JESK J Ergon Soc Korea 2016; 35(3): 185-191
http://dx.doi.org/10.5143/JESK.2016.35.3.185
http://jesk.or.kr eISSN:2093-8462

Resilience Engineering: A New Understanding of Safety

Erik Hollnagel

Professor, University of Southern Denmark Chief Consultant, Centre for Quality, Region of Southern Denmark

Corresponding Author
Erik Hollnagel
Professor, University of Southern Denmark
Chief Consultant, Centre for Quality,
Region of Southern Denmark
Email : hollnagel.erik@gmail.com

Received : June 21, 2016
Accepted : June 21, 2016

1. Introduction

When most people hear the word 'safety' they think of situations where something has gone wrong. It can be smaller events from their own experience, or major accidents they have read about in the news. 'Safety' makes us think about incidents and accidents - about (low probability) events with adverse outcomes, which is why safety traditionally is defined as a condition where nothing goes wrong as illustrated by the following typical definitions:

- Safety is the freedom from unacceptable risk (The American National Standards Institute).
- Safety is the freedom from accidental injury (U.S. Agency for Healthcare Research and Quality).
- Safety is the state in which harm to persons or of property damage is reduced to, and maintained at or below, an acceptable level through a continuing process of hazard identification and risk management (International Civil Aviation Organization).

The reason for this understanding of safety is not hard to find. It makes good practical sense to pay attention to situations where something has gone or may go wrong, both because such situations by definition are unexpected and because they may lead to harm or injury - the loss of life, materials, or money. It is therefore a natural response to try to find out why something went wrong so that steps can be taken to prevent it from happening again.

1.1 The causality credo

Explanations of how accidents happen usually refer to a set of assumption about how causes lead to effects, often expressed as an accident model. The simplest accident model is the Domino model (Heinrich, 1931), which relies on simple linear causality using the analogy of a set of domino pieces that fall one after the other. According to the logic of simple linear causality, the purpose of event analysis is to reason backwards from the injury to find the 'root cause'. Similarly, the purpose of risk analysis is to look for whether a specific component may fail or malfunction, either by itself or in combination with another failure or malfunction, and thereby become the first domino piece to fall.

Simple linear models were in the 1980s complemented by composite linear models, of which the best known example is the Swiss cheese model (Reason, 1997). These models explain adverse outcomes as combinations of active failures (or unsafe acts) and latent conditions (hazards conceived as degraded barriers or weakened defences). An event analysis thus looks for how degraded barriers or defences interacted with active (human) failures. Similarly, risk analysis looks for conditions where combinations of single failures and latent conditions may result in an adverse outcome.

Common to all accident models is the unspoken assumption that outcomes can be understood as effects that follow from prior causes. Since that corresponds to a belief in the laws of causality, it may be called a *causality credo*. This can be expressed as follows:

- Adverse outcomes happen because something has gone wrong. Adverse outcomes have causes. There is also a valence or value congruence between causes and outcomes, so that bad outcomes have bad causes and vice versa.
- It is possible to find these causes provided enough evidence is collected. Once the causes have been found, they can be eliminated, encapsulated, or otherwise neutralised.
- Since all adverse outcomes have causes, and since all causes can be found, it follows that all accidents can be prevented. This is the vision of zero accidents or zero harm that many companies find attractive.

The causes that we look for, and therefore also the causes we find, necessarily corresponds to our current view of how the world functions. In the earliest days, accidents were explained by invoking 'acts of god' or 'acts of nature', and therefore as being beyond human control. As humans gradually became masters of technology, especially after the second industrial revolution around 1750, 'acts of nature' were replaced by technological failures or shortcomings. The role of technology as the main source of problems, and paradoxically also of solutions, was generally accepted until 1979, when the accident at the Three Mile Island nuclear power plant demonstrated that safeguarding technology was not enough. Technological malfunctions were replaced by 'human error' as a convenient cause, although the latter was harder to improve or eliminate than the former. In 1986, only seven years later, the loss of the space shuttle Challenger, reinforced by the accident in Chernobyl, required explanations that went beyond 'human error'. This led to the introduction of concepts such as organisational failures and safety culture. While this sufficed for a while, there has lately been a trend to adopt yet another cause, namely 'complexity'.

The *causality credo* not only represents the firm human belief that there must be a reason or explanation for everything, but also represents three biases that go along with the belief in causation. The first is the proportionality bias, which says when that something big happens then something big must have caused it. The second is the valence bias, which says that negative consequences have negative causes and vice versa. And finally, there is the intentionality bias, which says that when something ambiguous happens then it must have been intended.

