Complexity for complexity—How advanced modeling may limit its applicability for decision-makers

Ale, Ben J.M.; Slater, David H.

DOI
10.1111/risa.14261

Publication date
2023

Document Version
Final published version

Published in
Risk Analysis

DOI: 10.1111/risa.14261

PERSPECTIVE

Complexity for complexity—How advanced modeling may limit its applicability for decision-makers

Ben J. M. Ale[1] | **David H. Slater**[2]

[1]Departement of technology, policy and management, Technical University Delft, Delft, The Netherlands

[2]Carey Dene, Carey, Hereford, Herefordshire, UK

Correspondence
Ben J. M. Ale, Technical University Delft, Mekelweg 5, 2628 CD Delft, The Netherlands.
Email: ben.ale@xs4all.nl

Abstract

As today's engineering systems have become increasingly sophisticated, assessing the efficacy of their safety-critical systems has become much more challenging. The more classical methods of "failure" analysis by decomposition into components related by logic trees, such as fault and event trees, root cause analysis, and failure mode and effects analysis lead to models that do not necessarily behave like the real systems they are meant to represent. These models need to display similar emergent and unpredictable behaviors to sociotechnical systems in the real world. The question then arises as to whether a return to a simpler whole system model is necessary to understand better the behavior of real systems and to build confidence in the results. This question is more prescient when one considers that the causal chain in many serious accidents is not as deep-rooted as is sometimes claimed. If these more obvious causes are not taken away, why would the more intricate scenarios that emanate from more sophisticated models be acted upon. The paper highlights the advantages of modeling and analyzing these "normal" deviations from ideality, so called weak signals, versus just system failures and near misses as well as catastrophes. In this paper we explore this question.

KEYWORDS
causal chain, complexity, FRAM, sociopolitical context

1 | INTRODUCTION

Assessing the efficacy of safety-critical systems in today's complicated engineering applications is a challenging responsibility. Traditionally, this has been done by establishing logically the causal links between specific components and the consequences of their failure. The standard way of doing this is either by inspection, qualitatively (failure modes and effects analysis, FMEA), or more rigorously by logic, or decision trees, which can be quantified using Boolean algebra such as fault tree analysis. Given these predetermined and fixed relationships, observance of such effects could be assumed to have occurred as a consequence of the appropriate component failure. If we know the reliability of these components, we can then establish a system integrity level for our system. If the safety system itself is a component, then a more extensive set of fault trees is assumed to predict the overall integrity or safety of the whole system.

But today's systems have tended to become ever more complicated and qualify as at least complex (Snowden & Boone, 2007), if not occasionally chaotic, with the increasing involvement of artificial intelligence in their control and safety management. If we add to this the fact that human intelligence is also intimately involved in these issues, although often conveniently ignored, then the now complex sociotechnical systems have yet more—human—factors to control. The challenge of now analyzing formally what goes on in these so-called complex sociotechnical systems is a major impediment to our achieving an adequate understanding of how they behave and how safe they are in operation. To do this, it is suggested by Leveson (2023) that we need a paradigm shift in our approach by moving the focus from preventing errors to enforcing constraints on the behavior of components and interactions between components. In doing so, the variability is reduced and so is the complexity (Ale et al., 2019a, 2019b; Baumgarten & Malakis, 2023; Franchina, 2023).

In these more complex systems, there are at least two new issues to address. First, the simplistic fixed linear cause and effect relationships are no longer reliable predictors of performance or accurate symptoms for diagnosing "faults" from observed effects. This is primarily a result of more "complex" behaviors, in which the decidedly nonlinear interdependence of interacting agents with the inherent variabilities in conditions in the real world means we have to allow for sometimes unexpected and unpredicted "emergent" outcomes in these systems. Second, effects can be seen to have no obvious relationships to the tidy sequential linear logic diagrams used in their design.

But if we want to establish quantitatively, the reliability and hence safety of these systems, we must have some way of legitimately and accurately modeling these systems and their inevitable interdependencies and interactions with each other and the real world in which they operate.

One of the options is to utilize the Functional Resonance Analysis Method (FRAM; Hollnagel, 2012) to include nonlinear relationships and predict emerging behaviors in such systems. These models can then be used as the equivalent dynamic logic trees to traditional FTA (Fault Tree Analysis) and can be quantified using reliabilities as conditional probabilities of success or failure. These are now dynamic Bayesian nets, rather than static Boolean gates, allowing for Markovian development of instantiations of the models to predict these emergent effects (Hanea et al., 2012).

This is very much in line with the traditional engineering approaches to assessing and assuring the safety of systems. The only difference is that we have now evolved the static Boolean fault trees into the more advanced Markovian and Bayesian belief nets (BBNs; Slater, 2023), and we can use AI (Artificial Intelligence) technologies and big data (Ale, 2016) to feed the models we have created.

As the topic we want to discuss emerges in part from the SAFETY II/resilience development, we need to summarize and mention some possibly unexpected and unwanted side effects of this development because the shift of attention to success entails the need for more complex modeling.

This is all progress, but two questions need to be raised regarding whether this progress will do us any good, when doing good is defined as allowing us to make the world a safer place: Does complexity just replace complexity; and will we be inclined to act on the results emanating from the models?

2 | COMPLEXITY REPLACES COMPLEXITY

The development of risk analysis methods originated from attempts to eliminate all failures from a future rocket launch. (Ericson, 1999; Watson, 1961). Fault trees ordered the paths to potential failure in a consistent and logical model, which lends itself to quantification. Even in the 1960s, a rocket and its launch was not a simple linear system. Even though keeping a rocket upright is governed by well-understood physics, actually doing it involves the interaction of several systems: the engine, the gyroscopes, the crosswind, gravity, and people, to name a few. What the early developers tried to do was to eliminate nonlinearity as much as possible, account for common mode and common cause failures, and reduce the number of paths to failure to just the unavoidable. In doing so, they adhered to existing engineering standards such as removing single points of failure and removing the common cause failures when their analysis allowed them to detect one.

For the quantification of fault trees and the final estimate of the probability of failure, data on the failure rates of components were needed. Originally, they were point estimates. Later, it was possible to treat these as distributed variables in BBNs such as in models for air transport safety (Ale et al., 2009). This did not resolve the issue of uncertainty, as technicians, statisticians and experts can differ widely on the value estimates they derived from historical data.

Human beings involved in the processes were treated as components with their own probability of failure. Therefore, the quantification of human failure rates was a necessary component of the analysis (Swain & Guttman, 1983). However, Swain et al. (1963) remarked that the quantification of human error predictions will continue to be piecemeal and years behind hardware system reliability analysis. So, it is more useful for designers to continue to design the human out because they feel that human–machine reliability cannot be determined as accurately as hardware reliability. The main problem they recognized was that it is not sufficient just to have data on the errors or failures, and one also needs data on the opportunity for failures and/or successes. This is called the denominator problem. One needs to know the number of opportunities that result in a failure as a fraction of all opportunities in order to estimate the importance of the action. One also needs to know the fraction of a particular action, error, failure, or deviation in the population of failures as compared with the fraction of the same action in the population of success to be able to judge whether that action is indeed in the causal chain of the particular failure. As an example, one needs to know in how many cases of clipping the wall with a Formula 1 car damages the drive train and in how many cases one gets away with it before one can decide that clipping the wall constitutes a problem. According to Swain et al., without denominator data, the probability of human failure leading to an accident can be significantly overestimated. This in turn led them to the development of the technique for human error-rate prediction; Bray, 1962; Shapero, 1960).

