

Understanding Accidents - From Root Causes to Performance Variability

Erik Hollnagel

*CSELAB, Department of Computer and Information Science, University of Linköping, Sweden
Email: erik.hollnagel@ida.liu.se*

Abstract--When an accident happens, it is obviously important to understand what caused it in order to take effective preventive measures. Accident analysis always implies an accident model, i.e., a set of assumptions of what the underlying “mechanisms” are. Over the last 50-75 years there have been significant changes in accident models, leading to changes in the methods and goals of accident analysis. In parallel to this development the understanding of the role of humans in accidents, and of the nature of “human error”, has also changed. This paper provides an overview of the developments, and outlines the consequences for contemporary accident analysis and prevention.

Index terms--Accident analysis, accident prevention, action failure, blunt end – sharp end, epidemiological model, sequential model, systemic model

I. THE NEED OF ACCIDENT MODELS

It is a truism that we cannot think about something without having words and concepts that describe it, or without having some frame of reference. Very often the frame of reference represents an unspoken but commonly held view that is part and parcel of a specific technical culture. The advantage of using a frame of reference is that communication and understanding become more efficient, because so many things can be taken for granted. The disadvantage is that it strongly favours a single point of view, which rarely is questioned. This makes it more difficult to be thorough in analysis, in the sense of considering alternative explanations.

The frame of reference is particularly important in thinking about accidents, because it determines how we view an accident and in particular how we view the role of humans. I shall refer to this frame of reference as the accident model, i.e., a stereotypical way of thinking about how an accident occurs. Although there are many individual models, they seem to correspond to one of the three types characterised below.

A. Sequential Accident Models

The simplest types of accident models describe the accident as the result of a sequence of events that occur in a specific order. This has been expressed as the First Axiom of Industrial Safety, which reads:

“The occurrence of an injury invariably results from a completed sequence of factors – the last one of these being the accident itself. The accident in turn is invariably caused or permitted by the unsafe act of a person and/or a mechanical or physical hazard.”

(Heinrich, Petersen & Roos, 1980; org. 1931, p. 21)

This Axiom was also called the domino theory and visualised in terms of a set of dominos. As everyone knows, if one domino falls it will knock down those that follow. If the dominos therefore represent accident factors, the model represents how these factors constitute a sequence of events where the linking of cause and effect is simple and deterministic.

Another and much later example is the Accident Evolution and Barrier model (Svenson, 1991, 2001), which describes an accident in terms of a sequence of events – or rather barriers – that failed. This description puts the focus on what went wrong, but in doing so leaves out additional information that may be potentially important. More generally, sequential models represent the accident as the outcome from a series of individual steps organised according to their order of occurrence. Sequential models need not,

of course, be limited to a single sequence of events but may be represented in the form of hierarchies such as the traditional event tree or networks such as Critical Path models or Petri networks. They may represent either the scenario as a whole, or only the events that went wrong. Figure 1 shows a typical example of a sequential model known as the “anatomy of an accident” (Green, 1988).

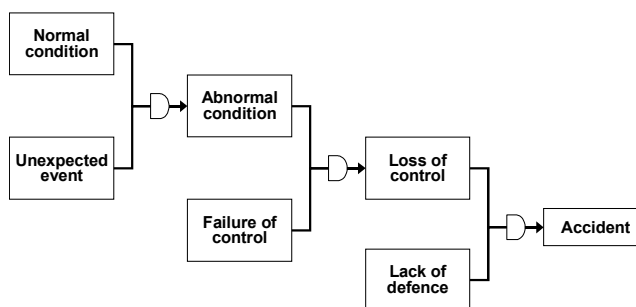


Figure 1: Anatomy of an accident

Sequential models are attractive because they encourage thinking in causal series rather than causal nets (cf. Dörner, 1980). They are furthermore easy to represent graphically, which facilitates communication of the results. While the sequential models were adequate for the socio-technical systems in the first half of the 20th Century, they turned out to be limited in their capability to explain accidents in the more complex systems that became common in the last half of the century. The need of more powerful ways of understanding accidents led to the class of epidemiological accident models, which began to gain in popularity in the 1980s.