New types of accidents have, historically speaking, always led to new types of causes but without challenging the underlying assumption of linear causality. We have therefore become so used to explain accidents in terms of cause-effect relations that we no longer notice it. And we continue to do so although it has becomes increasingly difficult to reconcile with reality.

2. Safety-I: Avoiding That Things Go Wrong

The historical development of safety thinking in combination with the *causality credo* leads to the view of safety illustrated by Figure 1. According to this, unacceptable outcomes happen because of preceding failures and malfunctions (cf., the intentionality bias), while acceptable outcomes happen because everything - including people - worked as it should. This implies a *'hypothesis of different causes*', namely that the causes or 'mechanisms' of adverse events are different from those of events that succeed. If that was not the case, then the elimination of the causes and the neutralisation of the 'mechanisms' would also reduce the likelihood that things could go right, hence be counterproductive.

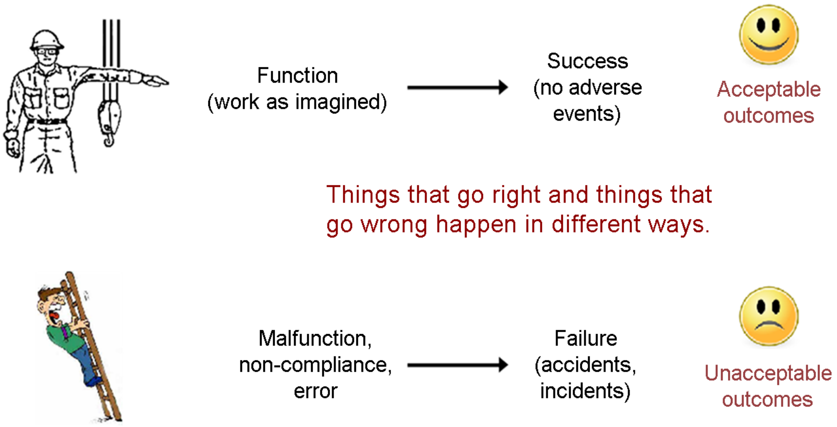

Figure 1. The Safety-I view of failures and accidents

This way of thinking corresponds to a definition of safety as a condition where the number of adverse outcomes (accidents / incidents / near misses) is as low as possible. Since this is also the traditional way of defining safety, it can be called Safety-I (The alternative proposed by resilience engineering is called Safety-II). Safety-I tacitly assumes that systems work because they are well designed and carefully maintained, because procedures are complete and correct, because even minor contingencies have been foreseen and anticipated, and because people behave as they are expected to. The purpose of safety management consequently becomes how to achieve and maintain that state. When people's performance is different from what has been predicted this unavoidably leads to an emphasis on *compliance* in the way work is carried out.

The definition of Safety-I is however problematic in several ways. One is that safety is defined by its opposite, by what happens when it is missing. Another that safety is measured indirectly, not by its presence or as a quality in itself, but by the consequences of its absence. And finally that the study of safety focuses on accidents and incidents, which are situations where safety by definition is absent.

3. Looking at what Goes Wrong Rather than Looking at what Goes Right

Resilience engineering has since its inception argued that the Safety-I perspective is both oversimplified and misdirected. Resilience engineering rejects the hypothesis of different causes and instead argues that things happen in basically the same way regardless of whether they go right or go wrong (Hollnagel et al., 2011; Hollnagel, 2014). Resilience engineering also points out that things as a rule go right and only exceptionally go wrong. Indeed, when something goes wrong it is more than likely that it has been done many times before or has happened many times before and that it always has gone right. It therefore makes sense to try to understand how acceptable outcomes happen as a basis for trying to understand how unacceptable outcomes happen. The consequences of looking at what goes wrong rather than looking at what goes right are illustrated by Figure 2, which represents the case where the (statistical) probability of a failure is 1 out of 10,000. In other words, for every time we expect that something will go wrong (the red line), there are 9,999 times where we should expect that things will go right and lead to the outcome we want (the green area).

