is in their interests.

- In the first case, they might glance at the bill, but they are likely to pay without looking too closely. If their expectations are roughly correct, this works fine. If their expectations about the cost of the meal are far too high or far too low, this could mean overpaying or underpaying.
- In the second case, they are likely to look at the bill and check to see if there was a mistake. Perhaps they got the bill for another table, perhaps the waiter charged them for really expensive wine they didn't order, or perhaps they accidentally ordered really expensive wine.
- In the third case, a very scrupulous person might be about equally likely to check the bill for errors as they would be in the second case. If they do, they probably look for items the waiter forgot to charge them for, or check to see if they gave them a bill for the table of two sitting next to them.

In catastrophe modelling, the concept of an error is not completely straightforward. Rather than thinking of catastrophe models as right or wrong, some may find it more appropriate to consider the models as more or less useful, or better or worse approximations of reality, relative to a particular purpose for which the model is being used. In this paper, we will be ecumenical about our use of the word "error", and include as errors both straightforward mistakes (such straightforwardly incorrect exposure data) as well as more subtle deviations from what best fits the purpose for which the model is being used on a particular occasion.

A simple framework for understanding biased error search

The example above is a special kind of confirmation bias which we'll call *biased error search*. Biased error search generally involves some prior expectations, an imperfect estimation process, and a decision to check for errors that is biased in favor of looking for errors when the estimate doesn't fit with prior expectations.

Suppose we are interested in the value of some quantity X. For illustrative purposes, we'll continue to consider the case where X is how much the group owes the restaurant, but X could be the expected loss of writing an insurance contract or the 99.5% value at risk on a portfolio for a certain class of business. They have some prior expectations or assumptions about the value of X. These expectations or assumptions might be specific (such as "exactly £150"), general and vague (such as "between £100 and £200"), or general and precise (such as "my subjective probability over X is lognormally distributed and I'm 95% confident the cost is between £100 and £200"). The true value of X might, be, roughly speaking, about what they expect, be surprisingly low, or be surprisingly high, as in Figure 1.

When they hear about an imperfect estimate of X, as they do when they get their dinner bill, they make a decision about how much weight to put on that estimate and whether to check for errors in the estimate. This tends to go one of three ways:

1. If the estimate of X generally fits with their prior expectations about X, (e.g., if the total on the bill is about what they expected) this will typically reinforce their prior assumptions about X, and is unlikely to result in careful scrutiny of the estimate, as in Figure 2. This also tends to reinforce any assumptions they were using to reach this estimate.

2. If the estimate of X is significantly higher than their prior expectations about X, then they spend more of their effort looking for ways in which the estimate might be too high (e.g., checking to see if they got charged for items they didn't order), as in Figure 3. If they fail to find errors, this will tend to undermine their belief in any assumptions they were using to reach this estimate.

3. If the estimate of X is significantly lower than their prior expectations about X, then they spend more of their effort looking for ways in which the estimate might be too low (e.g., checking to see if they didn't get charged for some items they did order), as in Figure 4. If they fail to find errors, this will tend to undermine their belief in any assumptions they were using to reach this estimate.

Figure 1: Expectations

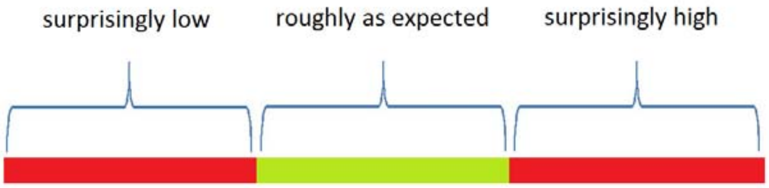


Figure 2: Estimate fits expectations, limited/no search for errors

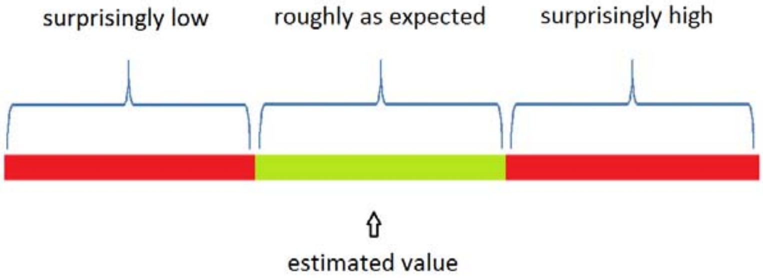


Figure 3: Surprisingly high estimate, search for errors that would result in an overestimate

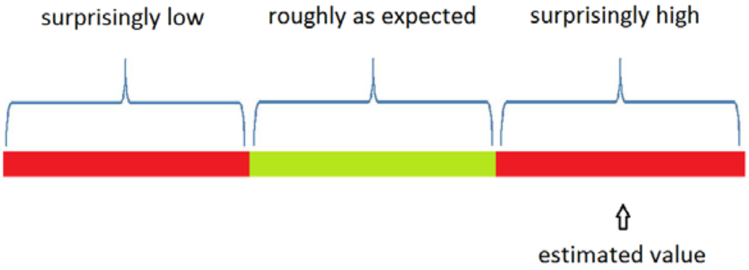
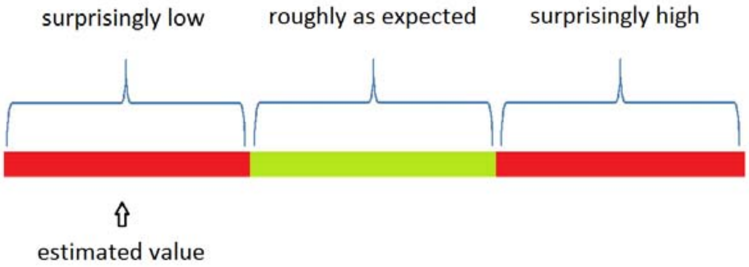


