Variability: Threat or Asset?

Ben J.M. Ale, Technical University Delft, PO Box 5015, 2600 GA Delft
Des N.D. Hartford BC Hydro, 6911 Southpoint Drive, Burnaby, BC, V3N 4X8, Canada
David H. Slater, Cardiff University, School of Engineering, Queen's Buildings, 14-17 The Parade, Cardiff CF24 3AA

In the philosophy of SAFETY-I variability is seen as a threat, because it brings with it the possibility of an unwanted outcome. Variability of hardware is curtailed by precise specifications, controlled manufacturing and installing. Variability of human behaviour is curtailed by training and selection of personnel and by regulations, prescriptions and protocols. In the philosophy of SAFETY-II variability is seen as an asset. In SAFETY-II, humans are seen as able to cope with the variability and imperfections of technology and the variability of circumstances to keep systems working.

In SAFETY-II this capacity of coping has been often designated as resilience. Recently the meaning of resilience has been further stretched to include the ability of restoring the operational state after an excursion into the realm of inoperability, or failure. Artificial intelligence allows systems to evolve by processing information acquired by sensing the result of their actions and variable environment in which they operate. This makes such systems intrinsically more variable than deterministic systems and therefore less predictable. For operators of these systems it is essential that they understand and are able to deal with this variability in order to keep systems operational and adaptive on the one hand and prevent excursions into unwanted territory on the other. The SAFETY-II philosophy seems to be more suitable to such an environment. At the same time it increases uncertainty about potential future states. The belief that humans will cope if an unexpected situation may arise, will reduce the emphasis on defensive, prevention measures that can limit the probability that the system may behave in an unwanted, unsafe manner.

The stretched meaning of resilience exacerbates this problem, because there is no real limit of what systems or society using these systems may bounce back from. A highway bridge that collapses can be re-built. Thus society is resilient against bridge collapses. The question is however, should society accept a situation in which there is a significant probability that such a bridge collapses as safe or safe enough.

The philosophies behind SAFETY-II and resilience engineering promote safety by exploiting self-correcting mechanisms in technology and the ingenuity of humans to keep systems within the desired operating envelope. In this approach, a form of trial, error and correct, the prior occurrence of the error, or deviation is essential. Unfortunately the error may also be fatal or catastrophic: maybe not for society as a whole, but surely for an individual, a group of individuals or a company. With an increasing tendency to evaluate every decision in terms of – preferably monetarized – costs and benefits, striking a balance between a SAFETY-I, a SAFETY-II and a resilience approach is not made easier by the inherent vagueness of the definition of success and the essentially qualitative nature of the latter two concepts.

In this paper we explore how Safety I, Safety II and resilience can be cast in a way that one levers off the strengths of each one to compensate for the weaknesses of the other.

Introduction

There was a time when nuts and bolts were forged individually by a blacksmith. Every nut fitted a particular bolt. This was completely satisfactory when each connection was unique and repairs were done by the same blacksmith. A good blacksmith could make a series of nuts and bolts that were close to the same, but since there were no standards, there was little chance that products were interchangeable. As late as in the 19th century Joseph Whitworth thought of standardizing the threads so that in the future products made by different manufacturers would fit each other. By curtailing variability productivity was improved.

Variability has been the driving force behind the evolution of nature and mankind. Successful variations of species lead to improved survival characteristics. Unsuccessful variations become extinct. Variations on the other hand also lead to unpredictability and uncertainty. The weather of today is not the same as the weather of tomorrow. To promote chances of survival, predictability is a valuable tool. To know in advance when winter will end, when seeds can be sown and when crops can be harvested, has led to the development of astronomy and many of the other sciences. Uncertainty can be further reduced by following in other people's footsteps, along paths that have been proven to be safe; first literally, leading to worn out footpaths, carriage tracks and roads, later figuratively, by following examples. These examples were often coded into practices, notes, drawings. To prevent houses collapsing, for example, such codes were then converted into rules and regulations.

These developments were all aimed at enhancing predictability and reducing variability. Timbers needed to be of certain dimensions and the wood of certain qualities.

In military applications, multiple layers of defence were designed to cope with expected and unexpected breaches of any one of them. Systems of outer walls, moats with drawbridges, inner walls and keeps formed the "defence in depth" of many fortresses. The idea of building such elaborate defences sometimes often survives beyond their useful life. The walls of the fortress in Sedan, France are 30 m thick and have never have been breached. The walls of the powder storage however, could not resist the experiments of the powder master. The arrival of the airplane in the 20th century, made all these walls useless although one would expect that it would have been a real surprise for the designers of the fortress to see that they had not been made obsolete, but have survived for over 400 years.

Over the centuries the emphasis has been swinging back and forth between curtailing variability and defence against external threats. But in their strategic thinking, there always has been a combination of the two. A fortress had walls, artillery and a guard inside to take care of intruders; and a fire brigade to deal with stuff that was thrown over the walls.

SAFETY-I

Safety always has been one of the primary concerns of mankind. It started to be a recognizable separate science during the industrial revolution of the 19th century (Swuste et al). The dominant motivation in society was progressive, encouraging technological development and the discovery of the new, as a reaction to the conservative reverence of the existing in the previous centuries. The development of machinery made it possible to harness energy on a larger scale and make it much more widely available than was possible earlier. This in turn led to an increase of number of situations where the force of the machinery exceeded the resistance of the human body. The number of injuries and deaths thus grew to worrying proportions. These concerns were enhanced because these deaths were not distributed evenly over society. There had always had been injuries and deaths during work on farms, in construction and accidents caused by trips and slips; but now the deaths were more localized, as in a single factory, with an identifiable owner. It is therefore no surprise that the owners put the blame on the workers themselves. They were held to be careless, or even accident prone. This however did not last. The owners of the factories themselves were held responsible. However, this was not deemed enough. A state commission set up by the Dutch government in 1886 concluded in its report following a parliamentary inquiry into the conditions in factories and workshops, "that legal provisions in the interest of the safety and health of the workers could not be missed; indeed, the entrepreneurs often did not pay sufficient attention to this when designing their companies ". In the 1840's the UK passed the first of its "Factories Acts", empowering Factory Inspectors to regulate industrial "Health and Safety", with criminal sanctions. But it wasn't until the 1970's in the aftermath of the Aberfan disaster, that the Robens Report [1] opened the way to requiring that managers owed a statutory duty of care to their employees.