At the time, the organization around rocket launches was indeed single-minded, in the sense that there were no other objectives than to have a successful launch; and money was definitely not an issue. In the further development of the space program, though this has changed, saving money, efficiency and commercial incentives were a significant influence on the organizational side of the program, and hence a significant influence on the chances of success (Vaughan, 1996). These additional complexities led to the notion that accidents in these systems are "normal" (Perrow, 1999) and thus unavoidable.

Nevertheless, the probability of failure in many of these systems is low, suggesting that it could be a better idea to look at the reasons for success rather than the reasons for failure. This was also suggested by Swain et al. They observed that the rate at which a human action led to a failure was much smaller than the error rate predicted by psychologists, implying that there was a recovery mechanism that needed to be taken into account. It was assumed that the "linear" analysis methods were insufficient to support further improvements in situations where there are so many interactions between humans and their organizations on the one hand and technology on the other and that they were inseparable.

This focus on success and resilience led to new analysis methodologies, such as Functional Resonance Analysis Model (Hollnagel, 2011) and more convoluted modeling than the more or less straightforward forward fault trees, BBNs, and event trees. These additional methods, however, resuscitate the problem of definition. Is success the absence of failure (Ale et al., 2019a, 2019b), or is there something between failure and success just as there was something between human error and human failure in the 1960s? And similarly, what constitutes success: completing a mission or survival of a crew, such as in the case of the Apollo 13 mission (HTTP1)? Another persistent problem is again that of the denominator. Take the case of signal passed at danger (SPAD) incidents in rail transport; here, one can count the number of cases in which a train encountered a signal at danger and stopped as a fraction of the total that approached. But obtaining data on the total number of missions, which would allow the estimation of the probability of failure or of success in terms of fractions per mission is difficult and expensive (Ale, 2006; Ale et al., 2006) if not impossible. Most of the failure data are in the form of the number of failures per unit time. They can be counted, at least in principle, and given a definition of failure. But how many successes are there in the same period of time? (Rashidy et al., 2018).

The use of artificial intelligence to extract data from accident and incident reports—in the light of the blatant absence of more than anecdotal success stories—creates an additional problem. The AI is another expert, but this time without any professional credentials. How are we to value the estimates made by AI, and how can we validate these estimates, even if only by a qualitative description of inherent uncertainties (Oviedo et al., 2023).

The idea behind creating models of a system is to try and overcome the intractability of man–machine systems. Such a model should provide insight into what could be done to make the system more successful, if desired. However, the more the model becomes like the real system, the more it also acquires the properties of the real system: And in the way it functions, it becomes similarly intractable. In addition, the data with which the model are fed also becomes intractable when there is an AI system between the real world and the data used. Therefore, there is the possibility that we may need another, more simplified model, to understand the behavior of the model itself tractable.

Nevertheless, even an intractable model may be useful as it can be run in accelerated time. Thus, those failures that have a low frequency of occurring still may occur during the running of the model, and something can thus be done about the causes before the failures appear in the real world. It would create hindsight before the fact. That is, assuming that the model is not so intractable that even the causes of an emergent failure can no longer be found, and then yet another model would be needed to describe and trace the actions of the event.

Investigating potential future behavior of complex systems, by developing an even more complex model might look like an appealing idea. However, the indication of anomalous behavior, which can be a signal (indication or warning presented by a potentially adverse outcome from a complex non-linear interaction in the system), may be much weaker than the simpler cause/effect paths to disaster traced out by an FTA analysis (which only describe "linear" interactions), even if these logic trees ignore more complex resilience effects present in the system.

Such models can still cover all events from things that go wrong, using so-called "big data" (Sing & Van Gulijk, 2023; Sing et al., 2023) and produce insight into measures to increase the number of successes, thereby decreasing the number of SPADs (Rashidi et al., 2018), given that the total number of signal passages remains constant (Hollnagel, 2014, Chap. 9). The question that remains then is, whether or not, observations of these signals and identification of measures to increase the likelihood of success are likely to be converted into action.

Even complex systems can be protected by (old-fashioned) engineering barriers. However, in complex systems, it is difficult to detect the failures of these barriers because they do not always immediately lead to an obvious failure. This difficulty is often referred to as the weak signal problem. Modeling these systems beforehand should help to detect these weak signals. The question still remains as to whether they will. There are many examples of weak signals. We will discuss the problems with detecting them in the next sections. The near miss is another example of a signal than can be weak or strong, depending on its consequences and its context, and therefore again, warrants a separate description.

3 | WEAK SIGNALS AND NEAR MISSES

What constitutes a weak signal is ill-defined. Guillaume (2011) gives a definition depending on a number of elements:

- Weak signals are qualitative, ambiguous, and/or fragmented information so that their threat to safety is neither clear nor direct.
- They are characterized by ambivalence, to the extent that they are hardly noticeable, but they offer the potential of anticipation of future accidents.

- Weak signals arise from the system of management of technical failures; the relative inadequacy of the bureaucratic organization works and limits to the proactive and reactive analyses of risks and the distribution of that information to those needing it.
- The overarching root of the "weak signal problem" is attributed to organizational complexity that leads to underestimation of risks.

In any case, the difference between work as done and work as planned, or technology as it works and technology as it was designed to work, constitute signals that are strong when they result in a detrimental incident and have the tendency to stay weak and become increasingly weaker when they do not result in an incident.

Nevertheless, at each deviation, the question should be raised as to why it was convenient or necessary to deviate from the "norm." If there is a more convenient way to do the work without entailing increased or even unacceptable risks, then that alternative way of doing the work should be incorporated into the rules. If the work cannot be done other than by violating the rules, then the whole description of the work including the rules needs to be revised, or the work should be abandoned. Allowing deviations from rules because they are "unworkable" is contagious and therefore may spread to rules and regulations that are essential for the delivery of the work, such as the quality of products and the safety of workers and clients and therefore should not be violated (Ale et al., 2019a).

In systems with defense in depth, especially, it is unlikely that the breach of a single layer of defense will result in an incident. That is what the defense system was designed for. However, persistent breaches of defenses increase the probability of simultaneous breaches of multiple and possibly all layers of defense, resulting in an incident or accident.

Therefore, deviations should be taken seriously, and organizations should be aware of the potential consequences of ignoring them (Rasmussen, 1997, Slater & Ale, 2022).

3.1 | Weak signals are difficult to detect

As remarked before, the problem with near misses, violations, deviations, and all other differences between work as done and work as planned, designed or foreseen, is that they might end up in a post-accident causation sequence. In that case, the question will be raised as to whether their occurrence in that sequence is because of hindsight bias—it is relatively easy to trace an accident back along a causal path—or the result of pre-sight bias: the tendency to dismiss deviations that are for now considered as not having the potential to cause harm just because until now they have been inconsequential.