Figure 4: Surprisingly low estimate, search for errors that would result in an underestimate



Biased error search as a risk of modelling in insurance

An example of biased error search from the history of science

Biased error search is also a familiar problem in science, as illustrated by this example from Richard Feynman:

> "Millikan measured the charge on an electron by an experiment with falling oil drops, and got an answer which we now know not to be quite right. It's a little bit off, because he had the incorrect value for the viscosity of air. It's interesting to look at the history of measurements of the charge of the electron, after Millikan. If you plot them as a function of time, you find that one is a little bigger than Millikan's, and the next one's a little bit bigger than that, and the next one's a little bit bigger than that, until finally they settle down to a number which is higher.
>
> Why didn't they discover that the new number was higher right away? It's a thing that scientists are ashamed of—this history—because it's apparent that people did things like this: When they got a number that was too high above Millikan's, they thought something must be wrong—and they would look for and find a reason why something might be wrong. When they got a number closer to Millikan's value they didn't look so hard. And so they eliminated the numbers that were too far off, and did other things like that." (Feynman 1974, p. 12)

As Feynman suggests in all but name, this is an example of biased error search because the scientists looked harder for errors—and tended to find them—when they got answers that were further away from their prior expectations, where their prior expectations were heavily informed by previous work. In this case, a group using biased error search took longer to converge on the correct estimate of the charge of an electron than a group using an unbiased search for errors.

Biased error search as a variety of confirmation bias

Confirmation bias is a tendency, usually subconscious, to seek, interpret, and recall evidence in a way that confirms one's current opinions. In contrast, a perfectly rational approach to processing evidence with no time constraints would consider all the evidence for and against all hypotheses under consideration, without disproportionately seeking evidence in favor of what one currently believes, disproportionately remembering evidence in favor of what one believes, or disproportionately processing evidence in ways that tend to confirm what one already believes.

In the context of correcting insurance models and data sets for errors, the tendency to seek information in ways that justifies one's current opinions seems to be the most relevant. Citing (Koriat, Lichtenstein, & Fischhoff, 1980) in a review of the literature on confirmation bias, Nickerson (1998, p. 177) found that:

> "People tend to seek information that they consider supportive of favored hypotheses or existing beliefs and to interpret information in ways that are partial to those hypotheses or beliefs. Conversely, they tend not to seek and perhaps even to avoid information that would be considered counterindicative with respect to those hypotheses or beliefs and supportive of alternative possibilities,"

In our dinner bill example, this would mean that if they expected the dinner bill to be between £100 and £200 and they got a bill saying £300, they'd be more likely to look for evidence that they had been overcharged (e.g., because they charged them for food they didn't buy or gave them a bill from a table with more people) and less likely to look for evidence that they had been undercharged (e.g., because they forgot to charge them for food that they did buy or they gave them a bill from a table with fewer people). In an insurance context, it might mean that if a model gives an expected loss estimate for some exposures in Miami that seems too high, they'll be more likely to look for evidence that the estimate is too high (e.g., because the damageability was set too high or because historical losses are lower than the estimate assumes) and less likely to look for evidence that the estimate is too low (e.g., because the damageability was set too low or because historical losses are higher than what the estimate assumes).

Motivated cognition vs. honest biased searches

Confirmation bias can be influenced by *motivated cognition*, i.e. a selective search for evidence that is biased in favor of getting answers that they want to hear. For example, if the estimate of X is more favorable to their interests (e.g., if the total on the bill seems too low and they don't care much about being honest), then they'll be less likely to check for errors than they would be in a case where the estimate of X was less favorable to their interests. In an insurance context, a broker preparing a submission and hoping for a low price on a contract will be less likely to search for errors in a low estimate of loss than a high estimate of loss. Error searches can also be sensitive to market dynamics. For instance, models may be more likely to be checked if margins are very thin due high estimated loss costs. Like all forms of confirmation bias, motivated cognition will often be subconsciously driven.

A group's decision about whether to check for errors can also be influenced by an honest search for information that pays attention to how much time they have to check for errors and the probability of finding errors. In more detail, biased error search can save time and help find errors if prior expectations are reasonable. If a group had all the time in the world and they really just wanted an accurate estimate of X, they would check for ways in which their estimate of X might be too high and the ways in which it might be too low in all three cases (surprisingly high estimate, estimate roughly fits expectations, surprisingly low estimate). In the real world, it costs time and effort to make these checks, so people focus on the cases where a check is more likely to result in finding an error. If they have accurate enough prior expectations—so that, e.g., if the estimate of X seems too high then it's more likely to really be too high than it is in a case where it seems to low—then they can save time through biased error searches. Still, there is no free lunch. As long as prior expectations are imperfect, biased error search implies a smaller chance of finding unexpected errors.

