Human complexity bias when choosing safety models and methods

***For* N/A**

January 2025

AB Risk
Limited

www.abrisk.co.uk

Contents

1 INTRODUCTION

Why are we attracted to complexity over simplicity? Could our natural biases fool us? How are we to avoid falling for strong but wrong models and methods?

We use models and representations to better understand system structure, behaviour, and interactions. These methods help us identify opportunities for improvement. Our focus may be on productivity, efficiency, or reliability. The quality of the model becomes especially critical when it is used to study safety risks.

1.1 Natural attraction to complexity

Like a moth to a flame, humans seem naturally drawn to complexity while eschewing simplicity. For example, when purchasing a domestic appliance (e.g., a washing machine), we are often persuaded to select a more sophisticated model with state-of-the-art features. Yet, more often than not, we quickly restrict ourselves to a small subset of the available options (e.g. always using the fast wash cycle because it proves sufficient for most loads).

Strangely, we tend to repeat this pattern with nearly every similar purchase. We might argue that simpler models are rarely available in stores, but instead of using our prior knowledge to seek out simpler options, we are inextricably drawn to the more sophisticated models.

1.2 What does this mean for system models?

Could our natural predispositions lead us to make unwise choices when selecting a system model? Could they distort our perception of the benefits we derive from the model? The following three statements highlight potential causes for concern:

- “All models are wrong, but some are useful” - Statistician George Box (1976).
- "There are two ways of constructing a software design: one way is to make it so simple that there are obviously no deficiencies, and the other way is to make it so complicated that there are no obvious deficiencies" - Computer scientist Tony Hoare (1980)
- “Of two competing theories, the simpler explanation of an entity is to be preferred” - Occam's razor, William of Ockham (14th century).

These quotes provide compelling reasons to exercise caution when working with complex models. It is also important to judge models on usefulness rather than precision.

1.3 Where do complex models come from?

Complex models are often developed by academics motivated by a desire to improve our understanding of how systems work. The resources needed to develop these models are secured by highlighting the shortcomings of existing approaches and the potential risks these pose. Individuals or teams then undertake research to explore options and then develop a new approach, which they test with a limited number of pilot studies.

We might wonder if the human predisposition toward complexity influences funding organisations when deciding which projects to support. Do complex projects sound more interesting than simple ones?

Perhaps academics should not be criticised for creating complex models. However, what about the individuals or organisations that apply these models to real-world scenarios? In high-hazard systems, the stakes can be significant. How objective are people when choosing a method, and what factors might lead them to make poor choices?

2 HUMAN BIASES

Observations over many years have shown that humans display systematic patterns of irrationality or lack of objectivity in the way we perceive situations, make judgments and decide how to act. These biases often arise from cognitive shortcuts that help us navigate the world without requiring a full understanding of every detail. Many are shaped by our experiences, emotions and societal influences.

Biases might explain our affinity for complexity but because they are inherent to human nature they cannot be avoided and we must work diligently to prevent them from adversely influencing our behaviour. The following are examples of biases that could influence our choices of models and methods:

Bias	Definition	Explanation
Action bias	Tendency to favour action over inaction	Using a complex model will require far more work than using a readily available simple alternative. You may believe this will translate into better results. However, using the simpler option may give you more time to think and reflect, which is likely to produce more useful results. This bias is likely to be particularly prevalent in a culture that rewards people who appear to be busy.
Ambiguity effect	Tendency to avoid ambiguous or incomplete options.	As Tony Hoare pointed out, simple models have obvious deficiencies. Complex models also have deficiencies but they are likely to be hidden, giving the illusion of them being the more complete and accurate solution.
Apophenia	Tendency to perceive patterns in random occurrences.	All models are based on known information. The more detailed representation provided by a complex model may give an illusion that additional data points from the past will provide a better prediction of the future. This may be why data correlation is often incorrectly accepted as causation. When patterns emerge in data, people may assume that one variable directly influences another, overlooking other possible explanations such as coincidence, confounding factors, or underlying systemic issues.
Authority bias	Tendency to be influenced by people of authority.	If you believe academics are more intelligent than people in operational roles you are likely to trust their models over the ones that may have a more practical background. The use of clever buzz words and sophisticated spiel may lead you to agree with the academic’s view of the world.
Availability heuristic	Tendency to have greater trust in ideas that come to mind more easily or are more available in your memory.	The detailed representation created using a complex model may give the illusion that it covers every possibility. This may lead to more effort being put into things already known, distracting from considering credible low likelihood events.