In aviation and to a certain extent in the medical environment, the reporting of near misses is an accepted practice. But deviations are much harder to spot. Not only because people do not consider them a problem if the work gets done, but also because the "work as the designers thought it would be done" often is not known to the workers (Van Gulijk et al., 2009). They are convinced that they are doing the right thing right. Even in aviation—with the long practice of reporting in a just culture—an extensive program is needed to detect deviations before they become the cause of an accident (Patriarca et al., 2022), or rewrite the "work as imagined" to include a deviation as normal (work as done), so that it is not a deviation anymore.

If we set out to monitor these signals, we can potentially gain a significant advance in managing safety in complex systems. Monitoring can allow us to learn the routine operations of the system and to establish patterns, which can allow the detection of weak signals or anomalies. These days, machine learning algorithms can make this straightforward. Having learned what anomalies look like, it is then possible to predict what consequences will develop from these excursions. This allows us to anticipate problems and respond by taking actions to mitigate or compensate and return the system to "normal." This fits in with a harder, better definition of "resilience," as having this ability to learn, anticipate, and adapt, as formal design requirements for complex systems (Hollnagel, 2014).

The obvious question is then how to make the distinction. What is still an acceptable variation on the theme and what constitutes a dangerous action or practice? What certainly does not help is that people are intrinsically optimistic (Hoorens, 2014). The bearers of bad news are rarely praised, and it is difficult to convince people that an action can be harmful, especially if until now it has worked well. "This will not happen to us" is engraved in the human brain (Ale, 2003). Even a program to observe the work as done can only be successful if the work as imagined is adequately described; the observers have a clear understanding of the work as imagined and have a sufficiently unbiased imagination, to envisage possible disasters, if any, resulting from the work as done, as they observe it. It seems a reasonable assumption that those who "imagined" the work, have designed the operational envelope of hardware, software, and people in such a way that the residual calculated risk of collateral damage is acceptable. Whether there has been an investigation into the consequences of an egress outside this operational and sufficiently safe envelope is rarely known to the operator. Therefore, when work as done differs from work as imagined, there is good reason to investigate carefully, whether these deviations create additional risk, regardless of the reasons why the deviation occurred. If they do not, the operational envelope can be extended. There is sufficient evidence that this sort of deviation is more likely to be the result of the desire to be more "efficient," Hollnagel's Efficiency Thoroughness Trade-off—ETTO (HTTP16), that is, have the work done faster and/or make more money, than from necessity, that is, because the work can be done more safely this way. On the other hand, Hollnagel might have underestimated the imagination of the designers (Leveson, 2020). The scholars in SAFETI II and resilience engineering also seem to overestimate the potential for resilience, for a person unilaterally "breaking the rules" and who might be killed due to his

assumption that he can vary the set procedures. Resilience is about recognizing and designing in the formal ability to learn, anticipate, and adapt to real-life conditions and incorporating the learnings in a better design. Doing something more efficiently is not an error, it is smart, but it may also be a step on the road to disaster. Making the right choice in the ETTO is easy to do when a wrong choice manifests itself in a disaster immediately. The trade-off is much more difficult when the probability of a mishap is low. It is even more difficult when the probability of a mishap is high or even certain, but the effect is delayed and the choice for efficiency is profitable (Hollnagel, 2009).

In fact, work as prescribed is often not as efficient as work as done; and sometimes work as prescribed is impossible in practice. Therefore, in the more complicated sociotechnical systems, there are always defects, errors, faults, and deviations. The result of industrial action in the form of "working to rule" or "according to the book" is a constant reminder of this fact. The observation that there does not seem to be a tendency to change "the rule or the book" after the industrial action is over and that the deviations from "the book" continue to be allowed and praised may be the strongest "weak" signal of all. The multilayered defenses or defenses in depth, including the ingenuity of operators, protect the system against accidents caused by these deviations and make it resilient, but that only goes so far (Cook, 2000).

3.2 | A near miss is also a signal

A near miss is also called a close call and can be defined as an unplanned event that has the potential to cause but does not actually result in damage or human injury. The most noticeable near miss is one in which we are ourselves a participant. One could call it the "oops" moment; Something that you instinctively recognize at the time. It usually comes with a surge of adrenaline and relief (HTTP2). An example is braking just in time before a child emerges, suddenly crossing the road, or a Formula 1 driver clipping the wall but still able to continue racing.

More layered instances are involved in surviving a car crash, where the driver forgot to wear a seat belt, but the airbag saved his life, or a person on a ladder mounting solar panels on a sloping roof, losing his balance and the user being saved by his fall arrester. In the latter case, four people not using a fall arrester lost their lives in the first half of 2023 alone (HTTP7, HTTP8).

The more complicated and less obvious near misses occur in multilayered safety systems: a system with two relief valves of which one is found to be inoperable for a long time or even multiple layers of defense being broken, but the last one prevented a disaster. Similarly unused high, high-level alarms in tank overfilling safety systems are rarely used and can be in a failed state unnoticed until they are needed, such as what happened in the Buncefield accident (HTTP3).

When we ourselves are involved in a near miss, we can judge how close we came to an incident, or accident, and whether or not we want to take the same risk again: that is, drive fast through a city street or drive centimeters from a crash barrier. But even if we decide to do better next time, we tend to forget the lesson as time passes, until a new near miss occurs, or the event is not a miss but an accident. If we survive, we may have learned an even more costly lesson, if we do not, there is nothing more to learn for ourselves, but the accident will be noticed and may serve as a lesson for others.

More problematic are the technical near misses, in which a significant number of the layers of protection are found breached, or are actually improvised temporary fixes, rather than like for like as designed. These can be repaired and whether or not this leads to the conclusion that it was a near miss and whether these are reported is at the discretion of the operator.

The reporting of near misses is problematic per se as will be discussed in the next section.

3.3 | How to promote near-miss reporting

If a fault, an error, or a small violation of a rule leads to a near miss, the reporting is at least embarrassing. It takes a lot of convincing and possibly a lot of time before a near miss and error reporting is accepted, such as in the air transport industry (Gnoni et al., 2022). Nevertheless, near-miss reporting is deemed an important tool in the quest for a safer world (Haas et al., 2020).

To encourage near-miss reporting, it is advised that organizations have a "just" culture as opposed to a "blame" culture (Dekker, 2009, 2012; Dekker & Breaky, 2016). In the latter, errors and violations are punished. In a just culture, people will not be punished for their errors and mistakes. They are then supposed to be able to freely report near misses. Here, the reporting of a near miss is rewarded (Boysen, 2013; Edwards, 2018). As always, the positive goal of creating a just culture and encouraging the reporting of errors and near misses may also have unintended consequences. Events that are not really a near miss may be reported as such, or near misses are "invented" for the reward, making any assessment of their frequency worthless as Goodhart's Law predicts (HTTP4). Creating a just culture takes time and money (Groeneweg et al., 2018). A no-blame culture can also have some unwanted side effects. In a just culture, no blame is or can be attached to specific people, and therefore there is no reason to investigate the role of the humans involved. While this usually is a good idea for the frontline workers involved, it sometimes is not such a good idea for the managers who made the decisions. It is easier and will not invoke any protest, if blame is put on organizations, equipment failure, paperwork, the act of a deity, or the unavoidable black swan (Sherratt et al., 2023).