Automation bias as a driver of biased error search

In addition to confirmation bias, automation bias can be another driver of biased error search. Automation bias is defined in terms of the concept of an *automated decision aid*, meaning devices that support human decision-making in complex environments. Examples of automated decision aids include aviation systems like the Traffic Conflict and Alert System and the Ground Proximity System. These systems monitor the environment and provide specific recommendations, like "Pull up! Pull up!" to pilots. Automated decision aids in medicine can recommend treatments or drug dosages (Parasuraman and Manzey 2010, pp. 390-391). *Automation bias* is the tendency to become overly reliant on the automated decision aid, at the expense of vigilant information seeking and processing (Mosier and Skitka 1996, p. 205).

Automation bias has mainly been studied in aviation, navigation, and medical decision-making contexts. An everyday example of automation bias is the case of someone missing a turn because their GPS malfunctioned. Such a person might well have made the turn on their own if they hadn't grown accustomed to the automated system. An example of the opposite type would be someone who took a wrong turn because their GPS malfunctioned, where they would have made the right turn if they hadn't been using the GPS (Parasuraman and Manzey 2010, p. 391).

In an insurance context, the automated decision aid is the insurance model. Automation bias would involve modellers and underwriters becoming overly reliant on models, at the expense of carefully considering and processing other relevant information. Our search of the literature found no published work discussing automation bias in an insurance context.

Insofar as the major problem in biased error search is neglecting to search for errors in cases where estimates fit well enough with prior expectations, automation bias is likely to exacerbate the problem. For a more detailed discussion of problems from automation bias and potential solutions, see Armstrong et al. in this volume.

Section 3:
Evidence of biased error search in insurance

In June 2013, we interviewed modellers, underwriters, actuaries, and other employees at a major insurance company about a variety of topics related to the use of models in insurance. We found evidence of (i) the existence of strong prior expectations, (ii) a tendency to carefully look for errors in estimates that differed from prior expectations, and (iii) a much less significant tendency to scrutinize estimates that fit with prior expectations, all consistent with biased error search playing an important role in insurance.

Regarding the existence of prior expectations, underwriters and modellers use models and data sets to help estimate the expected losses associated with writing insurance contracts. Underwriters use models and data sets as a starting point when deciding the price at which they are willing to write a contract, but they also rely on a variety of other factors including historical losses, specific knowledge about the exposures, personal intuition based on experience, relationships with brokers, knowledge of last year's estimates, and general knowledge of the business environment. These factors create prior expectations.

Regarding a tendency to search for errors in cases where an estimate seems wrong, interviewees estimated that when modellers are reviewing a submission from a broker, the estimate from a model will seem wrong around five to ten percent of the time, and this will lead to careful examination. Usually, this happens when the estimated loss differs substantially from the estimate from the previous year, especially in cases where the model has significantly changed. Error searches tend to focus on exposure data first, checking the performance of models on scenarios next, and sometimes talking with model vendors. There is a healthy mix of resolutions in talking with the model vendors.

The most frequent resolution was learning that the model was being applied incorrectly, though bugs in the model are often found at this stage and modellers from the insurance company are often convinced that the surprising estimate of losses was, in fact, reasonable.

However, this tendency to carefully search for errors is selective. We asked several people for examples of cases where (i) a model's estimate fit the modeller's or underwriter's prior expectations, and (ii) the model was inspected for errors. No one could recall such a case, though everyone we asked could supply a case of the type discussed in the above paragraph. This suggests that models and data sets which fit with prior expectations are rarely carefully checked by employees like the modellers and underwriters we spoke with.The tendency to adjust estimates primarily when the produce inconvenient results—and not to modify them when the results are convenient—appears to be common in insurance. In their "Philosophy of Modelling," Edwards and Hoosain (2012, pp. 59-60) appear to agree:

> "There is also the danger of using the opportunity of feedback adjustments to tweak the model in order [to] generate the results that one expects, or, even worse, that one wants (clearly not a danger confined to the feedback stage). As a general point, we seem to be more inclined to modify models where they generate inconvenient results."

Another way to react to a surprisingly high or low estimate—which doesn't neatly fit into the concept of "error search" but is continuous with the phenomenon—is to find a different model with an estimate that fits better with prior expectations. For example, our interviewees told us that in 2011 the model vendor Risk Management Solutions (RMS) released estimates of losses for European windstorms that were substantially higher than what underwriters' prior expectations. Many companies in the insurance industry reacted by using RMS 2011 or models from other vendors (which were closer to their prior expectations) in order to estimate losses from European windstorms. The reactions of the users of model vendors can affect the development of the models themselves; the following year, RMS released a model with loss estimates for European windstorms that were much closer to the estimates in RMS 2010.

Section 4: Consequences of biased error search in insurance

The material covered above illustrates what biased error search is and that it plays a role in insurance. But what are the consequences? This question is substantially less straightforward than it seems at first. This section explains why biased error search helps save time and find errors in cases where prior expectations are generally reliable, but leads to fewer errors found and misguided searches for errors in cases where expectations are less reliable. However, without further research (described in section 5 and 6), it is impossible to tell how the costs and benefits compare.

When expectations are about right, there are two types of cases to consider, each with a positive effect:

1. If the estimated value seems about right and really is (as in Figure 5), biased error search implies limited scrutiny and acceptance of the estimate. That saves time because there isn't an unnecessary search for mistakes.