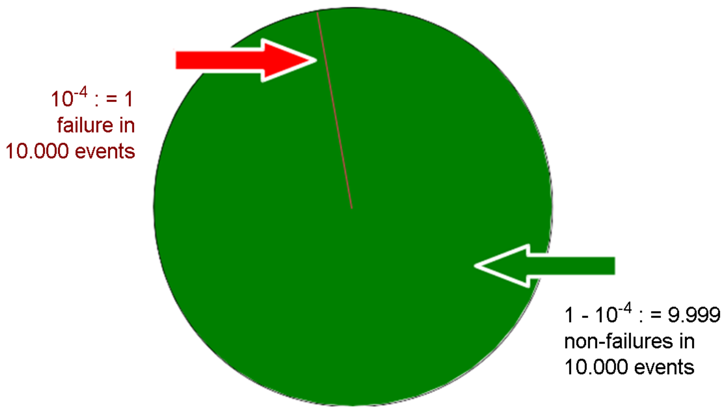

Figure 2. The imbalance between things that go right and things that go wrong

The focus on what goes wrong is required by regulators and authorities, supported by models and methods, documented in countless databases, and described in literally thousands of papers, books, and conference proceedings. The net result is a deluge of information both about how things go wrong and about what must be done to prevent this from happening. The recipe is the simple principle known as 'find and fix': look for failures and malfunctions, try to find their causes, and try to eliminate causes and/or improve barriers.

The situation is quite different when it comes to that which goes right, i.e., the 9,999 events out of the 10,000. The focus on what goes right receives little encouragement; it is not required by authorities; there are few theories or models about how human and organisational performance succeeds, and few methods to help us study how it happens; actual data are difficult to locate; it is hard to find papers, books or other forms of scientific literature about it; and there are few people who even consider it worthwhile. In other words, we spend nearly all our time and resources to understand why things go wrong, but very little to understand why they go right. We stubbornly study the absence of safety, the one event of of 10.000 that goes wrong, but neglect to study the presence of safety, the 9.999 times out of 10.000 that go right!

4. Why do things go Right?

Resilience engineering proposes that things go well because people are able to adjust what they do to match the conditions of work. Furthermore that this happens on all levels of an organisation, from board meetings to floor sweeping. People learn to identify and compensate for design flaws and functional glitches, they learn to recognise the actual demands and adjust their performance accordingly, and they learn interpret and apply procedures to match the conditions. People also notice when something is about to go wrong, and try to intervene before the situation becomes grave. This can be described as performance variability, not in the negative sense of deviations from some norm or standard, but in the sense of the smooth adjustments that are necessary for safety and productivity.

Performance adjustments are a *sine qua non* for the functioning of complex socio-technical systems. Attempts to prevent unacceptable outcomes or failures by eliminating or constraining performance adjustments, for instance by insisting on compliance, are counterproductive since they will affect the ability to produce the desired acceptable outcomes. Instead efforts should be made to facilitate the necessary performance adjustments by making it easier to perceive the resources and constraints of a situation and by making it easier to anticipate the consequences of actions. Performance adjustments should be managed by attenuation (dampening) if they seem to go in the wrong direction and by strengthening (amplification) if they seem to go in the right direction. In order to do so it is necessary first to acknowledge the inevitability and necessity of performance variability, second to find ways to monitor it, and third to find ways to control it.

5. Safety-II: Ensuring That Things Go Right

Our socio-technical environment continues to develop and become more complicated, not least due to the temptation of ever more powerful information technology. The thinking represented by Safety-I is therefore increasingly unable to deliver the required and coveted 'state of safety'. The common solution is to 'stretch' Safety-I tools and concepts even further until they 'break'. Another solution is to change the definition of safety from 'avoiding that something goes wrong' to 'ensuring that everything goes right' - or more precisely to the ability to succeed under varying conditions, so that the number of intended and acceptable outcomes (in other words, everyday activities) is as high as possible. This can be called Safety-II (Figure 3). The basis for safety and safety management now becomes an understanding why things go right, which means an understanding of everyday activities.

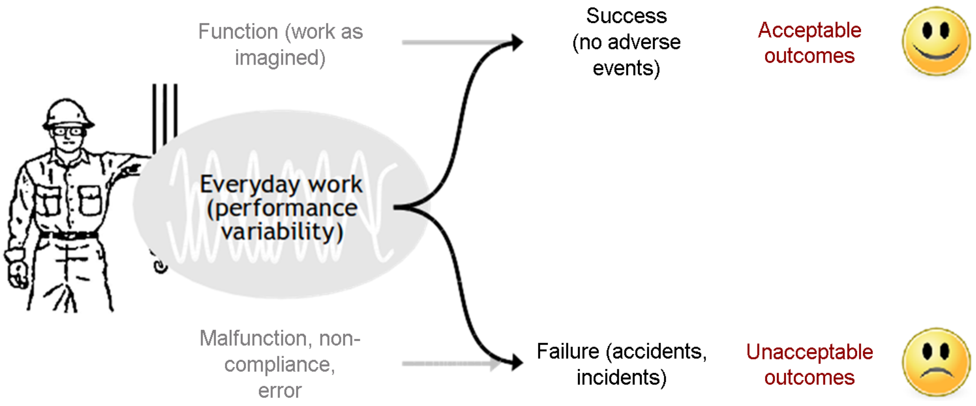
Function (work as imagined)
Everyday work (performance variability)
Malfunction, non-compliance, error
Success (no adverse events)
Failure (accidents, incidents)
Acceptable outcomes
Unacceptable outcomes