Sometimes a just culture can also lead to the notion that rules, standards, regulations, or standard operating procedures can be violated without any consequence. This is the extreme of the practice of making a deal in the US justice system. Although Dekker (2009) makes the point that "The

issue is not to exonerate individual practitioners but rather what kind of accountability promotes justice and safety," practice does not necessarily follow good intentions (Van Bijsterveld, 2023). Make a deal, confess your sin, or make excuses and your punishment will be reduced or you will be forgiven entirely (de Bruijn, 2007). This effect may invoke the continued practice of violations because it is convenient at the moment, without any regard to the potential of adverse outcomes in the future (Williams, 2018).

Despite these problems, near misses can be powerful signals of potential problems, and the reporting of near misses therefore is indispensable and having a just culture can promote it.

3.4 | A near miss is sometimes seen as a success

A near miss is a miss and therefore not an incident or an accident. This often leads to the conclusion that the system is resilient enough to accommodate the faults, errors, and mistakes that caused the near miss, and therefore things can stay as they are. In the SAFETY II approach, a near miss is a success because it does not result in a failure; and in case of success, nothing needs to change, although it is hoped that the adaptation needed to avoid the consequences will be noticed and formalized in procedures going forward.

The positive outcome could be regarded as proof that the system is resilient against errors and mistakes and the violations are what is needed to keep the system working. Even if the miss is the result of luck rather than design, there does not need to be a reason for change. If the odds are in favor of a good result and the probability of the action leading to an accident is sufficiently small, not spending resources on preventing the action that led to the near miss may seem to be a good choice.

However, when the near miss is the result of a violation of rules, standards, regulations, or standard operating procedures, one has to recognize that these rules standards, regulations, or standard operating procedures are there to prevent harm to equipment, installations, or the humans involved, in which case, enforcement efforts can do no harm. These can be in the form of reminding people that these rules, standards, regulations, or standard operating procedures are there for a reason and that compliance is generally regarded as essential to prevent harm, even when occasionally a violation does not lead to an immediate disaster. In many cases rules, standards, regulations, or standard operating procedures are part of a multilayered protection system. In such a multilayered system, the breach of one layer of protection need not compromise the system as a whole, but recurrent breaches increase the probability that all layers will be breached, with disaster as a consequence. A multiplicity of "barriers" or defenses is assumed to result in favorable odds and "guarantee" that the defense system will be upheld. Compliance therefore is often the last line of defense against uncontrolled changes in the operation of a system. Should the occasion occur that following the rules consistently is more harmful than not, the rules should be changed. This change should be subject to careful "management of change" considerations (Hale & Borys, 2013a, 2013b).

Proponents of SAFETY II and resilience engineering often argue that compliance actually can cause more harm than good in certain situations.

There is a general principle of law that necessity breaks the law (HTTP13). The scholars on the subject of SAFETY II are vague on whether violations of the laws, regulations, and prescriptions fall within the normal variability of human behavior. In any case, Hollnagel (2014) states that this variability needs to be curtailed if it threatens to get out of control. Violations of rules and regulations should be taken as signs that things are starting to get out of control.

The maintenance handbook for a typical car may be as large as 2300 pages. Although the variability in ways mechanics maintain such a car is considerable, adhering to the manual ensures that the client receives a car that is properly put together again after it has been in the workshop (HTTP14).

The consequences of "variability" in aircraft maintenance can be found for instance in the following example. A mechanic who had to refit the cockpit windows of the plane dropped the bolts. Being resilient he went to the workshop to fetch a new set. With these he refitted the window, so the operation was a local success. Unfortunately, the ones he took were slightly smaller in diameter. So, although the operation of fitting the window was a work-around success, when the aircraft took to the sky, the bolts did not fit enough to hold the window, which blew out. Although one could also say that the plane landed successfully, that nobody was killed, that the "system" was sufficiently resilient, and that the event was a success in SAFETI II terms, therefore not an accident and this went right (HTTP15).

In other articles, we have given several other examples where ignoring rules and regulations seems to be the right thing to do but proves disastrous later, such as letting a single point of failure exist in B737 (Ale et al., 2010) and in the B737max (Ale et al., 2021), or ignoring the rule that a new catalyst needs to be tested before use (Ale et al., 2018).

The scholars of SAFTY II extract their examples mainly from situations where the operator/worker can have sufficient expert knowledge to judge the consequences of their actions. It should be noted however that in complex industrial installations, any change in hardware, software, or modus operandi needs to go through a thorough management of the change process to make sure that such a change does not have unwanted usually long-term negative consequences.

Alternatively, one has to conclude that the rules are superfluous and discontinue them to prevent a contagious practice of violation that can spread to rules that are useful and necessary.

Unfortunately, at least in this context, human beings are intrinsically optimistic (Sharot, 2011). However, SAFETY II scholars do not define what "things go right" means (Leve-

son, 2020). The interpretation in practice is that a near miss when "things go right" after all is a success. Therefore, a near miss can lead people and their organizations to think that the favorable result is as it should be, and will be in the future, and that therefore an error, mistake, deviation from operating procedures, or hardware malfunctions is not so serious after all (Ale et al., 2020). In this context, the notion of hindsight bias is often used as an argument to not do something about an element in the causal chain constructed after the fact, for instance, because the causal chain is assumed to be unique for the fact. Also, the claim of hindsight bias is often incorrect. The claim that one could not know with the knowledge before the fact, that the causal path existed and the logic and inevitability of the fact, given a particular causal element, was only revealed after the fact, has been proven to be false in almost every incident report the authors are aware of. In most cases, the potential consequences of a particular element were known and reported before the fact but dismissed as unlikely, fatalistic thinking, or similar qualifications. Qualifying a near miss as a success contributes to the dismissal of these inconsequential elements and makes these into a weak signal.

In practice, this means that, despite the urgings of SAFETY II and resilience engineering scholars (Hollnagel, 2014), additional measures are deemed disproportional or too expensive (Helsloot, 2023, 2010), which is unfortunate but not uncommon; this could be seen as collateral damage of the perceived meddling and criticism, with undoubtedly good intentions, of these SAFETY-II and resilience engineering scholars.

4 | THE USEFULNESS OF SIGNALS

This raises the question of whether hunting for weak signals does any good in promoting safety. De Bruijn (2007) observed that recommendations given by investigation bodies such as public inquiries in the United Kingdom or the Onderzoeksraad voor Veiligheid in the Netherlands are often ignored. He lists a number of reasons:

1. The investigation was aimed at finding a single and simple explanation. In their search, some other investigative bodies resort to investigating causal relationships and find causes in violations of existing rules and regulations. Other bodies resort to investigating the system and organizations and find complexity and trust issues. The former usually leads to hard conclusions, and the latter usually to softer conclusions often blaming the big bad world and legitimizing the process as is.
2. The investigation concentrates on what went wrong, ignoring that similar actions and decisions were made in the past and even in hindsight seem right.
3. Recommendations to change something are made without consideration of the potential negative effects and collateral damage of the change itself.