2. If the estimate value seems too high and really is too high (as in Figure 6), biased error search implies searching for errors that would cause an overestimate. That is an efficient way to bring the estimate back to where it should be. Something analogous happens if the estimate seems too low and really is too low.

When expectations are wrong, there are two more types of cases to consider, each with a negative effect:

3. If the estimate seems about right but the true value is surprisingly high or surprisingly low (as in Figure 7), then biased error search implies no search or a limited search for errors, which is likely to result in no errors found and an incorrect estimate.

4. If the estimate is surprisingly high and the true value really is surprisingly high (as in Figure 8), then biased error search implies a search for errors which, if found, would decrease the estimate. In this case, it's a search for the wrong type of errors. If there are often errors or simply judgment calls that could push the estimate down, the search for errors can be actively misleading. Something analogous happens if the estimate is surprisingly low and the true value really is surprisingly low.

Cases 1-3 are relatively straightforward, but case 4 is not, so it deserves a more thorough explanation. Suppose there is a model estimating the expected loss on the portfolio for a whole class of business. Suppose that the model (and/or exposure data) has been updated recently, and the new model says there is more correlated risk than was previously expected. And suppose that, in fact, the new model is right. The new model's estimate is likely to be seen as surprisingly high, and it will be scrutinized more carefully than it would be in a case where the new model gave an answer that was similar to the old model.

The people scrutinizing the new model (and/or exposure data) will especially be on the lookout for ways in which the risk may have been overestimated. Ideally, such a search couldn't be misleading. However, there a couple reasons that it could be. First, the estimate could have errors pushing in opposite directions that leave the estimate approximately correct. Searching only for errors of one kind could make the estimate worse, as even the best of model-based estimate is likely to have *some* imperfections. Second, the estimate could rely on inputs that rely on judgment calls. These judgment calls may be especially likely to be questioned in cases where they could be adjusted downward.

The net effect considering all four types of cases is hard to discern from first principles, and we know of no relevant tests of how they trade off against each other. Therefore, it is unclear how much biased error search is hurting or helping, and whether surprisingly high and surprisingly low estimates are getting the right amount of scrutiny in comparison with estimates that come out roughly as expected.

Figure 5: Truth and estimate as expected

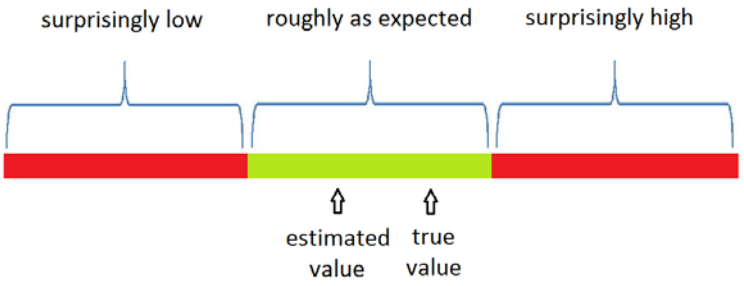


Figure 6: Truth fits expectations, estimates don’t

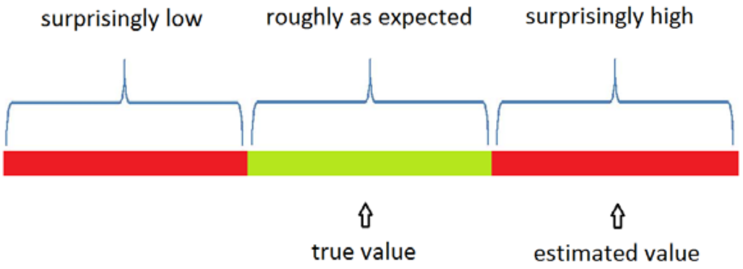


Figure 7: Estimates fit expectations, but truth doesn’t

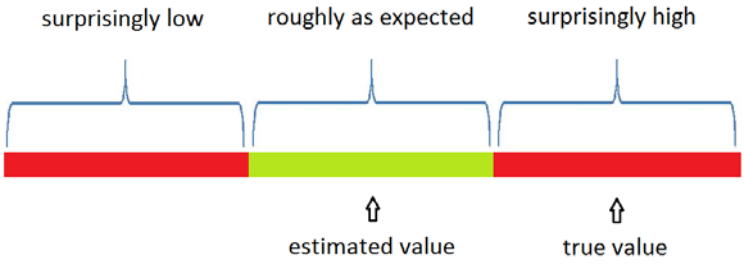
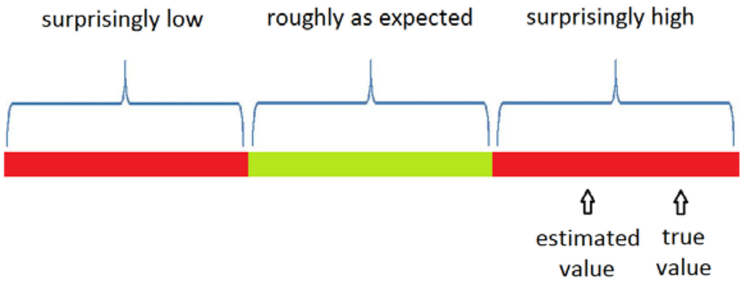


Figure 8: Estimates and truth contrary to expectations



Some important unknowns for determining the consequences of biased error search

To weigh up all these consequences of biased error search, we need answers to the following questions:

1. How do losses estimated by a model (adjusted or unadjusted) compare with estimated losses by underwriters and with real losses?