In this context, it is not surprising that there was a demand for the design of protective safety measures, which became another branch of the engineering profession; but where safety was not necessarily an integral part of the design. More often than not, safety measures were just add-ons to an existing technology. This also is understandable. It was already difficult enough to make a machine such as a loom, or a press, work efficiently. Consideration that people who use these machines could be injured and killed often came later; and this led to the design and addition of guards, barriers, kill-switches and other devices that should prevent personnel being killed, or injured. The philosophies behind these designed defences, however, were not new at all. If you need to make sure a structure stays upright when a support fails, use multiple supports. If it is necessary to be able to leave a building if one of the exits is blocked, make multiple exits. If a system needs to be safe if one of the safety systems fail, introduce redundancy and defence in depth.

These developments made systems safer, but also more complex. Assuring safety, slowly became detached from assuring functionality. Nevertheless the development of tools for the analysis of failure was originally meant to "assure" operators that systems would work. Fault-tree analysis for example, was introduced to make sure that the Minuteman missile would actually arrive at its target, which would happen if nothing went wrong. In the decades that followed, various other techniques were introduced such as Hazard and Operability studies, Failure Modes and Effects Analysis (FMEA) and finally Quantitative Risk Analysis. In these developments increasingly, the definition of success was that there were to be no failures.

This was largely caused by two factors. The first is that there is much more known about failures than about success. To construct a fault-tree or perform a FMEA, no quantified information is necessary. Anecdotal and analytical information is sufficient. In ancient times it was sufficient to know that the artillery of the enemy could penetrate a 10 feet wall. No further analysis was needed to decide to make the walls 15 feet thick. Many of today's methods to improve functionality and safety, work the same way; and are the explicit aim of accident investigations: prevent a repeat of the same accident in the future. Fault-Tree Analysis (FTA) and Failure Modes and Effects Analysis, (FMEA) are aimed at preventing specific "accidents" that may not have happened yet. The recognition that individual failures, which by themselves would not jeopardize a mission, could combine in a situation that could, is especially important for systems, such as missiles, for which repair on the go, or remedial action during a mission, is impossible.

The need to know about accidents, incidents and near misses, in order to take measures to prevent a repeat of an unwanted situation, led to the volume of data available on these events to grow to the extent that statistical analysis could be performed, resulting in probabilities that single, or combined failures could occur. This development was needed after it was recognized that no technology is perfect and, that even after all preventive measures were taken, the potential for a mishap was not guaranteed to be eliminated.

It is certainly needed when measures to eliminate a mishap are deemed too expensive. In that case one has to decide whether or not to continue the activity. Usually, for a change in a technology for which a mishap may involve the loss of human health, or life, a demonstration is required that the updated system is at least as safe as the old configuration. A quantified FTA then helps to support the argument.

Since with the expansion of society and technology, the chance that one experiences a mishap first hand, or can learn everything there is to learn by following one's peers, decreases. The specifications of systems, safety measures and parts need to be codified in rules, regulations, building codes and codes of practice.

There are however two problems with this fault centred and fault eliminating approach.

Accidents happen because systems do not conform to regulations.

Accidents happen despite systems conforming to regulations.

The latter is more problematic than the former, especially when systems conforming to regulations are declared safe to the extent that "nothing bad can happen".

Deviations

There are many reasons for why a system may not behave as expected. Some of these are real surprises, but these are rare indeed. We will return to this subject later. Most deviations are caused by non-compliance with rules, regulations, or codes of practice; in short, by not fully applying the lessons from the past. These deviations are predominantly associated with human behaviour. This is not necessarily human error. It can also be deliberate. Deliberate actions can be for the good or for bad reasons. However for most of the time humans are just humans and their actions are variable.

The problem with this focus on deviations is that most of the data about deviations are derived from analyses of failure. What is largely unknown, is how many deviations do not lead to failures. And if deviations exist without leading to a failure, what is the cause of it. Was it because the deviation was not important, because it still was within the safety margins of the design? Was there another factor that prohibited the deviation to become a failure? ,Or was the deviation spotted in time and rectified?

If indeed a deviation leads to a system failure, or an accident, then such a deviation would not often be repeated. If a deviation does not lead to a system failure, or the probability of the deviation leading to a system failure is low, or the deviation does not lead to a system failure immediately, or in an observable time frame, it is likely that such a deviation will persist and be repeated. When a traffic accident can be traced to a defect in a type of automobile, more often than not tens of thousands of cars of the same type need to be called back, because they have the same defect, but not an accident.

Therefore what really needs to be known, is whether the deviation is more common in the population of systems, or people that have an accident, than in the population that does not have an accident. This is also called the denominator problem. [2, 3]. This information is very difficult to acquire, as was demonstrated in the development of the Occupational Risk Model (ORM) [4, 5, 6] and of the Causal Model for Air Transport Safety (CATS) [7]. In the ORM project extensive surveys among workers were performed to acquire data on compliance to rules and regulation and on the underlying causes of non-compliance. This information was the used to determine what deviations should be addressed as a priority. In the CATS project the Accident/Incident Reporting (ADREP) [8] database was used to determine the total number of certain deviations in the population of commercial aircraft, as it was believed the reporting would be complete.