It has been mentioned above that a hard conclusion often is met with objections such as that the world is more complicated than what the investigators present and that one should look at the wider system. It has also already been mentioned that a soft conclusion lets everybody off the hook and usually nothing changes.

The problem with similar actions that even in hindsight seem right is that although it seems right in hindsight today, that hindsight may change after the next accident. Therefore, although the invocation of the hindsight excuse may help to maintain the status quo, changing it may be better in the long term, recognizing that foresight or precaution may ensure that there will remain an opportunity for hindsight in the future.

Although the costs of a disaster in many cases exceed the amount it would have cost to avoid them (Kletz, 1988), they are almost always underestimated before the fact. Therefore, although the disaster is predictable and there are clear signs that it is coming, the warnings are ignored in favor of continuing business as usual (Wucker, 2016).

On the one hand, investigators point to the importance of a just culture and near-miss reporting (Helsloot, 2023), while on the other hand, showing a composite picture of the effect of a boiling liquid expanding vapor explosion on a city center to an audience of decision-makers about to authorize large-scale transport of liquid petroleum gas by train through a central station has been dismissed as scaremongering by single-minded safety experts (Helsloot & Schmidt, 2012).

The major contributor to the spread of the fire in the Grenfell Tower, which killed 72 people, was the flammability of the outside cladding (HTTP6). These fires had occurred earlier, such as in Summerland on the Isle of Man in 1973 (50 casualties), Knowsley Heights in 1991, Lakanal House in Southwark in 2009, and the fire in The Marina Torch in Dubai in 2015 in the United Arab Emirates (HTTP5).

The main contributor to the fire in the prison at Schiphol Airport (2005, 11 dead; OvV, 2006) was the omission of prescribed barriers against the spread of an initial blaze.

The main contributor to the crash of Turkish Airlines 1951 (2009, nine dead; Ale et al., 2010) and to the crashes of the B737-max in 2018 and 2019 (346 dead) is their vulnerability to a single point of failure and ignoring previous—inconsequential—failures (Ale et al., 2021).

These accidents all have different circumstances and different causal chains: And each causal chain was identified in hindsight. What they have in common is that in all these cases, basic engineering principles were violated; and it was known at the highest level of management that these principles had been violated, principles that have long been established as a generic barrier to prevent the accidents before they escalated. These are examples of the pre-sight bias mentioned before, that is, the potential for an accident is dismissed before the fact, or the risk is deemed acceptable before the accident, but not after. There was also a common cause—short-term profit was prioritized over safety.

Although the focus on near misses and weak signals is established in the airline industry, even there, the larger

signals mentioned above were dismissed. This raises the question as to whether attention to weak signals and sophisticated system modeling will really help to reduce the probability of accidents and promote safety. From the modeling point of view, it may be beneficial to extend the modeling to include the stakeholders, decision-makers, and the, primarily money-driven, environment in which the decision-making system operates. From the safety professionals' side, it should not be assumed that what seems good for safety is not automatically considered good for a company or even for society. In the effort to get to grips with the complexity of sociotechnical systems, sight again may be lost of the context in which the results of these efforts are communicated as warnings of potential harm or recommendations for improvements (Otway & Wynne, 1989).

Woods (2005) recommends changing organizations and giving technical experts a more explicit and more influential place in the decision-making process. From the above examples, it seem to appear that the world has moved in another direction, making the influence of technical expertise in many decision-making processes even smaller (Ale, 2022).

Hindsight is not as biased as sometimes suggested—although hindsight bias exists. However, the signals indicating future problems are often too weak to be taken seriously or are outweighed by other arguments. Declaring the techniques of "yesteryear," such as FTA and QRA (Qualtitative Risk Analysis), obsolete does not help to change the direction of development. SAFETY II is said to be redirecting the focus on to why things go right, whatever the definition of right may be (Ale et al., 2021), which leads to the demand for models that are closer to simulation models. However, real systems normally do not fail. If a model simulates a real system, it will normally not show a path to failure either. Therefore, the safer the system, it becomes increasingly unlikely that rare signals of potential upsets will be detectable, even in multiple instantiations of the model.

5 | DISCUSSION

The current emphasis in most approaches to coping with complexity in today's systems tends to be focused on discerning and preventing system failures. Patching up systems in this way is attractive as it involves finding and fixing relatively rare occurrences. This is often much simpler than having to really understand exactly how the whole system "normally" operates successfully, most of the time. The paradigm shift thus required is to heed Ackoff's warnings (Ackoff. 1988) that system analysis (of separate components) can give you knowledge but not necessarily the understanding needed and is necessary to change the perspective from defensive, reactive, to proactive adaptive (Hollnagel SAFETY II).

Our thesis is that this new mindset requires us to recognize that these attributes of resilience, interaction, and interdependence of components and functions and the unique adaptive abilities of humans, as well as allowing for their fallibilities, are vital and largely unrecognized, preconditions to coping with systemic complexity. This is a fundamental requirement currently, which will only become more urgent as we progress into the next industrial age of AI.

There is still the problem of incorporating all these factors. If "models" succeed in describing real systems more accurately, they tend to become as intractable as the real world. Since these models were meant to help us better understand and predict the behavior of the real systems, their usefulness diminishes even as they are improving and becoming as intractable as the real world.

In the Cynefin classification of system behaviors (Snowden & Boone, 2007), simple and even complicated systems are amenable to the traditional methods of system analysis and modeling. Here, relationships between entities in the system are clear and unambiguous and they behave predictably. So, we can build mathematical "models" to describe how the system is expected to function in different situations.

Bar-Yam (2004) defined "complex systems" as systems that "have multiple interacting components, whose collective behaviour cannot be simply inferred from the behaviour of components."

So, the traditional system modeling approach of decomposition into components and attempting to build up a picture of system performance from individual pieces is no longer appropriate for this class. The problem is that most of the methods we currently employ, and particularly the methods that attempt to make the case that these systems are safe, rely on this decomposition into components and attempt to make the parts more reliable individually. Even the more systemic approaches encourage adding layers of protection, barriers, strengthened control loops, and better safety-critical systems, in the hope that they are making the systems safer. But without a valid model to test the effectiveness of these add-ons, we cannot be sure that they will not have an opposite effect to make the systems even more complex, unreliable, and unpredictable.

This intractability of a model can itself lead to a disaster. In the Netherlands, people can obtain an allowance for childcare, which is paid out by the tax office. In their efforts to detect fraud, the tax office uses statistical methods to detect whether certain properties of people indicate an elevated probability of fraud. To enhance this method, they introduced AI and self-learning capabilities. One of the features was that the program selected the indicators for fraud itself from a large set of properties of the people who applied for an allowance. These properties were extracted from the database of taxpayers and the general database of the population. This program subsequently "learned" that the highest probability of fraud was related to low-income, single parents of color. On this basis, the payments of approximately 30,000 people were revoked plunging these families into poverty and social misery, completely unjustifiably. (HTTP9). In his testimony before the parliamentary enquiry committee, the responsible consultant testified that he was aware of the unpredictability of this behavior and that he therefore had advised to continuously monitor the behavior of the program to catch unwanted

effects in time, that is, before actions indicated by the results of the program were taken toward citizens. Unfortunately, this advice was ignored. In practice the tax office would have been better off with a simple statistical analysis in which the variables to be analyzed were completely under their control (HTTP10, HTTP11).