This question seems important for assessing how much value is added by modellers and underwriters at different stages of the process. We did not discover whether this data is being systematically collected and compared, though our understanding is that underwriters writing a contract must write down estimated losses and that this information is included along with technical price, and reviewed during the risk review process.

2. How often do models and data sets contain important, detectable errors in cases where model-estimated losses fit with prior expectations?

This question can't be answered with existing data because estimates which fit with prior expectations are rarely closely scrutinized. However, some evidence suggests that there may be errors in these cases. For example, when a new person builds a model from the ground up when an old model already exists, the models often disagree.

This question is important for assessing the impact of rarely searching for errors in cases where model-based estimates fit with prior expectations, as our interviews suggest modellers and underwriters currently do. If models and data sets rarely contain important, detectable errors when they fit the experience and prior expectations of underwriters and modellers, there is comparatively little benefit in searching for errors in such cases. However, if they contain important, detectable errors frequently enough in such cases, then rarely checking for errors in such cases would be problematic.

3. Are there generally significant errors in model-based estimates that could be corrected, regardless of whether the estimate looks wrong? How often?

4. Would someone looking for an error generally look for different things if they thought a model's estimated losses were too high than they would if they thought the estimated losses were too low? E.g., would someone looking for errors check damageability, building coding, historical loss records, and so on either way, or would they check some if they thought it was too high and others if they thought it was too low?

5. How do people searching for errors decide when to stop looking for errors?

Better answers to these questions could be found by following the recommendations outlined in the section below.

Tuned models, tuned expectations

These last three questions are important together because they help indicate how much danger there is of tuning models and data sets to prior expectations during the error correction process. Suppose that there were significant correctable errors in most model-based estimates and suppose that someone looking for errors in a model-based estimate generally looks for errors in cases where the model doesn't fit prior expectations, generally looks for errors that would bring the model in line with their prior expectations, and is unlikely to find errors that would go in the other direction unless they were purposely looking for them. Suppose further that people searching for errors had a strong tendency to keep looking for errors as long as model-based estimates didn't fit their prior expectations, and tended to stop searching for errors once model-based estimates did fit their prior expectations.

If all of these things were true, then biased error search would result in an error detection and correction process that substantially tuned the model-based estimates to fit with the prior expectations of modellers and underwriters.

A similar process would unfold with the model users' subjective judgments and expectations. Their experience would be that whenever their expectations differed significantly from the model's, they then were able to find model errors that brought the model back in line and *confirmed that their expectations were correct*. This provides continued feedback seeming to confirm their own expectations, and would cause them to develop high overconfidence in their estimates – more than the evidence warrants. This overconfidence would then be applied to the next time they need to decide whether and where to search for model errors.

Section 5:
How to assess the impact of biased error search and prevent the worst of it

In previous sections, we covered what biased error search is, gave evidence that it happens in the insurance industry, discussed its possible costs and benefits, and identified key questions for getting a clearer picture of their importance and magnitudes. This section covers some methods for reducing the impact of biased error search and further illuminating its consequences.

How could we mitigate biased error search driven by confirmation bias?

The literature on mitigating confirmation bias suggests a number of methods which might reduce the impact of bias error search. For a summary, see the table below:

Among these interventions, asking people who use models to offer reasons that their use of the model might be mistaken, and educating staff about cognitive biases seem stand out because they are simple to implement and test without changing the way people do their jobs.

Table 1: Interventions with experimental evidence suggesting that they mitigate confirmation bias

Intervention	Who claims it works?
Prime people for counterfactual thinking	Kray and Galinsky 2003
Ask people to consider alternate possibilities	Lilienfeld et al. 2009 citing (e.g., Anderson, 1982; Anderson & Sechler, 1986; Hirt & Markman, 1995; Hoch, 1985; Lord, Lepper, & Preston, 1984), (Hoch, 1985; Koriat, Lichtenstein, & Fischhoff, 1980)
Ask people to give reasons justifying their choices	Wheeler and Arunachalam 2008
Ask people to give reasons they might be wrong	Koriat et al. 1980
Present relevant evidence in a graphical layout	Cook and Smallman 2008
Talk to people with different opinions	Schulz-Hardt et al. 2002
Educate people about cognitive biases	Lilienfeld et al. 2009 citing (Evans, Newstead, Allen, & Pollard, 1994; Kurtz & Garfield, 1978; Mynatt, Doherty, & Tweney, 1977; Newstead, Pollard, Evans, & Allen, 1992; Tweney et al., 1980)

Ask people to give reasons they might be wrong

In two experiments, Koriat et al 1980 gave paid volunteers tests of general knowledge. The authors describe the questions as follows:

> The questions covered a wide variety of topics including history, literature, geography, and nature. All had a two-alternative format. For example, "the Sabines were part of (a) ancient India or (b) ancient Rome." (p. 109)

Their abstract describes their methods and results as follows:

> "Exp I presented Ss with 2-alternative questions and required them to list reasons for and against each of the alternatives prior to choosing an answer and assessing the probability of its being correct. This procedure produced a marked improvement in the appropriateness of confidence judgments. Exp II simplified the manipulation by asking Ss first to choose an answer and then to list (a) 1 reason supporting that choice, (b) 1 reason contradicting it, or (c) 1 reason supporting and 1 reason contradicting. Only the listing of contradicting reasons improved the appropriateness of confidence."