These investigations led to insights into what deviations were more important than others, and which deviations had been of consequence so far. This should not lead to the conclusion that these deviations should be allowed to persist. If these deviations mainly pertained to secondary and tertiary defences and the primary defence is mostly effective, the probability of these defences to be challenged may be small, but there has been a decision at some point that these defences were nevertheless necessary. Moreover, allowing non-compliance and the persistence of deviations is contagious. Postponing a paint job on a white storage tank allowing a little rust to persist may seem harmless until some years later everybody seems to be used to this tank to be brown.

Although some deviations may have natural causes, most deviations ultimately are the result of human decisions and actions. Letting them persist after they have been noticed, is exclusively the result of a decision by a human. Of all the components in any system the human is the most variable.

Human actions

In the 18th, 19th and 20th centuries, the golden age of engineering relied on the seemingly immutable laws of Newton and his descendants, to specify how systems would work: and humans were trained to operate them and expected to behave in a similarly specified way. The operator was thus an add-on, an extension of the engineering design [9].

Since, in the mind-sets of the designers of these relatively simple systems, humans were a problem, it became logical to ascribe the “cause” of malfunctions of the systems to the most unreliable part, the human factor. It is then a very human response to “blame” the human for the results. People can be characterized as being accident prone, either by their personality, or because of their living conditions [10,11]. It is also very convenient. It eliminates the need for further thought about the intrinsic properties of a system and it avoids the need for protection measures that cost money. In incidents with serious consequences, this has often major implications for the involved parties. Each party thus tries to identify and prosecute the party, or person, to blame, to the relief of all the other parties. At the Flixborough Inquiry, there were no less than six different legal teams, each attempting to prove the other’s people were “at fault”.

On the other hand, there is a contradiction in defining the problem this way. Why would the designer be intelligent and infallible, but the user, or operator, be stupid and make errors. On a more abstract level, humans could be said to be the root cause of most problems, as it is ultimately from human decisions that systems are designed and technology used.

In any case, the variability of human behaviour needed to be curtailed. This was not a new idea. Before the technological revolution you already could not be a carpenter without being a member of the carpenter-guild, and you could not be a member of that guild if you did not pass your master’s test. Drivers need a driver’s-license and there are rules about what to do on the road, such as driving in the right side of the road, which is not necessarily the right side of the road.

Standard operating procedures and rules limited the decision space of an operator to the extent sometimes, that a protocol, or procedure, reads like a computer program. If one could automate the actions of the human operator one would do that. In many industries the human operator is there, because there is no machine yet who could do the task, or a human is cheaper, but should behave as a robot. However human intervention can lead to disasters if these humans do not understand the designed workings of systems, as was abundantly demonstrated by the Three Mile Island incident.

Interpreting human error as operator error, however, ignored what emerged to be a deeper problem, which is the circumstances created by supervisors and managers and the decisions they take. It has become increasingly clear that the root of deviations and non-compliance lies much higher in the organization and often much earlier in time.

What operators and managers have in common is that they are driven by many more forces than safety, and the long term functionality of the system alone. Among these factors are peer to peer recognition, rewards, power and above all, money. [12] Where these drivers lead to, depends on the situation, but people appear to be prepared to take significant future risks to gain of immediate satisfaction. People smoke, drink and ride motorcycles. They exceed speed limits to be "on time". And they ignore rules and regulations if there is a short (time) gain to be had.

Unfortunately, bending the rules is often necessary to keep things moving, as is demonstrated time and again when industrial actions are performed by following the letter of all of the regulations. This proves that human ingenuity in interpreting and bending the rules, regulations and protocols, is often needed to make things work.

The question then is, what is the right balance to allow for human ingenuity, on the one hand, and defending against human fallibility at the other, in order to make things work. Looking at a system and analysing it in terms of making it work and making it work reliably and safely, rather than predominantly looking at avoiding deviations, has been designated as SAFETY-II

Variability in Safety I

The SAFETY-I approach primarily focusses on staying out of the unwanted zone. Safety measures can be characterized by curtailing variability and adding barriers to catch situations in which a parameter gets in the unwanted zone. Reducing overall variability does not take away the need for curtailing and defensive measures.

SAFETY II

Towards the end of the 20th century, people were starting to realise that engineering advances were producing systems that were far from simple. So that as well as humans being fallible, the sheer complexity of these modern systems made it difficult to understand how all the personnel and functions fitted together to make them work. Systems became and are still becoming increasingly intractable, which makes an "à priori" analysis of what might go wrong increasingly impossible. Systems that incorporate artificial intelligence and thus have a mind of their own, complicate things even further. Perrow [13] termed these systems as "stiff" and that in such systems it was quite "normal" to have incidents due to misunderstandings of the required interactions and interdependencies.

Perrow's examples of solutions included military discipline and operational priorities drilled into expert teams to offset this complexity. So again, the human factor was identified as the problem to be addressed and people trained and organisations resourced sufficiently to deal with potential upsets.

The Challenger disaster, however, reminded us tragically, that these failings were perhaps inevitable as a result of pressures in a large organisation such as NASA and the lessons learned from the inquiry were developed as an "engineering" solution for these large organisations. It recommended treating the organisation as a system in which there should be similar and sufficient functions, as in engineered hardware, to exert the controls needed to "regulate" the behaviour of the human components. This System Theoretic Approach (e.g. STAMP), is still popular with some of the large military, regulatory and aerospace organisations.

Having identified the nature and significance of some of the factors causing difficulty in operating large complex systems satisfactorily and reliably, many safety professionals have become convinced that we may need to address these issues from a different perspective, if we want to continue to operate ever more complex systems "safely'. Some of these are triggered by the perceived incomprehensibility of low probability – high consequence events. Some of these again, are triggered by the notion that analysis of causality seems to have no end; and some by the more legalistic discussion on whether a probabilistic progression of a sequence of events should lead to a negation of the certainty of the cause after the fact. Thus the matter of causality is a highly philosophical question [14].