But as far as models are concerned, the ultimate example of a model that is at the same time mathematical explicit and intractable is the uncertainty principle by Heisenberg. The models used by safety practitioners are usually much simpler, but the purpose of models like FRAM is to capture the emergent behavior of a real system. Emergent behavior may be retractable after the fact, but is intractable before the fact, by definition.

The recent incident in the Air Traffic Control System for UK airspace illustrates this point well. A single wrongly filed flight plan caused disruption to flights across the world and lasted for days (HTTP12). The organization responsible, the National Air Traffic Service, is undoubtedly a highly experienced, responsible, and very professional and effective body. So, they would have done their safety studies diligently. They would have applied the procedures laid down in their codes and standards. The software equivalents of FMEAs and fault trees used in the process industries would have been done by the book and almost certainly flagged non-conforming data as an issue. But the safety-critical system they put in place, to deal with this "failure," seems to have allowed the "total system" (not just the control towers) to fail.

So, such basic safety disciplines as fault trees and FMEA's, while fundamentally important to identify and ensure specific identified failure cases are prevented and or mitigated, can be misleading. Just adding more and better safety-critical systems, redundant components, and "barriers" can lead to the opposite effect if the resulting complex interactions and interdependencies are not addressed. This can lead to the kind of unexpected and emergent behaviors that they necessarily would not have seen before.

To deal with this, we need an understanding of how the whole system actually works and responds to perturbations and less-than-ideal situations, and hence a change in mindset is needed, not just in design but in the assurance and safety approaches applied to comply with performance standards. Perhaps something more aggressive is appropriate in the nature of the Security Chaos Engineering approach suggested by Rosenthal and Jones (2020).

6 | CONCLUSION

The scientific world tries to help decision-makers to cope with increasingly complex and intractable sociotechnical systems by building ever more complex models that increasingly themselves become intractable. This does not necessarily help the decision-maker or make the world any safer. It is crucially important to make the decision-makers aware of the potential consequences of this complexity.

Extensive progress has been made in developing instruments to get to grips with the complexity of sociotechnical systems. Exploiting AI in doing so arguably enhances the capabilities of these systems. They can learn to detect the variations in operation under the influence of time and pressures such as efficiency and efficacy. If these systems could be framed with a definition of what constitutes the safe envelope (Hale et al., 2007), these could generate explicit warnings that the operation has evolved into uncharted and potentially dangerous waters (Woods, 2005).

However, these systems will have little effect if the message does not get through to the decision-makers or is dismissed. Although "follow the money" seems an easy answer, valuing in monetary terms, what are also called the imponderables and agreeing on these valuations, proves to be difficult. Therefore, even economic evaluations are contextual.

In conclusion, we repeat a quote we used in previous papers: "And in this sense, we, who claim to work in the area, should accept that we are dealing with a topic that, in itself, is no simpler, nor more complex than that of any other aspect of how people experience and model their worlds and then act on these representations. But it is a topic that is, nevertheless, essentially a political matter" (Otway & Thomas, 2006).

ACKNOWLEDGMENTS

We thank the unnamed reviewers for their comments, which improved our paper considerably.

ORCID

Ben J. M. Ale https://orcid.org/0000-0002-9634-3002

REFERENCES

Ackoff, R. L. (1989). From data to wisdom, presidential address to ISGSR, June 1988. *Journal of Applied Systems Analysis*, *16*, 3–9.

Ale, B. J. M. (2003). *Dat Overkomt Ons Niet. Oratie ter gelegenheid van ambstaanvaarding*. TU-Delft.

Ale, B. J. M. (2006). *The occupational risk model*. TU-Delft. ISBN 9056381571.

Ale, B. J. M. (2016). Risk analysis and big data. *Safety and Reliability*, *36*(3), 153–165. https://doi.org/10.1080/09617353.2016.1252080

Ale, B. J. M. (2022). *Third-party risk policies in the Netherlands: A historical sketch*. Cambridge Scholars Publishing.

Ale, B. J. M., Bellamy, L. J., Cooper, J., Ababei, D., Kurowicka, D., Morales, O., & Spouge, J. (2010). Analysis of the crash of TK 1951 using CATS. *Reliability Engineering & System Safety*, *95*(5), 469–477.

Ale, B. J. M., Bellamy, L. J., Oh, J. I. H., Whiston, J. Y., Mud, M. L., Baksteen, H., Papazoglou, I. A., Hale, A., Bloemhoff, A., & Post, J. (2006). Quantifying occupational risk. *3rd International Conference on Working on Safety*, Zeewolde, The Netherlands.

Ale, B. J. M., Bellamy, L. J., van der Boom, R., Cooper, J., Cooke, R. M., Goossens, L. H. J., Hale, A. R., Kurowicka, D., Morales, O., Roelen, A. L. C., & Spouge, J. (2009). Further development of a causal model for Air Transport Safety (CATS): Building the mathematical heart. *Reliability Engineering & System Safety*, *94*(9), 1433–1441.

Ale, B. J. M., Hartford, D. N. D., & Slater, D. H. (2019a). Variability: Threat or asset? *IChemE Symposium Series No 166, HAZARD 29*, Birmingham, UK. https://www.icheme.org/media/19419/hazards-29-paper-39.pdf

Ale, B. J. M., Hartford, D. N. D., & Slater, D. H. (2019b). Variability, asset or curse. *ESREL 2019 29th European Safety and Reliability Conference*, Hannover, Germany.

Ale, B. J. M., Hartford, D. N. D., & Slater, D. H. (2020). Resilience or faith. *E-proceedings of the 30th European Safety and Reliability Conference and 15th Probabilistic Safety Assessment and Management Conference (ESREL2020 PSAM15)*, Venice, Italy.

Ale, B. J. M., Hartford, D. N. D., & Slater, D. H. (2021). Prevention, precaution and resilience: Are they worth the cost? *Safety Science*, *140*, 105271, https://doi.org/10.1016/j.ssci.2021.105271

Ale, B. J. M., Kluin, M. H. A., & Koopmans, I. M. (2018). Safety in the Dutch chemical industry 40 years after Seveso. *Journal of Loss Prevention in the Process Industries*, *49*, 61–67.

Bar-Yam, Y. (2004). *Making things work: Solving complex problems in a complex world*. Knowledge Press.

Baumgartner, M., & Malakis, S. (2023). Just culture and artificial intelligence: Do we need to expand the just culture playbook? *Eurocontrol, Hindsight*, *35*, https://skybrary.aero/sites/default/files/bookshelf/34372.pdf

Boysen, P. G., II (2013). Just culture: A foundation for balanced accountability and patient safety. *The Ochser Journal*, *13*(3), 400–406. PMID: 24052772; PMCID: PMC3776518.