Asking for one reason a person might be wrong is a very simple procedure, and this makes it attractive to test in a context where confirmation bias may be problematic.

Educate people about cognitive biases

Lilienfeld et al. 2009 reviewed cognitive science research on how to mitigate confirmation bias. They reported:

> "In some but not all studies, basic education about specific cognitive biases (e.g., brief and nontechnical tutorials on confirmation bias) also decreases participants' tendency to fall prey to certain errors, including confirmation bias (Evans, Newstead, Allen, & Pollard, 1994; Kurtz & Garfield, 1978; Mynatt, Doherty, & Tweney, 1977; Newstead, Pollard, Evans, & Allen, 1992; Tweney et al., 1980)."

However, they emphasized that this kind of solution was unlikely to be a cure-all:

> "Nevertheless, the question of whether instruction alone is sufficient to disabuse people of confirmation bias and related errors is controversial. Arkes (1981) maintained that psychoeducational methods by themselves are ''absolutely worthless'' (p. 326), largely because people are typically oblivious to cognitive influences on their judgments. In contrast, others (e.g., Parmley, 2006) believe that psychoeducational programs may often be efficacious. For example, Willingham (2007) argued that although critical-thinking programs are, at best, modestly effective, the most successful methods teach participants ''metacognitive rules,'' such as reminding them to consider alternative points of view in pertinent situations." P. 393

How could we limit automation bias's role in biased error search?

The two most helpful papers for addressing this question were literature reviews by Parasuraman and Manzey 2010 and Goddard et al. 2012. They found a number of methods for limiting the role of automation bias:

Table 2: Interventions with experimental evidence suggesting they mitigate automation bias

Intervention	Who claims it works?
Decrease level of automation	Parasuraman and Manzey 2010 citing (Crocoll & Coury, 1990; Rovira et al. 2007; Sarter & Schroeder, 2001)
Display the decision aid's context-relative reliability	Parasuraman and Manzey 2010 citing (McGuirl & Sarter, 2006); Goddard et al. (2012) citing (McGuirl & Sarter, 2006)
Increase accountability	Parasuraman and Manzey 2010 citing (Skitka, Mosier, and Burdick 2000)
Make the aid less reliable	Parasuraman and Manzey 2010 citing (R. Parasuraman et al. 1993, May, Molloy, and Parasuraman 1993, and Bagheri and Jamieson 2004)
Decrease task load	Parasuraman and Manzey 2010 citing (Parasuraman et al. 1993); Goddard et al. 2012 citing (Grubb et al. 1995, Dixon and Wickens 2006, and McFadden et al. 2004)
Increase variation in reliability	Parasuraman and Manzey 2010 citing (Parasuraman et al. 1993); Goddard et al. 2012 citing (R. Parasuraman et al. 1993 and R. Parasuraman et al. 1996)
See a recent automation failure	Parasuraman and Manzey 2010 citing (Lee and Moray 1992 and Lee and Moray 1994)
See an automation failure early	Parasuraman and Manzey 2010 citing (Molly and Parasuraman 1996)
Decrease task complexity	Goddard et al. 2012 citing (Bailey and Scerbo 2005)
Decrease time constraints	Goddard et al. 2012 citing (Sarter and Schroeder 2005)
Increase understanding of how the decision aid works	Goddard et al. 2012 citing (Dzindolet et al. 2003)
Give information rather than a recommendation	Goddard et al. 2012 citing (Sarter and Schroeder 2005)

Among these interventions, increasing the accountability of people relying on models and including information about the robustness of insurance models alongside the models themselves stand out, (though the second may be more of a task for the model vendors than insurance companies). The first intervention stands out because it seems good from a common-sense perspective and could be done with only small changes to existing business processes. Implementing the second might be more work, but it also seems strong from a common sense perspective and the effect size was large in experimental studies. Regularly showing employees examples of failures of insurance models and training new staff with examples of insurance model failures also seem fairly promising ideas to test. Experimental results supporting these recommendations are reviewed in the following paragraphs.

Increase accountability

Skitka et al. 2000 got 181 undergraduates to try a flight simulator where they had various objectives and used a decision aid. Some of the undergraduates were told that they would have to explain their choices in the flight simulator to a professional and that their discussions would be recorded for in-depth review later (accountability conditions). Others were told no such thing (control condition). The students in the accountability conditions made fewer errors than the students in the control condition. The authors' conclusion was that increasing accountability decreases automation bias.

It's common sense that holding people accountable for doing something generally makes them do it better, giving a reason to implement this change even without strong research proof.

This common sense perspective is generally upheld, though qualified and made more precise, in a series of research findings by Tetlock (e.g. Tetlock et al. 1989, Lerner and Tetlock 1999). Below, we offer some more detail about how this idea could be implemented to encourage more attentive use of models in insurance.