B. Epidemiological Accident Models

Epidemiological models, as the name implies, describe an accident in analogy with a disease, i.e., as the outcome of a combination of factors, some manifest and some latent, that happen to exist together in space and time. The term was used as far back as 1961, when Suchman proposed that an accident phenomenon is “the unexpected, unavoidable unintentional act resulting from the interaction of host, agent, and environmental factors within situations which involve risk taking and perception of danger” (Suchman, 1961; quoted in Heinrich, Petersen & Roos, 1980, p. 50). According to this view an accident will have observable and measurable effects, but the accident itself results from a combination of “agents” and environmental factors that create an unhappy setting. The epidemiological accident model was alluded to in the analysis of the Chernobyl accident, which contains the following passage:

“All man-made systems have within them the seeds of their own destruction, like ‘resident pathogens’ in the human body. At anyone time, there will be a certain number of component failures, human errors and ‘unavoidable violations’. No one of these agents is generally sufficient to cause a significant breakdown. Disasters occur through the unseen and usually unforeseeable concatenation of a large number of these pathogens.”

(Reason, 1987)

The concept of a pathogen or a specific causative agent is clearly taken

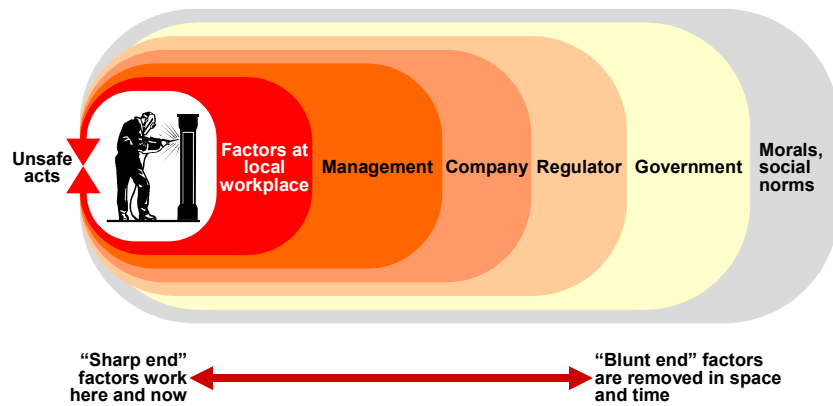


Figure 2: Sharp-end – blunt-end couplings.

from medical terminology, as is the notion of the pathogen being resident. It corresponds to the notion of latent conditions as proposed by Reason (1990). Other examples are models that consider barriers and carriers such as the well-known Swiss cheese analogy (Reason, 1997), models of “sharp end”-“blunt end” interactions (Woods et al., 1994), and models of pathological system (organisation) states.

Epidemiological models are valuable because they provide a basis for discussing the complexity of accidents that overcome the limitations of sequential models. The notion of latent factors simply cannot be reconciled with the simple idea of a causal series, but requires a more powerful representation – at least that of a causal network. This means that the analysis cannot be a search for simple causes, but must involve an account of more complex interactions among different factors. Unfortunately, epidemiological models are rarely stronger than the analogy behind, i.e., they are difficult to specify in further detail, even though the concept of pathogens allows for a set of methods that can be used to characterise the general “health” of a system (Reason, 1997). A third type of models was needed to overcome these limitations.