Bias	Definition	Explanation
Bandwagon effect	Habit of adopting behaviours or beliefs because other people do the same.	Complex models are more interesting to talk about, with multiple features to discuss. Simpler models create less interest and so can fade into the background. Academics who are expected to publish their ideas are particularly well positioned to share their views to a wide audience, hence creating ample opportunity for the bandwagon effect to occur.
Confirmation bias	Favouring or focusing on information that confirms your existing beliefs and preconceptions.	Because the deficiencies of a simple model are easy to recognise you may question whether it is fit for purpose. Someone promoting a more complex model can support their cause by reinforcing your concerns, especially if they fail to mention the hidden deficiencies of their complex approach.
Dunning-Kruger effect	Inability to recognise your own incompetence.	Having worked hard and diligently to generate a representation using a complex model you may be liable to believe it is correct. This may be reinforced if other people accept it and don’t highlight any deficiencies. But because the deficiencies are hidden this lack of critical feedback reinforces your over estimation of your own capability.
Effort justification	A type of cognitive dissonance that arises when we put a lot of effort into something.	If you have spent a lot of time building a representation using a complex model you are likely to justify this by convincing yourself that it was worth it, clouding your judgement on its accuracy and usefulness.
False consensus effect	Overestimation of how much other people agree with you or approve of your behaviour.	If people do not highlight the deficiencies with your representation you may believe that they accept it as correct. You may find they are more critical of simpler representations, but this is to be expected because the deficiencies are inherently more visible.
Hard-easy effect	Incorrectly predicting our ability to complete a task based on its level of difficulty.	This may mean you are over-confident in the value of your representation using a complex model because it was difficult to develop.
Illusion of validity	Tendency be over-confident in the accuracy of our judgements.	This may give you confidence that you can accurately represent a system. In doing so you will avoid simper models because they have obvious deficiencies. You may feel that the level of detail included when using a complex model is sufficient to overcome the inherent uncertainty that exists with any approach.
The IKEA effect	The tendency to value an object more if you made it yourself.	This may lead you to favour a detailed model that allows you to include multiple inputs to match your specific situation over an off-the-shelf tool because you do not feel you are more involved in its development.

Bias	Definition	Explanation
Zero risk bias	A preference for absolute certainty that risks have been eliminated.	You may perceive that a more complex model gives you more certainty that an undesirable event will not happen.

3 OBSERVED EXAMPLES - RISK MANGEMENT METHODS OF VARYING LEVELS OF COMPLEXITY

Safety cannot can be measured directly. To be able to manage it effectively requires methods of representing the way systems function so that we can identify interventions that will make a difference. Multiple approaches are available to do this. Some are little more than general concepts, which may be thought provoking but do not directly contribute to safety management. However, there are many methods that can have a more direct influence with varying levels of complexity.

3.1 Risk assessment

Risk assessment matrices are a well-established way of evaluating the two components of risk: consequence and likelihood. The simplest versions are typically 3 x 3 matrices, which may be criticised for lacking precision. The standard matrix has likely expanded to 5 x 5, which is generally considered reasonable. However, examples of 9 x 9 matrices now exist. These are clearly more complex to use, with questionable additional benefit. It seems likely that people start to believe that complexity provides more precision, that would be desirable for determining absolute risks. But the scale used on the rows and columns are often orders of magnitude, which even for a 5 x 5 matrix results in a risk range of eight orders of magnitude, which makes no sense. Simple risk matrices have been very effective at increasing understanding of risk as a product of consequence and likelihood but should only be used for risk ranking, typically low to high. Additional complexity does not contribute to this.

One of the most significant criticisms of risk matrices is their subjectivity. This is a valid but obvious limitation. More formal and complex approaches, such as Quantified Risk Assessment (QRA), are often perceived as being more objective and therefore more accurate. But QRA relies on numerical estimates, typically selected from a database, with the analyst using subjectivity to determine which values are most appropriate.

Risk assessment is inherently subjective. Whether risks estimated using a complex QRA are more accurate than those assessed with a simple risk matrix is debatable. However, the fact that subjectivity in QRA is concealed within the model is certainly a cause for concern.