Display the decision aid's context-relative reliability

McGuirl and Sarter 2006 put 30 instructor pilots through a flight simulator with a decision aid that was used to help the pilots detect icing events (which could cause stalls). The treatment group's decision aid gave the pilots information about how accurate the decision aid was likely to be, whereas the control group's decision aid did not. The treatment group was substantially less likely to stall. They stalled in 36% of cases, whereas the control group stalled in 64% of cases (McGuirl and Sarter 2006, p. 661). McGuirl and Sarter's conclusion was that having information about a decision aid's reliability decreases the effect of automation bias.

As in the previous case, it's common sense that people using a decision aid should be keeping in mind how reliable the decision aid is in the context in which they are using it, giving a reason to implement this change even without strong research proof. Including and prominently displaying this information in catastrophe models may be something for vendors to consider or something for insurance companies who use vendor models to consider adding to internal systems.

Recommendations

We recommend testing the following practices, comparing the results with a control group:

1. Keep a record of all manual adjustments from initial settings on catastrophe models and exposure data. Require modellers to write a one-sentence reason whenever they make a manual adjustment to a catastrophe model or the exposure data. Inform modellers that there will be random spot checks of adjustments and non-adjustments.

2. Require underwriters to write at least one sentence about why they might have overestimated the loss from accepting a contract, and once sentence about why they might have underestimated the loss. Pass this on to risk review, and inform the underwriters in advance that this will be done.

3. Keep a record of model-estimated losses, both before and after making all manual adjustments, as well as the estimated losses from underwriters, and then periodically compare these with actual losses. Let modellers and underwriters know that this will be done.

4. Randomly select some model-based estimates for detailed inspection in cases where nothing seems amiss in order to find base rates of error in such cases.

5. More closely review cases where unadjusted models, adjusted models, and underwriter loss estimates differ substantially.

6. Educate modellers, underwriters, and actuaries on how to recognize confirmation bias and avoid it.

7. Design modelling software that displays an educated estimate of the model's reliability for the case in which it is being used.

These recommendations need not be accepted in an all-or-nothing fashion. They could be tested on a case-by-case basis. Collecting this data (especially #4) would allow one to directly compare error detection rates under biased error search and unbiased error search, making it easier to distinguish between cases where biased error search is overall helpful and cases where it is overall harmful. #5 could focus this research on the most important cases.

Conclusions and further research

The key findings from this research are that:

1. People using models in insurance search for errors and adjust models in ways that favor pre-existing views of the risks being modelled.
2. This pattern of biased error search and adjustment is driven by confirmation bias and automation bias.

Some key remaining questions include:

1. When is biased error search helpful and when is it harmful? While many psychologists and cognitive scientists see confirmation bias as pathological (e.g. Lilienfeld et al. 2009), others see it as an efficient way for imperfect people to quickly arrive at "good enough" answers (e.g. Gigerenzer and Goldstein 1996). It is unclear which perspective is most applicable to biased error search in insurance.
2. Would the empirically-tested techniques for mitigating confirmation bias and automation bias help in insurance contexts? Little is known about the external validity of this research, and much of it focuses on undergraduates, challenges from other fields (such as aviation), and occurs in a controlled laboratory setting.

These questions could be resolved by following the recommendations outlined in the previous section. Studying other fields that use imperfect models as an important input to decision-making, and learning about how they handle these challenges could potentially help make more effective use of models in insurance. These fields could include, but would not necessarily be limited to: weather forecasting, finance, trading players in professional sports (especially baseball), business forecasting, climate modelling, and macroeconomic modelling.

There is also a substantial psychology and cognitive science literature comparing expert predictions with both (i) the predictions of simple statistical models, and (ii) the predictions experts make when they are allowed to see the predictions of simple statistical models and adjust them in ways they think are appropriate (e.g. Dawes and Corrigan 1974). This literature could have something to say about how to improve the use of models in insurance.

There is also a substantial literature on publication bias in science. Publication bias is the tendency of researchers and journal editors to favor positive or surprising scientific results, which can make the published literature fail to reflect the whole space of scientific results (including unpublished ones). Publication bias may be driven in part by a form of biased error search with motivated cognition, where scientists search for ways of analyzing a problem which would result in a more interesting or publishable research finding. A number of different techniques for mitigating publication bias have been recommended, and some of them could carry over to managing biased error search in insurance.

Closer study of these fields, with an eye to applications for biased error search in insurance, could shed light on the consequences of biased error search or generate new hypotheses for how to reduce it.

References

Anderson, C.A. (1982). Inoculation and counter-explanation: Debiasing techniques in the perseverance of social theories. *Social Cognition*, 1, 126–139.

Anderson, C.A., & Sechler, E.S. (1986). Effects of explanation and counterexplanation on the development and use of social theories. *Journal of Personality and Social Psychology*, 50, 24–34.

Bahner JE, Huper AD, Manzey D. (2008). Misuse of automated decision aids: Complacency, automation bias and the impact of training experience. *Int J Hum Comput Stud* 66, 688–99.

Bailey NR, Scerbo MW. (2005). The Effects of Operator Trust, Complacency Potential, and Task Complexity on Monitoring a Highly Reliable Automated System. *Dissertation Abstracts International: Section B: The Sciences and Engineering*. US: ProQuest Information & Learning.

Cook, M. B., & Smallman, H. S. (2008). Human factors of the confirmation bias in intelligence analysis: decision support from graphical evidence landscapes. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50(5), 745-754.

Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin*, 81(2), 95.