The discussion about the infinity of the chain of causality is an old one and goes back to the Greek atomists some 400 years BC [15]. The why question in this context can have two meanings: "to what purpose" and "with what cause". Both questions can only be answered within a bounded system, because they imply that there is something causing the system to exist. A bounded system can show behaviour that the makers did not anticipate. In most cases the cause of this behaviour can be found as a combination of behaviours of parts of the system that the makers of the system did not consider.

Projective analyses take time and effort, and efficiency demands these analyses to be limited. The fact that a behaviour was not anticipated does not imply that anticipation was impossible, merely that it was deemed impractical. Nevertheless, one could make the proposition that complex systems show emergent behaviour that is not only surprising, but could not be anticipated in principle. This proposition seems equal to proposing that the system is alive: as Chalmers put it:" *A system is alive if and only if it reproduces, adapts with utility 800 or greater, and metabolizes with efficiency 75 percent, or exhibits these in a weighted combination with such-and-such properties," we can simply note that if a system exhibits these phenomena to a sufficient degree then it will be alive, by virtue of the meaning of the term. If an account of relevant low-level facts fixes the facts about a system's reproduction, utility, metabolism, and so on, then it also fixes the facts about whether the system is alive, insofar as that matter is factual at all"* [16]: and although human beings may be part of such a system, the systems we are interested in, are put together by humans and run by humans, but are in themselves inanimate. [17, 18]. As regards causality in the "legal" sense, this is an issue that also plays a role in the discussion about flood defences: what causes a flood: high water or a low dike. This is a question much like, what is the contribution of the left hand to the noise when clapping hands?

More generally, a cause is the occurrence of a particular combination of the values of relevant parameters that give rise to an accident. Extremes of random variations of values of parameters can combine, such that their combined effect takes a system outside its – "safe" – operating envelope. In the case of accidents, the rare extremes of these independent variables can occur simultaneously by chance, such as in the – sometimes referred to as typically Dutch – problem of assessing the possibility and probability of extreme flood conditions. Here the unknown probabilities of extreme values of heights of water have to be deduced from the distribution of more moderate heights. The probability of extreme weaknesses of dikes has to be inferred from the more familiar state of the sea defences. These have to be combined to result in the probability of the simultaneous occurrence of the two, giving rise to a flood. [19] Causality therefore, can be established in – inanimate –systems in principle. Whether it is worth the effort is a cost-benefit question and therefore profoundly political with emotive moral and ethical dimensions.

Increasingly decision makers and scientists seem of the opinion that these analyses have no value for systems that are too complicated to be understood completely. Rather than making systems simpler they prefer to look at the system as an organic creature which may behave in unexpected and unforeseen ways. In this line of thinking, humans need to cope with these behaviours and rectify them, when they are unwanted, or unsafe.

There is also a different type of accident, specifically the type that occurs when a number of common conditions occur in a very uncommon way – the problem of accidents due to "unusual combination of usual conditions". These types of problems can arise in well-understood systems and should be dealt with in a similar but subtly different way – specifically correcting the deviations from the norm of the parts as soon as they occur. Often the deviations from the norm are minor and are in themselves of no functional consequence individually.

Thus we can see a pattern that we can recognise as perhaps embarrassing, that many systems are working despite shortcomings in the original designs and management arrangements. In the military context, it is recognised that no plan, no matter how well designed, survives first contact with the enemy, but we do need a plan. Similarly, no design, no matter how well thought out, can have foreseen every possible variation in environmental conditions, every possible interaction between components and variables and every possible interaction with humans. Adapt and adjust during operation and managing changes during construction, has long had a place in building works. In the so called SAFETY II approach, we try to learn from how these adjustments work and incorporate them methodically, in order to strive for continuous improvement. It is an iterative, complementary process, which needs an open mind to recognise the human as an essential part of getting it right, even if occasionally human imperfections negate the best laid plans of mice and men, and get it disastrously wrong.

Human Factors approaches, would encourage us to look at the human component in system design, from a different perspective. We should design the systems around our human strengths and weaknesses, not try and force fit this, inevitably individual (– like the blacksmith's nut and bolt), non-standard human component into the system. But having made that paradigm shift, it opens up and explains why, in the past, we have managed to make even the most challenging systems work. The construction of the leaning Tower of Pisa is a classic example of humans adapting to the realities of the environment to make it work, rather than sticking to the script [20].

This approach, however is not without its own challenges.

Functionality

Whereas it is usually clearly defined as to what constitutes a failure, what constitutes success is usually not clearly defined; and even more often, defined in terms of non-failure. A car that does not go is obviously failed. But can a car that still goes on a donut spare still be considered to function correctly, or a car, where the driver of which, has to continuously correct for asymmetric steering behaviour. In the latter case, the driver makes things work, but can it be called a success? In the rules of tennis, it is precisely defined when a ball is in – be it only for the white outer line. If the ball is not in, it is out. Nevertheless the most contested decision is, whether the ball is in or out.

For functionality one could distinguish three stages of performance [21]

(1) Functions as intended,

(2) Functions but not as intended and

(3) Does not function.

Functionality can also be seen as a continuous and distributed entity (Figure 1). The distribution ranges from normal functionality through subnormal functionality to failure, but the demarcations between the areas are not sharp. The problem with a state in which the system functions but not as originally conceived, or designed, or meant, is that there is no longer a precise understanding of what is going on. A broken part is replaced by a part that is not according to specification, but does the job. Is it now necessary to change the inspection regime for that part, or is it assumed that doing the job is a sufficient indicator for all the other demands that were put on the original part such as resistance to wear, weather, ageing?

For unexpected human intervention, this is even more problematic. The apocryphal guy in the blue overalls with the oil can, who keeps the system going may have been acceptable as an engineering solution for a steam locomotive, but regular intermittent halting of machinery, is not considered a valid engineering solution for airplanes.