3.2 Root cause analysis

The late Professor James Reason's developed his Swiss Cheese metaphor to illustrate, in a simple way, how multiple layers of protection work together to prevent accidents. No layer is perfect and when weakness coincide the control of hazards can be lost leading to undesirable consequences. The representation does not directly provide a practical method but the underlying principles are consistent with various methods of root cause analysis.

A commonly stated criticism of the Swiss Cheese metaphor is that it implies a linear progression of events. This seems to be unjustified because not all the slices of cheese, with several usually representing management and organisational factors.

More generally the whole concept of root cause analysis is often criticised by people who say that looking for 'a' root cause is too simplistic. The explanation is that accidents have multiple causes of different types. This seems to be a misrepresentation because root cause analysis has never been promoted as a method of finding a single root cause. Whilst a strict application of '5 whys' approach to root cause analysis is too simplistic, more free form causal trees can be very effective at identifying multiple causes of accidents, including root causes, whilst still be very simple to develop and understand.

Those who criticise root cause analysis are somewhat vague about what to do instead. They seem to favour applying vaguely defined concepts such as 'system thinking' to develop complex webs that illustrate interactions between multiple technical, human and organisational components. The implication is that you have to generate a unique complex model for every incident that occurs.

A number of methods have been developed to support incident analysis including:

- Failure Modes and Effects Analysis (FMEA)
- Fishbone diagrams;
- Human Factors Analysis and Classification System (HFACS)
- AcciMap
- Causal Analysis based on System Theory (CAST).

These can be useful but require skill and resource to apply. This results in fewer people being able to analyse incidents and without limited resources fewer incidents will be analysed. More significantly use of specific and complex methods may reduce the level of understanding within an organisation of why incidents are occurring and how they may be prevented.

3.3 Hazard analysis

A Hazard and Operability (HAZOP) study is a well-established method for performing hazard analysis. It uses the system's Piping and Instrument Diagrams (P&ID) as its foundation, which are familiar to the individuals involved in the study. The method employs a range of relatively simple guidewords designed to prompt discussions about potential hazardous scenarios.

HAZOP is sometimes criticised for its reliance on human judgment and its focus on single failures. There is a debate about whether, and how much, risk evaluation should be included. It is clearly an additional activity that may stall the process leading to a lack of engagement, especially if the HAZOP team does not have the appropriate skill set for risk evaluation. For these reasons, risk evaluation is typically best avoided or used as a relatively simple screening process, with more difficult territory referred to a separate risk evaluation process. This has led to suggestions that more sophisticated methods should be used because they offer better insights.

Systems-Theoretic Process Analysis (STPA) is often promoted as a superior alternative to HAZOP due to its more advanced safety control models, which can identify a broader range of potential hazards. It is reported that STPA can better represent the complexity of systems compared to simpler methods. However, it is a complex method and a theoretically more accurate representation is not necessarily more useful, especially if the additional content if related to remote factors outside the control of the organisation.

STPA and CAST (incident analysis – see above) are based on the System-Theoretic Accident Model and Processes (STAMP) framework, developed and advocated by a small group of academics. The documentation for these approaches spans hundreds of pages. Compared to the textual tables created by a HAZOP study, the representations produced by STPA and CAST could objectively be described as strong models. The academic advocates of the approach are quite quick to emphasise the limitations of other methods but do not seem to acknowledge any limitations with the STAMP framework.

3.4 Task analysis

Task analysis is used to understand how tasks are performed. In a safety context the main aim is to identify potential for human error and how the associated risks are controlled. This usually involves creating a structured representation of the task method, performing a human error analysis and evaluating the risk controls and Performance Influencing Factors (PIF).

There has been a push to adapt task analysis to align with a more academic and complex perspective on human behaviour. Classifying human error types is often advocated to fit within theoretical frameworks. Additionally, risk control is frequently divided into preventative and recovery measures. There has even been a challenge to the notion of human error, as it focuses on the individuals involved in a task, whereas the primary failures often stem from the systems and environments in which people work. Task analysis can be a simple process. These additional features add complexity with uncertain benefit.

The Functional Resonance Analysis Method (FRAM) is another approach developed and advocated by academics, who suggest it is a better method of representing the way tasks are performed and potential failures (errors) because it integrates human actions within the wider system. FRAM generates a complex model composed of functions and activities, represented using six basic characteristics. While academics appear confident in the method's capabilities, they remain largely silent about its limitations, aside from acknowledging that it requires significant resources.