C. Systemic Accident Models

The third type of models is the so-called systemic model. As the name denotes, these models endeavour to describe the characteristic performance on the level of the system as a whole, rather than on the level of specific cause-effect “mechanisms” or even epidemiological factors. Instead of using a structural decomposition of the system, the systemic view considers accidents as emergent phenomena, which therefore also are “normal” or “natural” in the sense of being something that must be expected. This is consonant with Perrow’s notion of normal accidents (Perrow, 1984). Systemic models have their roots in control theory (Sheridan, 1992), in chaos models, in coincidence models, and most recently in the idea of stochastic resonance. In general, systemic models emphasise the need to base accidents analysis on an understanding of the functional characteristics of the system, rather than on assumptions or hypotheses about internal mechanisms as provided by standard representations of, e.g., information processing or failure pathways. An accident can be described neither as a causal series nor as a causal net, since either representation is incapable of accounting for the dynamic nature of the interactions and dependencies. Systemic models deliberately try to avoid a description of an accident as a sequence or ordered relation among individual events or even as a concatenation of latent conditions, and are therefore difficult to represent graphically.

E. Accident Models And Accident Analysis

The three main types of accident models are summarised in TABLE 1. Each type carries with it a set of assumptions about how an accident analysis should take place and what the response should be.

TABLE 1: THE MAIN TYPES OF ACCIDENT MODELS.

Model type	Search principle	Analysis goals	Example
Sequential models	Specific causes and well-defined links	Eliminate or contain causes	Linear chain of events (domino) Tree models Network models

Epidemiological models	Carriers, barriers, and latent conditions	Make defences and barriers stronger	Latent conditions Carrier-barriers Pathological systems
Systemic models	Tight couplings and complex interactions	Monitor and control performance variability	Control theory models Chaos models, stochastic resonance, coincidence models

For sequential models the accident analysis is usually a search for specific causes and well-defined cause-effect links. The underlying assumption is that the accident is the result of a sequence of events and that causes, once they have been found, can be eliminated or encapsulated thereby effectively preventing future accidents.

In the case of epidemiological models, the accident analysis is usually a search for “carriers” and latent conditions, as well as for reliable indicators of general system “health”. On a more general level the search is for characteristic performance deviations, with the recognition that these can be complex phenomena rather than simple manifestations. In the case of the epidemiological models, the underlying assumption is that defences and barriers can be put into place and/or strengthened either to prevent future accidents from happening or to diminish the effects.

Finally, for systemic models the analysis is a search for unusual dependencies and common conditions that from experience are associated with accidents. This reflects the belief that there always will be variability in the system and that the best option therefore is to monitor the system’s performance so that potentially uncontrollable variability can be caught early on. Variability is, however, not inherently bad and the aim should not be to eliminate it at any cost. Quite to the contrary, performance variability is necessary for users to learn and for a system to develop; monitoring of variability must be able to distinguish between what is potentially useful and what is potentially harmful.

The distinction between the three classes of accident models proposed above does not imply that one is unequivocally better than the others. Although it is inadvisable to rely on a sequential accident model as the only basis for analysis and explanation, it should not be discarded outright. In cases where there are easily distinguishable causes it makes sense to try to eliminate them. Similarly, in cases where there is a multitude of contributory factors it may be better to apply preventive and protective barriers or to monitor closely the system to detect impending instabilities and coincidences. Although complexity is difficult to handle, both in theory and in practice, it should not be shunned.

II. ROLE OF HUMANS IN ACCIDENTS

Even though humans are no longer seen as the primary cause of accidents, they do play a role in how systems fail – as well as in how systems can recover from failure – simply because of the fact that they are an indispensable part of all complex systems. Whereas the focus used to be on the role of humans at the “sharp end”, i.e., during actual operations at the time and place where the accident happened, it has now become clear that humans play a role also at the blunt end – as well as everywhere in between. As

Karlene Roberts so poignantly has put it: “everybody’s blunt end is somebody else’s sharp end”. No technological system has created itself or can take care of itself, and humans are involved from the very beginning to the very end. The role of the human must therefore be considered at all levels, from the initial design to repair and maintenance – including also inspection, regulation, and dismantling in the end.