Functionality therefore is ill defined and usually, with the implicit assumption that human intervention is not outside the boundaries of the user, operator and maintenance instructions. Reference to the emergency procedures, is usually considered as a sign of unwanted deviations on a path leading to disaster, which although it is fixed by human operators should not repeat itself. If human intervention to keep things working requires bending the rules, or violating rules and regulations, the question

should be raised as to whether the total construct of design and prescriptions is still compatible with continued operation. The answer to that question should lead to a decision to change the design, change the rules, or enforce them.

The analysis tools associated with SAFETY-II, such as FRAM (Functional Resonance Analysis Model) [22], the graphical implementation of which is a simplified version of SADT (Structural Analysis and Design Technique) [23], are essentially qualitative. They can be used to describe the structure and behaviour of a system and the way it is supposed to function, but for a quantitative analysis of the potential variations of the behaviour of the system, a process simulator has to be put on top of these models.

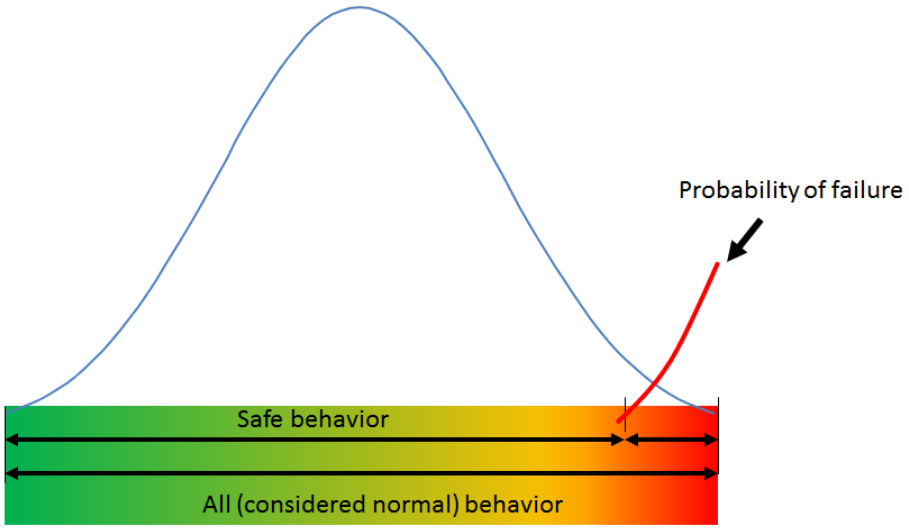

Figure 1 Model simulations show potential failure conditions

Resilience

Resilience engineering accepts that unexpected events can happen. Resilience engineering expects that an intelligent human being will intervene before it is too late. It tends to support this idea that systems should have sufficient "designed-in" capacity to resist and recover from unanticipated upsets. This sounds like a praiseworthy ambition, but again like safety, although in a different way, this seems to be a label which covers a multitude of concepts; which range from the use of more effective or layers of barriers (defences), to designing in some functionality to monitor, respond, adapt and learn from actual operational experiences. Again, inevitably, because this is applied without distinction to everything, from simple engineering systems to large organisations, it is difficult to get a consensus view as to exactly what it is. Let alone how we formally incorporate it into sound engineering practice. Nevertheless this must count as utilising these human skills at workarounds and adapting to the real world, as adding resilience to systems.

The unfortunate side effect of this line of thought is that it entices engineers to refrain from further analysis of possible deviations and their consequences and use these analyses as a basis for design changes, or the incorporation of further protective measures: be it in the form of additional hardware at one end of the spectrum, to additional emergency protocols at the other. The expectation that problems will be dealt with when they arrive, is common in politics and religion, but as Clausewitz [24] explains, this is not a good idea if one contemplates to engage in warfare. This "culture of coping" is remarkably prevalent in a wide range of domains, ranging from built infrastructure, to utilities such as electricity and water quality control. The ultimate outcome of resilience engineering is continuous improvisation. Even for a restaurant, this is not a good idea. It is true that new recipes are often the result of improvisation, but to make it to a Michelin starred restaurant one has to serve the clients consistent quality and the same taste every time a certain recipe is served. Variability produces innovation, but in the end consistency sells the product, as Deming [25] and Juran have pointed out before [26]

A common argument in favour of the SAFETY-II approach is that if the SAFETY I approach is successful and no accidents occur, doubt will be raised as to whether the investment in safety was justified, while in the SAFETY II approach, investments are made that promote productivity and, as a consequence, also promote safety. This argument however falters if humans are considered expendable commodities, which VOSL analyses [27, 28] often demonstrate.

Variability in Safety II

In the SAFETY-II approach reducing variability brings overall system performance closer to the desired optimal state. Variability – especially in human behaviour – gives room for out of bounds ingenuity and therefore curtailing variability and prohibiting moderate deviations from design values is less desired.

Precaution

One of the innate survival skills of the human operator is an ability to assess the safety of actions before embarking (or not) on them. This requires an ability to recognise, in real time, the practical system boundaries of operability that cannot be exceeded

safely. As Rasmussen, pointed out, over time, humans tend to push ever closer towards these boundaries, sometimes exceeding these safety limits. If there is uncertainty, or lack of any reliable measures for these limits, safety considerations dictate that we should err on the side of safety as a precaution. This precautionary principle is thus an implicit recognition of the complexity of the system, or the designers/ operators level of ignorance (or candour?).Perhaps this should trigger a closer look at what is at the heart of this uncertainty, and again intelligently adapt the designs, rather than just extending the limits arbitrarily, uncontrolled and in essence unpredictably. Precaution is often said to get in the way of progress, but in many cases, taking risk without precaution, leads to regret and blame after the unwanted event.