3.5 Bow-ties

Bow-ties are used to represent and communicate risk in a structured way. The name comes from its shape, which resembles a bow-tie, with a hazard at the centre, potential threats (or causes) on the left, and possible consequences on the right. Barriers are identified to reduce risk by preventing the progression from threat to consequence.

Much is made of the utility of bow-ties in facilitating communication, but the effectiveness of this communication depends on people being able to read and understand the bow-ties. A full understanding requires reading and interpreting each element, but other methods such as LOPA (see below) are usually better at conveying this level of detail. Using a bow-tie to give an overview of arrangements requires the number of threats, barriers, and consequences to be kept to a minimum. Attempts to create a more comprehensive representation or to include quantification adds complexity and makes the bow-tie much harder to interpret. Some have suggested creating multi-level bowties, which only exacerbates the issue.

Advocates of more complex approaches of representing system (including but not limited to the method described in this paper) tend to be dismissive of bow-ties, viewing their simplicity as a weakness rather than a strength.

3.6 Layers of Protection Analysis (LOPA)

LOPA is a risk assessment method used to evaluate the effectiveness of different controls known as layers of protection in preventing hazardous events. It helps determine whether existing risk reduction measures are sufficient or if additional controls are needed.

Guidelines used in LOPA typically include generic data for risk reduction achieved by different types of controls. These values are usually presented in orders of magnitude rather than precise probabilities. The aim is to provide a reasonable assessment of whether the controls are suitable and sufficient overall. For simpler systems with truly independent layers the approach can be equivalent to a full QRA. In many cases the relatively broad categories used mean that it is better viewed as being semi-quantitative, providing an input into a more wide ranging risk evaluation risk.

The validity of these generic guidelines provided by guidelines used in LOPA is often questioned, particularly in the context of human error, where generic data is considered inappropriate. Some suggest that a formal and more complex calculation method should be used instead, such as:

- Human Error Assessment and Reduction Technique (HEART)
- Technique for Human Error Rate Prediction (THERP)
- Standardized Plant Analysis Risk Human Reliability Analysis (SPAR-H)

All of these quantified techniques face the same challenges as QRA (see above). A more specific issue is that they generate data points that are inconsistent with the order-of-magnitude figures for technical failures provided in LOPA guidelines. Performing human reliability calculations takes significantly more time compared to using generic figures.

4 USEFULNESS

"All models are wrong, but some are useful," highlights that the value of a model or method should not be judged solely on its technical merits but also on its practicality. This is something often overlooked by those promoting complex approaches. Perhaps usefulness is not the primary concern of academics who develop these methods, but it certainly should be for those seeking to make a real impact on safety.

4.1 Judging usefulness

A useful model or method represents reality in a way that helps users make informed decisions. A model that only produces theoretical knowledge may fundamentally fail to do this or possibly lead people to be overconfident in the decisions they make. Numerical methods can be particularly problematic because people have a tendency to incorrectly associate the apparent precision with accuracy.

Unnecessary complexity makes it harder to understand and communicate what is important. Overly simplistic approaches give results that are too broad and vague to support decision making. This may seem at odds with Occam's Razor that says we should always select the simplest approach. However, we can easily expand the concept to say we should use the simplest useful approach.

4.2 Other effects on usefulness

A model that is difficult, expensive, or time-consuming to apply will inevitably be used less frequently and by fewer people. This limits its impact, particularly in time-sensitive or resource-constrained environments. The most effective models strike a balance between depth and usability, providing valuable insights without being burdensome to implement.

If the results are meant to be applied in practice, decision-makers must understand how the model works and why it produces certain outputs. Black-box models that generate conclusions without clear explanations can lead to distrust and hesitation in decision-making. A useful model should offer transparency, allowing users to trace the reasoning behind its conclusions.

4.3 Some examples of usefulness

Risk assessment matrices and the Reason's Swiss Cheese metaphor have had a profound effect on the way people view safety. This seems to be overlooked by people promoting more complex methods.

Without matrices the concept of risk is quite intangible. It can be described as a product of consequence and likelihood, but this is does not give people a method to determine risk in any objective way. Whilst risk assessment matrices can be criticised for relying on subjective ratings of both consequence and likelihood, and for combining them in a way that defies mathematical logic, they allow people to explain how they have decided whether a risk is acceptable or not. Other people may disagree with the result but the matrix allows them to explore the source of that disagreement. They are simple and flawed, but have proven to be useful.