Dixon SR, Wickens CD. (2006). Automation reliability in unmanned aerial vehicle control: a reliance-compliance model of automation dependence in high workload. *Hum Factors* 48, 474–86.

Dzindolet MT, Peterson SA, Pomranky RA, et al. (2003). The role of trust in automation reliance. *Int J Hum Comput Stud* 58, 697–718.

Dedwards, M., and Hoosain, Z. (2012). The Philosophy of Modelling. Presented to the Staple Inn Actuarial Society. URL:http://www.sias.org.uk/diary/view_meeting?id=SIASMeetingJune2012. Accessed: 2013-10-29. (Archived by WebCite® at http://www.webcitation.org/6Kjf04GVO)

Evans, J., Newstead, E., Allen, J., & Pollard, P. (1994). *Debiasing by instruction: The case of belief bias. European Journal of Cognitive Psychology*, 6, 263–285.

Feynman, R. P. (1974). Cargo cult science. *Engineering and Science*, 37(7), 10-13.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological review*, 103(4), 650.

Grubb PL, Warm JS, Dember WN, et al. (1995). Effects of Multiple-Signal Discrimination on Vigilance Performance and Perceived Workload. *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society, 1360–4.

Hirt, E., & Markman, K. (1995). Multiple explanation: A consider-an-alternative strategy for debiasing judgments. *Journal of Personality and Social Psychology*, 69, 1069–1086.

Hoch, S.J. (1985). Counterfactual reasoning and accuracy in predicting personal events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 719–731.

Koriat, A., Lichtenstein, S., & Fischoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 107–118.

Kray, L. J., & Galinsky, A. D. (2003). The debiasing effect of counterfactual mind-sets: Increasing the search for disconfirmatory information in group decisions. *Organizational Behavior and Human Decision Processes*, 91(1), 69-81.

Kurtz, R., & Garfield, S. (1978). *Illusory correlation: A further exploration of Chapman's paradigm. Journal of Consulting and Clinical Psychology*, 46, 1009–1015.

Lerner, Jennifer S.; Tetlock, Philip E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, 125(2), 255-275.

Lilienfeld, S. O., Ammirati, R., & Landfield, K. (2009). Giving debiasing away: Can psychological research on correcting cognitive errors promote human welfare?. *Perspectives on Psychological Science*, 4(4), 390-398.

Lord, C., Lepper, M., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology*, 47, 1231–1243.

Masalonis AJ. (2003). Effects of training operators on situation-specific automation reliability. IEEE International Conference on Systems, Man, and Cybernetics. Washington DC: IEEE Computer Society Press, 2, 1595–9.

McFadden SM, Vimalachandran A, Blackmore E. (2004). Factors affecting performance on a target monitoring task employing an automatic tracker. *Ergonomics*, 47, 257–80.

McGuirl JM, Sarter NB. (2006). Supporting trust calibration and the effective use of decision aids by presenting dynamic system confidence information. *Hum Factors*, 48, 4656–65

Mosier KL, Skitka LJ, Dunbar M, et al. (2001). Aircrews and automation bias: the advantages of teamwork? *Int J Aviat Psychol,* 11, 1–14.

Mynatt, C., Doherty, M., & Tweney, R. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology*, 29, 85–95.

Newstead, S., Pollard, P., Evans, J., & Allen, J. (1992). The source of belief bias in syllogistic reasoning. *Cognition*, 45, 257–284.

Parasuraman R, Molloy R, Singh IL. (1993).Performance consequences of automation-induced ‘complacency’. *Int J Aviat Psychol*, 3, 1–23.

Parasuraman R, Mouloua M, Molloy R. (1996). Effects of adaptive task allocation on monitoring of automated systems. *Hum Factors*, 38, 665–79.

Sarter NB, Schroeder B. (2001). Supporting decision making and action selection under time pressure and uncertainty: the case of in-flight icing. *Hum Factors*, 43, 4573–83.

Schulz-Hardt, S., Jochims, M., & Frey, D. (2002). Productive conflict in group decision making: Genuine and contrived dissent as strategies to counteract biased information seeking. *Organizational Behavior and Human Decision Processes*, 88(2), 563-586.

Silver, N. (2012). *The Signal and the Noise: Why So Many Predictions Fail-but Some Don’t*. Penguin Press.

Skitka LL, Mosier K, Burdick MD. (2000). Accountability and automation bias. Int J Hum Comput Stud 52, 701–17.

Tetlock, Philip E.; Skitka, Linda; Boettger, Richard. (1989). *Journal of Personality and Social Psychology*, 57(4), 632-640

Tweney, R.D., Doherty, M.E., Worner, W.J., Pliske, D.B., Mynatt, C.R., Gross, K.A., & Arkkelin, D.L. (1980). Strategies of rule discovery in an inference task. *Quarterly Journal of Experimental Psychology*, 32, 109–123.

Wheeler, P. R., & Arunachalam, V. (2008). The effects of decision aid design on the information search strategies and confirmation bias of tax professionals. *Behavioral Research in Accounting*, 20(1), 131-145.

Amlin plc
St. Helen's, 1 Undershaft
London EC3A 8ND
www.amlin.com

Tel +44 (0)20 7746 1000

In association with the Future of Humanity Institute,
University of Oxford

OXFORD
MARTIN
SCHOOL

UNIVERSITY OF
OXFORD