"We Athenians in our persons, take our decisions on policy and submit them to proper discussion. The worst thing is to rush into action before the consequences have been properly debated. And this is another point where we differ from other people. We are capable at the same time of taking risks and estimating them beforehand. Others are brave out of ignorance; and when they stop to think, they begin to fear. But the man who can most truly be accounted brave is he who best knows the meaning of what is sweet in life, and what is terrible, and then goes out undeterred to meet what is to come." (Thucydides,). [29]

Many of successful risk takers and entrepreneurs took meticulous precautions before embarking on any daring enterprise and therefore could be better characterized as successful risk mitigators: "if you are a risk taker, then the art is to protect the downside" (Richard Branson). Insurance and stocks and bonds were not meant to be vehicles of risk taking but instruments to share risks and limit the risks to the individual. Taking up insurance, or releasing bonds, were and are, precautionary measures.

Nevertheless precaution will also invoke a measure of regret about things that could have been done, but have not been done for the sake of safety. Because the advantages gained if these things had been done can never be known, the discussion about missed opportunities is a never ending one [30].

Precaution implies an extreme form of curtailing variability, as certain areas of variation, the areas where the large losses are, are just not entered into.

Big Data

With the use of Big Data technology, however, it is no longer necessary to define failures and failure paths beforehand. The full behaviour of a system can be inferred from actual operating data logs and learned (modelled), which will implicitly include the variability of its parameters. Such a model will show what Hollnagel [31] calls "emergent" behaviour, just as reality does. These model results could thus be observed and unwanted behaviour can be detected. This would be in hindsight after the analysis / calculation of effects, but before the unwanted behaviour emerges in reality. The causes identified can then be explored. These causes, which can be a combination of factors that are considered a bit extreme, have been accepted as normal. The result of modelling all behaviour, is conceptually depicted in figure 1. Average behaviour is then the most probable / expected behaviour. What is unsafe variability will thus reveal itself.

"Normal" chemical plants do not explode on average, nor do average airplanes fall from the sky, or average cars collide. But then, the world is not an average. The world, judged on the average of the currently observed universe should not exist, but it does. So, where, in many areas of technology, the probability of failure is low, but the consequences of failure could be catastrophic, it is no longer sufficient to look at predefined abnormalities. Unfortunately just looking at success, or why a system keeps working, is not sufficient either. Having airplanes diverted to small airfields was, and still is considered normal: just as having pilots and co-pilots with large differences in experience, air traffic controllers with limited command of the English language and taking off with limited visibility, is considered routine. These factors can all be designated as intelligent coping, which contributes to the speedy and successful operation of an airline, even in case of adverse circumstances. They nevertheless combined into the 1977 Tenerife disaster [32]. Using the correlation between success, failure and values of parameters in systems, these "emergent" behaviours can be explored. From that, it can be decided whether it's preferable, or practical to curtail variability, by selection, or exclusion of certain actions.

As described earlier, for this approach to be successful, a much larger dataset is needed, than on failures and near misses alone. Ideally, it would require the complete record of all – relevant – system parameters, during normal and abnormal conditions. A few projects exemplified below are currently aimed at achieving this.

Signals Passed at Danger (SPAD) is a codified deviation in the operation of railways that, at the same time is dangerous and occurs often. These occurrences are often interlinked with the operational requirements and the layout of the safety system. In a modern railway system, the movements of the trains are constantly monitored, creating a wealth of data. A project trying to exploit these data to understand the underlying causes and thereby reduce the number of SPADS and increase the reliability of the operation [33, 34] is expected to reduce costs and enhance the carrying capacity of the railways in the United Kingdom.

Similarly, the Platypus project is aimed at exploiting the automatic registration and storage of operational data in a chemical plant to understand beyond design deviations and remedy them by taking the causes away before they lead to a major process upset or a loss of containment. [35, 36, 37]

These projects combine the idea of resilience, SAFETY-I, SAFETY-II and precaution in a practical system that at the same time reduces the potential for incidents and accidents and improves the overall performance of the system.

Costs and benefits

The standard approaches in SAFETY-I, such as FMEA and Bow-Tie analysis, have in common, that they are all based on the principles of fault and consequence analysis. These analyses can be visualized as structures in the form of trees or directional acyclic graphs, which lend themselves for quantification, if the base events, roots or equivalent entities in the analysis can be given a quantitative expression, such as a probability, frequency, or extent of the damage predicted. As has been demonstrated

elsewhere, there is no need for systems to behave linearly for these methods to be employed. For a successful quantification, the metrics of these need to be consistent. The consequence metrics need not necessarily just be monetary. Often the consequences are expressed in a set of metrics, such as money, people injured and people killed. The desire to be consistent over multiple areas of policy, then leads to attempts to unify these metrics. The choice of the common metric then is usually money, which leads to the ethically unresolvable discussion about the value of a human life [27, 28]. Nevertheless the approaches employed in SAFETY-I, do lend themselves to quantification and therefore to a more objective comparison of costs and benefits. The costs are normally those of safety measures and the benefits those of avoiding incidents and accidents. As in all analytical techniques, there is the problem of uncertainty. In most cases the values of parameters are only known with limited accuracy; and thus the results of cost-benefit evaluations are uncertain as well. The major uncertainty is whether all possibilities have been covered. This is also known as the "black swan" problem. There may always be surprises. In some discussions, potential events that are deemed to have too low a probability to consider, are called "black swan" events as well, but the costs benefit relationship of these foreseeable, but disregarded, events, could have been considered and therefore are not truly, a surprise.

For SAFETY-II a cost benefit evaluation is more difficult because the methods employed are currently qualitative; perhaps with the exception of the cases where success is defined as the absence of failure. In the latter case SAFETY-II in essence reverts to SAFETY- I. Because the primary objective of SAFETY-II is to improve functionality, there is no à priori relation between the results obtained through this method and increased safety. There are many ways in which the functionality of a system can be improved without any influence on safety. One could, for instance, construct a better building and unintentionally increase the number of injured, or killed workers at the same time. An additional problem is the loose definition of a measure of the effectiveness of the functional system and exactly how improvements could be measured.