Similarly, before Reason published his Swiss Cheese metaphor people were inclined to view causes of accidents at a superficial level. Safety specialists may have understood the usefulness of delving deeper to find management, system and organisational factors but most people struggled to see how this would result in better safety outcomes. Although it does not directly result in a useful tool the Swiss Cheese metaphor has been widely used when explaining the concept of multiple layers of control and its graphical representation has proven highly effective at communicating the concept.

4.4 Devious uses

It would be a little naïve to assume that everyone wants transparency in the information used to make decisions. If someone has a vested interest they may view a complex approach that hides its deficiencies and can be manipulated to present a desired outcome as useful. They can be very effective at engineering the outcome if they present their methods and results in a way that will influence decision makers due to natural human bias. Numerical methods can be particularly strong in this regard, especially if precise results are quoted confidentially that align with what people want to hear. For example, being told an intervention could 'save up to £10millon per year' is likely to be more interesting than a range of possible and likely outcomes, especially if negative outcomes could be equally likely.

5 CONCLUSIONS

The purpose of this paper is not to suggest that complex models and methods should never be used but to highlight that human biases may lead us to prefer them over simpler alternatives. Our main interest should be usefulness and how an approach leads to better decisions and tangible improvements. It should not only diagnose a problem but also help users identify solutions and assess their effectiveness.

5.1 Multiple biases

To summarise, the ambiguity effect causes us to perceive complex models as superior because their deficiencies are less obvious. The availability heuristic can further reinforce this if a complex model appears to explain familiar issues, even though the less familiar issues could be the greater concern.

Those who promote complex models often benefit from authority bias because their perceived expertise makes their ideas more convincing. They further strengthen their position by emphasising the obvious flaws of simpler alternatives, reinforcing or even creating confirmation bias in their audience. Additionally, by using multiple channels for promotion and cultivating a dedicated following they create a bandwagon effect making their approach seem more widely accepted and credible.

Somewhat perversely, action bias means that the additional effort required to use complex models and methods increases our interest in them. The IKEA effect further reinforces this if we perceive the results as being directly the result of our own work. The hard-easy effect gives us confidence that the effort was worthwhile, while the Dunning-Kruger effect may lead us to believe that our approach was correct, even if we lack the expertise to accurately assess it. If no one challenges our approach, the false consensus effect further reinforces our belief in its validity, even though the complexity makes it difficult to fully understand what we have done or recognise its deficiencies.

5.2 Sunk cost fallacy

One observed trait that is not typically classified as a cognitive bias, but is particularly concerning, is the sunk cost fallacy. It causes us to persist with something we have already invested in, even when evidence suggests that it is not working or that an alternative approach would be more effective. Once we have dedicated time and resources to obtaining or developing a complex model or method, we naturally want to use it at every opportunity. This can lead us to overlook, or even actively avoid, simpler alternatives, even when they may be more effective.

5.3 Box, Hoare, and Occam

Combining the statements quoted at the start of the paper may lead us to the following statement.

Simple models expose their deficiencies making them easier to recognise and manage, enhancing their usefulness. Complex models conceal their deficiencies leaving them unknown and unmanageable, diminishing their usefulness.

5.4 Looking to the future

No discussion on models and methods in 2025 is complete without mentioning Artificial Intelligence (AI). AI applications are undeniably highly sophisticated, yet they also have well-documented limitations. Interestingly, one common criticism of AI, that its results tend to be verbose, provides some mitigation against the issues discussed in this paper because it allows us to explore the result and understand the reasoning. However, as AI technology continues to evolve, we must remain increasingly vigilant to recognise and counteract our natural biases.

5.5 Closing remark

Having read this paper you may feel a little foolish if you have at some time in the past invested your valuable time and resources into using a complex approach when a simper one was available. But you should take some reassurance that your inherent human biases made you vulnerable.

It is OK for you to consider and try different approaches being promoted but being aware of your biases and recognising the methods used to promote strong but wrong methods may help you in the future. A few catchy buzz words and criticism of simper alternatives should alert you to be cautious. Bearing in mind the time you need to understand their method leads to the sunk cost fallacy, especially if you decide to give it a try.

Albert Einstein said that “the definition of genius is taking the complex and making it simple.” Remembering this should at least help you repel authority bias.