Figure 3. The Safety-II view of failures and successes

Because everything basically happens in the same way regardless of the outcome, it is no longer necessary need to have one type of causes and 'mechanisms' for things that go wrong (accident and incidents) and another for things that go right (everyday work). The purpose of safety management is to ensure the latter, since by doing so it will also reduce the former. Safety-I and Safety-II therefore both lead to a reduction in unwanted outcomes, but use fundamentally different approaches with important consequences for how the process is managed and measured - as well as for productivity and quality.

From a Safety-II perspective, the purpose of safety management is to ensure that as much as possible goes right and that everyday work achieves its intended purposes. This cannot be done by responding alone, since that will only correct what has happened. Safety management must also be proactive. For this to work, it is necessary to think about what can happen and to have the appropriate means (people and resources) to do something about it. That in turn requires an understanding of how the system works, of how its environment develops and changes, and of how functions may depend on and affect each other. This understanding can be developed by looking for patterns and relations across events rather than for causes of individual events. To see and find those patterns, it is necessary to take time to understand work-as-done rather than to spend all resources on fire-fighting.

6. Conclusion

Safety efforts under Safety-I are usually triggered by some kind of adverse event or unexpected outcome. The larger and the more severe the event is, the more urgent and extensive is the response. The primary aim is to prevent the adverse event from ever happening again, either by trying to identify and eliminate the causes (hazards) or by finding ways to contain the outcome or consequences of the accident. This typically involves comparing what actually happened (work-as-done) to what had been assumed or prescribed (work-as-imagined) and categorising the actual actions as one form of non-compliance or the other.

Safety efforts under Safety-II lead to a different practice. First and foremost to look at 'work-as-done' - to pay attention to what happens when 'nothing' happens - and to recognise the habitual ways of working that are the basis for everyday productivity. Work well done is impossible with adjustments, large and small. The adjustments are the reason both for acceptable and unacceptable outcomes

In Safety-I learning is based on accidents and learning efforts are proportional to the severity of the outcome. In Safety-II learning should be based on the frequency of events rather than the severity. It is simpler and less costly to make small changes to everyday performance than to make large changes to rare performance. The benefits are also easier to see and to calculate, and give a better return on investment.

Overall, safety management can and should not be based on accidents and incidents, which represent snapshots of unacceptable performance. Safety management should instead care about what happens all the time, about the continuous flow of activities that constitute work-as-done. We nearly always know how many times something has failed or go wrong, but we rarely know how many times or how often something just works.

Resilience engineering does not argue for a wholesale replacement of Safety-I by Safety-II, but rather proposes a combination of the two ways of thinking. Safety-II is first and foremost a different understanding of what safety is, hence also a different way of applying many of the familiar methods and techniques. In addition to that it will also require methods on its own, to look at things that go right, to analyse how things work, and to *manage* performance variability rather than just *constraining* it. We cannot make things go right simply by preventing them from going wrong. We can only make things go right by understanding the nature of everyday performance and by learning how to perceive those things which we otherwise do not see.

References

Heinrich, H.W., Industrial accident prevention: A scientific approach. McGraw-Hill, 1931.

Hollnagel, E., Paries, J., Woods, D.D. and Wreathall, J. (Eds.). Resilience engineering in practice: A guidebook. Farnham, UK: Ashgate, 2011.

Hollnagel, E., Safety I and Safety II. The past and future of Safety management. Franham, UK: Ashgate, 2014.

Reason, J.T., Managing the risks of organizational accidents. Aldershot, UK: Ashgate Publishing Limited, 1997.

Author listings

Erik Hollnagel: hollnagel.erik@gmail.com

Highest degree: M.Sc. in psychology from the University of Copenhagen, and a Ph.D. in psychology from the University of Aarhus (Denmark)

Position title: Professor at the University of Southern Denmark and Chief Consultant at the Centre for Quality Improvement, Region of Southern Denmark. Adjunct Professor at Central Queensland University (Australia), Visiting Professorial Fellow, Faculty of Medicine and Health Sciences at Macquarie University (Australia), Visiting Fellow of the Institute for Advanced Study of the Technische Universitat München (Germany), Professor Emeritus at the Department of Computer and Information Science (IDA) at Linköping University (LIU), Sweden

Areas of interest: Resilience engineering, system safety, human reliability analysis, cognitive systems engineering, intelligent man-machine systems