Bray, C. W. (1962). Toward a technology of human behavior for defense use. *American Psychologist*, *11*(7), 527–541.

Cooke, R. I. (2000). *How complex systems fail*. Cognitive Technologies Laboratory, University of Chicago. https://how.complexsystems.fail/

De Bruijn, H. (2007). *Een gemakkelijke waarheid, Waarom we niet leren van onderzoekscommissies*. NSOB. https://www.nsob.nl/sites/www.nsob.nl/files/2019-10/NSOB-2007-Een-gemakkelijke-waarheid.pdf

Dekker, S. (2012). *Just culture balancing safety and accountability*. CRC Press. https://doi.org/10.4324/9781315251271

Dekker, S., & Breakey, H. (2016). 'Just culture:' Improving safety by achieving substantive, procedural and restorative justice. *Safety Science*, *85*, 187–193. https://doi.org/10.1016/j.ssci.2016.01.018

Dekker, S. W. A. (2009). Just culture: Who gets to draw the line? *Cognition Technlogy & Work*, *11*, 177–185. https://doi.org/10.1007/s10111-008-0110-7

Edwards, M. T. (2018). An Assessment of the impact of just culture on quality and safety in US HOSPITALS. *American Journal of Medical Quality*, *33*(5), 502–508. https://doi.org/10.1177/1062860618768057

Ericson, C. (1999). Fault tree analysis—A history. *Proceedings of the 17th International Systems Safety Conference*, Orlando, FL.

Franchina, F. (2023). Artifical intelligence and the just culture principle. *Eurocontrol, Hindsight*, *35*, https://skybrary.aero/sites/default/files/bookshelf/34372.pdf

Gnoni, M. G., Tornese, F., Guglielmi, A., Pellicci, M., Campo, G., & De Merich, D. (2022). Near miss management systems in the industrial sector: A literature review. *Safety Science*, *150*, 105704.

Groeneweg, J., Ter Mors, E., van Leeuwen, E., & Komen, S. (2018). The long and winding road to a just culture. *Paper presented at the SPE International Conference on Health, Safety, Security, Environment, and Social Responsibility*, Abu Dhabi, UAE.

Guillaume, E. G. (2011). *Identifying and responding to weak signals to improve learning from experiences in high-risk industry* [Doctoral dissertation, TU Delft]. https://repository.tudelft.nl/islandora/object/uuid:f455e8a0-cce5-4a36-8a98-f83371dc2a2a/datastream/OBJ/download

Haas, E. J., Demich, B., & McGuire, J. (2020). Learning from workers' near-miss reports to improve organizational management. *Mining, Metallurgy & Exploration*, *37*(3), 873–885. https://doi.org/10.1007/s42461-020-00206-9

Hale, A. R., Ale, B. J. M., Goossens, L. H. J., Heijer, T., Bellamy, L. J., Mud, M. L., Roelen, A., Baksteen, H., Post, J., Papazoglou, I. A., Bloemhoff, A., & Oh, J. I. H. (2007). Modelling accidents for prioritizing prevention. *Reliability Engineering and System Safety*, *92*(12), 1701–1715.

Hale, A. R., & Borys, D. (2013a). Working to rule or working safely? Part 1: A state of the art review. *Safety Science*, *55*, 207–221.

Hale, A. R., & Borys, D. (2013b). Working to rule or working safely? Part 2: The management of safety rules and procedures. *Safety Science*, *55*, 222–231.

Hanea, D., Hanea, A., Ale, B. J. M., Sillem, S., Lin, P. H., van Gulijk, C., & Hudson, P. (2012). Using dynamic Bayesian networks to implement feedback in a management risk model for the oil industry. *PSAM*, *11*, *ESREL 2012*, Helsinki, Finland.

Helsloot, I. (2023). Houd veiligheid uit de concurrentiestrijd. *HSEQ—BILFINGER & SAFETY BILFINGER magazine 2*.

Helsloot, I., Pieterman, R., & Hanekamp, J. (2010). *Risico's en redelijkheid*. Boom Juridische Uitgevers Den Haag.

Helsloot, I., & Schmidt, A. (2012). The intractable citizen and the single-minded risk expert. *European Journal of Risk Regulation*, *2012*(3), 305–312.

Hollnagel, E. (2009). The ETTO principle: Efficiency-thoroughness trade-off—Why things that go right sometimes go wrong. *Risk Analysis*, 30, 153–159. Routledge, ISBN 0754676781.

Hollnagel, E. (2011). *Introduction to FRAM: The functional resonance analysis method*. SDU. https://www.functionalresonance.com/FRAM-2_introduction_to_FRAM.pdf

Hollnagel, E. (2012). *FRAM: The functional resonance analysis method: Modelling complex socio-technical systems*. CRC Publications.

Hollnagel, E. (2014). *Safety—I and Safety—II- The past and Future of Safety Management*. Ashgate Publishers.

Hoorens, V. (2014). Positivity bias. In A. C. Michalos (Ed.). *Encyclopedia of quality of life and well-being research* (pp. 4938–4941). Springer.

HTTP1. (2023, July 23). Apollo 13. In *Wikepedia*. https://en.wikipedia.org/wiki/Apollo_13

HTTP10. (2023). *Parlementaire enquêtecommissie Fraudebeleid en Dienstverlening – openbaar verhoor de heer Koemans*. Tweede Kamer. https://debatgemist.tweedekamer.nl/debatten/parlementaire-enqu%C3%AAtecommissie-fraudebeleid-en-dienstverlening-%E2%80%93-openbaar-verhoor-de-hee-22

HTTP11. https://files.tweedekamer.nl/sites/default/files/2023-10/20231002%20PEFD%20openbaar%20verhoor%20Koemans.pdf

HTTP12. (2023). *NATS report into air traffic control incident details root cause and solution implemented*. NATS. https://www.nats.aero/news/nats-report-into-air-traffic-control-incident-details-root-cause-and-solution-implemented/

HTTP13. (2023, October 23). Necessity in English criminal law. In *Wikipedia*. https://en.wikipedia.org/wiki/Necessity_in_English_criminal_law

HTTP14. *Dacia Duster workshop manual 2009–2017*. https://cargeek.live/docs/Dacia_Duster_Workshop_Manual_2009_2017_aauacfg.pdf

HTTP15. (2023, October 25). British Airways Flight 5390. In *Wikipedia*. https://en.wikipedia.org/wiki/British_Airways_Flight_5390

HTTP16. (2023, October 24). Erik Hollnagel. https://erikhollnagel.com/ideas/etto-principle/

HTTP2. *What is OSHA's definition of a near miss?* OSHA.com. https://www.osha.com/blog/near-miss-definition

HTTP3. (2023, July 29). Buncefield fire. In *Wikipedia*. https://en.wikipedia.org/wiki/Buncefield_fire

HTTP4. (2023, October 25). Goodhart's law. In *Wikipedia*. https://en.wikipedia.org/wiki/Goodhart's_law

HTTP5. (2023, July 30) Grenfell Tower fire. In *Wikipedia*. https://en.wikipedia.org/wiki/Grenfell_Tower_fire

HTTP7. *Easi Dec Solar Bridging Ladder—For installing and maintaining solar panels*. Simplified Safety. https://simplifiedsafety.co.uk/brands/easi-dec/easi-dec-solar-ladder

HTTP8. (2023, January 24). Inspection: Many accidents during installation of solar panels, 4 deaths. Blikopnieuws. https://www.blikopnieuws.nl/nieuws/297191/inspectie-veel-ongelukken-bij-installatie-zonnepanelen-4-doden.html

HTTP9. (2023). *Parlementaire ondervraging kinderopvangtoeslag*. https://www.tweedekamer.nl/sites/default/files/atoms/files/20201217_eindverslag_parlementaire_ondervragingscommissie_kinderopvangtoeslag.pdf

Kletz, T. (1988) *Learning from accidents in industry*. Butterworth. ISBN 10: 0408026960 ISBN 13: 9780408026963.