For resilience, a cost benefit evaluation is even more challenging. When resilience is interpreted as having redundant defences and defences in depth, resilience engineering reverts to SAFETY-I. When resilience is interpreted as assuming that any problems will be successfully solved by human ingenuity, when and if they arrive in the future, there is an immediate cost saving, as complicated in-depth analyses of potential faults and their consequences, is no longer necessary. Another cost reduction results, because potentially costly measures to take away problems that could, or would emerge from these analyses, are not necessary either. Resilience engineering therefore is an attractive alternative to a SAFETY-I approach. Since the paradigms behind resilience engineering implicitly, or explicitly, assume that future problems will be solved successfully, the problem of "black swan" events disappears. In fact, many of the potential events that would be discovered by SAFETY-I analyses may now be future surprises.

The precautionary approach is probably the most expensive of the four approaches discussed in this paper. On the cost side there are the costs of the measures taken and also the costs of missed opportunities to consider. On the benefit side, nothing can be proven, as it has been avoided. Precaution thus foregoes the evaluation of probabilities. Only the potential negative outcomes are considered sufficient motivation to take precautionary action [38]. Therefore the risk cannot be evaluated and thus a cost estimation is impossible. However, this may be deceptive. If the activity with these potential adverse consequences is still undertaken and the adverse consequences arrive anyway, the costs could be catastrophic and so would be the regret.

Conclusion

Currently four main streams of safety engineering can be identified: SAFETY-I, SAFETY-II, Resilience Engineering and Precaution. Each of these try to deal with the effects of the variability of nature and of human behaviour. SAFETY-I and Precaution limit this variability by prohibiting the system entering into a state that, à priori, is designated to be unsafe. If an excursion into an unsafe state cannot be avoided, additional barriers are added to the system to mitigate the effects of the excursion.

SAFETY-II and Resilience Engineering, on the contrary, accept variability and exploit the variability and ingenuity of human beings to cope with problems as they arrive.

SAFETY-I and Precaution designate variability as, in principle, unwanted. In SAFETY-II and Resilience Engineering it is desirable. However, the latter implicitly assumes that future variations will stay within bounds and the system does not stray so far outside the safe operational envelope, that recovery is not possible. They share with the former two the aim of avoiding loss of health, or life, and/or catastrophic losses and ruin.

Of the four, only SAFETY-I approaches currently allow quantitative evaluation of costs and benefits. The other three rely on qualitative estimates in spirit ranging from "things work out fine" to "doom will be upon us unless". Although softer approaches can seem more attractive than hard-core technological approaches to organizations with budgetary constraints, there are a few caveats. In the short term, using humans to prevent unwanted events can often seem less expensive, than hardware solutions, leading to less technological defence in depth. Softer qualitative approaches do not seem to require as complete an understanding of technology, nor demand in depth analyses, neither before, nor after, the accident, which saves time and money.

However as Taleb [39] shows in his book on Black Swans, variability induces uncertainty and the propensity of so-called "outliers" and simultaneous occurrence of extreme values within otherwise acceptable ranges is often if not always underestimated [40]. In the end therefore, the same holds for safety, as for quality control and the stock exchange: variability and the associated uncertainty in future states, needs to be reduced to a minimum as much as possible.

References

1 Lord Robens, Safety and Health at Work. Report of the Committee. 1970–72, HMSO, Cmnd 5034

2 B.J.M. Ale, L.J. Bellamy, R.M. Cooke, L.H.J. Goossens, A.R. Hale, A.L.C. Roelen, E. Smith. (2006) Towards a causal model for air transport safety—an ongoing research project - *Safety Science 44 (2006) 657–673*

3 Roelen, A. L. C., Lin, P. H., & Hale, A. R. (2011). Accident models and organisational factors in air transport: The need for multi-method models. *Safety Science*, 49, 5-10

4 B.J.M. Ale The Occupational Risk Model, TU-Delft/TBM RC 20060731, ISBN 90-5638-157-1, Delft, 2006

5 L. J. Bellamy, B.J.M. Ale, J.Y.Whiston, M.L. Mud, H. Baksteen, A. Hale, I.A. Papazoglou, A. Bloemhoff , J.I.H Oh. (2006) The software tool Storybuilder and the analysis of the horrible stories of occupational accidents, *Working on Safety, 12-15, September 2006*

6 Ioannis A. Papazoglou and Ben J.M. Ale 2007 A logical model for quantification of occupational risk •*Reliability Engineering & System Safety*, *Volume 92, Issue 6, June 2007, Pages 785-803.*

7 Ale, B.J.M , L.J. Bellamy, R. van der Boom,R.M. Cooke, L.H.J.Goossens, A.R. Hale, D. Kurowicka, P.H. Lin, O., Morales, A.L.C., Roelen, J., Cooper, J., Spouge. (2006) CATS final report, Ministry of Transport and Water management, The Hague, The Netherlands, ISBN 10: 90 369 1724-7; ISBN 13: 978 90 369 1724-7

8 ICAO (2000), International Civil Aviation Organization, Accident/Incident Reporting Manual (ADREP), ICAO, Montreal, Canada

9 Denis Besnard, Erik Hollnagel. (2014) I want to believe: some myths about the management of industrial safety. Cognition, Technology and Work, Springer Verlag, 2014, 16 (1), pp.13-23. <10.1007/s10111-012-0237-4>. <hal-00720270>

10 Heinrich, H., (1931), Industrial Accident Prevention a Scientific Approach, first ed. McGraw-Hill Book Company, London.

11 Paul Swuste, Coen van Gulijk, Walter Zwaard, (2010) Safety metaphors and theories, a review of the occupational safety literature of the US, UK and The Netherlands, till the first part of the 20th century, *Safety Science 48 (2010) 1000–1018*

12 2012 Lin, Pei-Hui; Hanea, Daniela; Ale, Ben; Sillem, Simone; Gulijk, Coen; Hudson, Patrick (2012); Integrating organisational factors into a BBN model of risk; ; *PSAM 11, Esrel 2012, Helsinki 25-29 juni 2012*

13 Perrow, C., 1984. Normal accidents. In: Living with High-Risk Technologies. Princeton University Press, Princeton, NJ

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

15 Russell Bertrand (1946), History of western philosophy. London: George Allen & Unwinn; 1946

16 Chalmers DJ. The conscious mind: in search of a fundamental theory. Oxford: Oxford University Press; 1996.

17 ArshinovVladimir,,Christian Fuchs,editors (2003),.Causality,emergence,self- organisation, /http://www.self-organization.org/results/book/Emergence- Causality Self-Organisation.pdfS, 2003

18 Goldstein Jeffrey (1999) Emergence as a construct: history and issues. *Emergence 1999;1(1):49–72*

19 van Gelder, Pieter (2007) Quantitative methods for flood risk management, statistical extremes and environmental risk. Faculty of Sciences, University of Portugal, Lisbon, Portugal, February 15–17, 2007

20 Hollnagel, E., R.L. Wears, J. Braithwaite, (2015) From Safety-I to Safety-II: a white paper. https://www.england.nhs.uk/signuptosafety/wp-content/uploads/sites/16/2015/10/safety-1-safety-2-whte-papr.pdf (last visited 18-11-2018)

21 B.J.M. Ale, L.J. Bellamy, A.L.C. Roelen, R.M. Cooke, L.H.J.Goossens, A.R. Hale, D. Kurowicka, E. Smith (2005) Development of a causal model for air transport safety, IMECE 2005, 79374, *Proceedings of IMECE, 2005 ASME International Mechanical Engineering Congress and Exhibition, Orlando, Florida, nov 5-11, 2005, ISBN 0-7918-3769-6*

22 E. Hollnagel (2012) FRAM, The Functional Resonance Analysis Method, CRC Press, ISBN 9781351935968

23 D. Marca, C. McGowan (1987), Structured Analysis and Design Technique, McGraw-Hill, 1987, ISBN 0-07-040235-3

24 Clausewitz, Carl von, (1992) On War, translated by Howard, Princeton university press, 1992, ISBN 0691018545

25 Deming, W. E. (1982), Quality, productivity and competitive position, Massachusetts Institute of Technology, Cambridge, ISBN 0911379002

26 Juran, J. M. and Gryna, F. M. (Eds.) (1988), Quality control handbook, (4th. ed.), McGraw-Hill, New York ISBN-10: 9780070331761

27 B.J.M. Ale, D.N.D. Hartford, D. Slater (2015) ALARP and CBA all in the same game, *Safety Science 76 (2015) 90–100*

28 B.J.M. Ale, D.N.D. Hartford, D.H. Slater (2018) The practical value of life: priceless or a CBA calculation? *Medical Research Archives, vol. 6, issue 3 (2018) ISSN: 2375-1924*

29 Thucydides, 431, http://www.historywiz.com/primarysources/funeraloration.htm

30 R.B. Jongejan, B.J.M. Ale, H.J. Pasman(2006) The precautionary principle and industrial safety regulation, I

31 Hollnagel 2006, Resilience Engineering, Ashgate, ISBN 978-0-7546-4904-5p12

32 https://en.wikipedia.org/wiki/Tenerife_airport_disaster

33 Peter Hughes, David Shipp, Miguel Figueres-Esteban, Coen van Gulijk, (2018) From free-text to structured safety management: Introduction of a semi-automated classification method of railway hazard reports to elements on a bow-tie diagram, *Safety Science 110 (2018) 11–19*

34 Rawia Ahmed Hassan E.L. Rashidy, , Peter Hughes, Miguel Figueres-Esteban, Chris Harrison, Coen Van Gulijk (2018), A big data modeling approach with graph databases for SPAD risk, *Safety Science 110 (2018) 75–79*

35 Ben Ale, Coen van Gulijk, Anca Hanea, Daniela Hanea, Patrick Hudson, Pei-Hui Lin, Simone Sillem Towards BBN based risk modelling of process plants, *Safety Science 69 (2014) 48–56*.

36 Gulijk, C.van, Ale, B.J.M., Ababei, D., Steenhoek, M.(2014) Comparison of risk profiles for chemical process plants using PLATYPUS Safety and Reliability: Methodology and Applications - *Proceedings of the European Safety and Reliability Conference, ESREL 2014 2015, Pages 1363-1368 European Safety and Reliability Conference, ESREL 2014; Wroclaw; Poland; 14 September 2014 through 18 September 2014; Code 107147*

37 C. van Gulijk, D.H. Hanea, K.Q. Almeida, M. Steenhoek & B.J.M. Ale, Dan Ababei (2013) Left-hand side BBN model for process safety, *Safety, Reliability and Risk Analysis: Beyond the Horizon – Steenbergen et al. (Eds) pp 1867-1873, Taylor & Francis Group, London, ISBN 978-1-138-00123-7*

38 B. Ale, T. Aven, R. Jongejan, (2010) Review and discussion of basic concepts and principles in integrated risk management, in Reliability, Risk and Safety: Theory and Applications – Briš, Guedes Soares & Martorell (eds)© 2010 Taylor & Francis Group, London, ISBN 978-0-415-55509-8-

39 Taleb, N.N. (2007) The black swan: The impact of the highly improbable. London: Penguin, ISBN 978-1400063512

40 http://functionalresonance.com/brief-introduction-to-fram/index.html (last visited 27-11-2018