Leveson, N. (2020). *Safety III: A systems approach to safety and resilience*. MIT. http://sunnyday.mit.edu/safety-3.pdf

Leveson, N. (2023). *A paradigm change for safety and for system engineering*. Presentation to the Object Management Group MBSE workshop https://omgwiki.org/MBSE/lib/exe/fetch.php?media=mbse:incose_mbse_iw_2023:0.0.2023-01-28.iw2023_mbse_workshop_plenary_leveson_incose_keynote_talk.pdf

Otway, H., & Thomas, K. (2006). Reflections on risk: Perception and policy. *Risk Analysis*, *2*(2), 69–82.

Otway, H., & Wynne, B. (1989). Risk communication: Paradigm and paradox. *Risk Analysis*, *9*(2), 141–145.

Oviedo-Trespalacios, O., Peden, A. E., Cole-Hunterm, T., Costantini, A., Haghani, M., Rud, J. E., Torkamaan, H., Yariq, A., Newton, J. D. A., Gallagher, T., Steinert, S., Filtness, A., & Reniers, G. (2023). The risks of using ChatGPT to obtain common safety-related information. *Safety Science*, *167*, 106244.

OvV. (2006). *Brand cellencomplex Schiphol-Oost*. OvV. https://www.onderzoeksraad.nl/nl/page/395/brand-cellencomplex-schiphol-oost

Patriarca, R., Leonhardt, J., & Licu, A. (2022*). Unearthing weak signals for safer and more efficient socio-technical systems, the SECA method*. Eurocontrol. https://skybrary.aero/sites/default/files/bookshelf/32715.pdf

Perrow, C. (1999). *Normal accidents: Living with High risk technologies*. Princeton, NJ: Princeton University Press.

Rashidy, R. E. L., Hughres, P. H., Figueres-Esteban, M. F., Harrison, C., & Van Gulijk, C. (2018). A big data modeling approach with graph databases for SPAD risk. *Safety Science*, *110*, 75–79.

Rasmussen, J. (1997). Risk management in a dynamic society: A modelling problem. *Safety Science*, *27*(2-3), 183–213. https://doi.org/10.1016/S0925-7535(97)00052-0

Rosental, C., & Jones, N. (2020). *Chaos engineering; System resiliency in practice*. O'Reilly.

Shapero, A. (1960). *Human engineering testing and malfunction data collection in weapon system test programs* (WADD Technical Report 36). Wright Air Development Division, Wright-Patterson Air Force Base, Ohio, Feb. 1960.

Sharot, T. (2011). *The optimism bias: A tour of the irrationally positive brain*. Pantheon Books.

Sherratt, F., Thallapureddy, S., Bhandari, S., Hansen, H., Harch, D., & Hallowell, M. R. (2023). The unintended consequences of no blame ideology for incident investigation in the US construction industry. *Safety Science*, *166*, 106247.

Singh, P., & van Gulijk, C. (2023). Digital safety delivery: How a safety management system looks different from a data perspective. In C. van Gulijk, E. Zaitseva, & M. Vassay (Eds.), *Reliability engineering and computational intelligence for complex systems. Studies in systems, decision and control*(Vol. 496, pp. 145–157). Springer. https://doi.org/10.1007/978-3-031-40997-4_10

Singh, P., van Gulijk, C., & Sunderland, N. (2023). Online process safety performance indicators using big data: How a PSPI looks different from a data perspective. *Safety*, *9*(3), 62. https://doi.org/10.3390/safety9030062

Slater, D. (2023). Complex system modelling 1.1. *Complex system modelling conference: Safer complex working group*, Safety Critical Systems Club. https://doi.org/10.13140/RG.2.2.27522.04806

Slater, D. H., & Ale, B. J. M. (2022). Organisations: Drifting or dysfunctional? In M. Chiara Leva, E. Patelli, L. Podofillini, & S. Wilson (Eds.), *Proceedings of the 32nd European safety and reliability conference (ESREL 2022)* (pp. 3173–3180). Research Publishing. https://doi.org/10.3850/978-981-18-5183-4_S32-02-197-cd

Snowden, D. J., & Boone, M. E. (2007). A leader's framework for decision making. *Harvard Business Review*, *85*(11), 68–76.

Swain, A. D., Altman, J. W., & Rook, L. W. (1963). *Human error quantification*. SANDIA.

Swain, A. D., & Guttmann, H. E. (1983). *Handbook of human-reliability analysis with emphasis on nuclear power plant applications. Final report* (Report No: NUREG/CR-1278; SAND-80-0200 ON: DE84001077). NRC, Web. https://doi.org/10.2172/5752058

Van Bijsterveld, K., & Verhaegh, A. (2023). Reconciling criminal law enforcement with just cuture. *Eurocontrol, Hindsight*, *35*, https://skybrary.aero/sites/default/files/bookshelf/34372.pdf

Van Gulijk, C., Vroom, M. B., Binnekade, J. M., Tepaske, R. J., Dongelmans, D. A., Kurk, M. J., Gans, M., Schipper, C. V., Koornneef, F., & Ale, B. J. M. (2009). *Management van Patientveiligheid*. TU-Delft.

Vaughan, D. (1996). *The challenger launch decision: Risky technology, culture, and deviance at NASA*. University Of Chicago Press.

Watson, A. (1961). *Launch control safety study, Section VII, VOL1*. Murray Hill, NJ: Bell Labs.

Willams, N. (2018). *Gross negligence manslaughter in healthcare*. https://assets.publishing.service.gov.uk/media/5b2a3634ed915d2cc8317662/Williams_Report.pdf

Woods, D. (2005). Creating foresight: Lessons for enhancing resilience from Columbia. Lessons from the Columbia accident. In W. H. Starbuck & M. Farjoun (Eds.), *Organizations at the limit, lessons from the Columbia disaster* (pp. 1–6). Blackwell.

Wucker, M. (2016). *The gray rhino: How to recognize and act on the obvious dangers we ignore*. St. Martin's Press.

How to cite this article: Ale, B. J. M., & Slater, D. H. (2023). Complexity for complexity—How advanced modeling may limit its applicability for decision-makers. *Risk Analysis*, 1–11. https://doi.org/10.1111/risa.14261