The consequence of this position is not that accident analysis should attempt to follow the antecedents back to the origin of the system – or even beyond that. This would be a misinterpretation of the sharp-end, blunt-end model, since it does not imply that there must be a causal relation between the blunt-end (as causes) and the sharp-end (as effects). The model should rather be understood as saying that actions or decisions at any level may have effects that only manifest themselves much later and in indirect ways, as described by the latent conditions that are part of the epidemiological accident model.

In trying to understand the role of humans in accidents, a first step is to acknowledge that human actions cannot be described in binary terms, i.e., as being either correct or incorrect. The correctness of actions can only be judged in hindsight, i.e., with knowledge of the outcome (Woods et al., 1994). It must be assumed that people always try to do what they think is right at the time they do it. For example, no one would for a moment suggest that the operators in Chernobyl or at Three-Mile Island tried to bring about a nuclear accident. Their actions turned out to be disastrously wrong because the operators did not understand the situation, not because they intended to do wrong. Yet the actual actions may sometimes differ from what was intended for a variety of reasons such as distraction, inadequate interface design, fatigue, lack of knowledge, work overload, etc. They may furthermore realise this either when they do it, immediately after, or sometimes later and as a consequence try to recover from the action failure. This view leads to a classification of the variety of human actions along the following lines (Amalberti, 1996).

Actions that are correctly performed, i.e., where the intended and actual outcomes correspond.

Actions where the failure is detected and successfully recovered. The recovery may either be immediate, as when we notice a typo, or at some later time – depending on the nature of the process and how forgiving the system is.

Actions where the failure is detected but tolerated. This usually happens because people believe that the consequences will be minor, or that they will be able to recover at a later time.

Actions where the failure is detected but not recovered. The failed recovery can happen because the process is not reversible, because there are insufficient resources – in particular insufficient time, because the detection comes too late, because there are no options for recovery, etc.

Finally, there are actions where the failure is not detected, for instance because the effects are latent. This is typically the case for failures during maintenance. In these cases the actual and intended outcomes will evidently not correspond.

The different types of actions are illustrated in Figure 3. This figure shows that it is only if actions are classified in terms of their outcome (correct, incorrect) that they themselves can be seen as right or wrong. Doing so will, however, be a gross oversimplification, which not only disregards the nuances between different types of actions but also makes it impossible to develop

effective responses. It is only by knowing what people did and why they did it that effective solutions can be devised.

A. Action Types And Accident Models

Since humans play an important role in how accidents occur, it is appropriate to consider how the variety of human actions corresponds to the different accident models. In other words, what are the assumptions about human actions implied by each model?

The sequential accident model basically assumes that a chain of events causes the accident, as illustrated by the First Axiom of Industrial Safety quoted above. Each event in this chain is considered as either being done correctly or as having failed, and this goes for human actions as well as anything else (cf. Barnes et al., 2002). There is therefore no room for the multi-faceted view of human actions that Amalberti (1996) has proposed. This is in itself sufficient reason for not basing an accident analysis primarily on the sequential model. The sequential accident model furthermore makes it difficult to consider more complex descriptions of accidents, such as the detection-recovery functions (Kanse & van der Schaaf, 2000), which require that actions can be undone, i.e., that the consequences of a possibly incorrect action are not immutable.

In the epidemiological models, human actions – or unsafe acts – at the sharp end are the triggers but not the causes of accidents. The accidents come about because of the “unforeseeable concatenation” of unsafe actions and latent conditions, the latter themselves being the outcome of actions removed in space or time – i.e., at the blunt end. These conditions have many origins: regulatory narrowness, incomplete procedures, mixed messages, production pressures, responsibility shifting, inadequate training, attention distractions, deferred maintenance, clumsy technology, and so on. The latent conditions can come about because people tolerate faulty actions or are unable adequately to recover, as well as for other reasons. There is therefore no conflict between the epidemiological model and the concept of varied human actions, although the model does not directly account for how they can come about.

The systemic model is based on the notion that human actions are variable, and that the variability – rather than human failures – is the central issue. It is therefore important to understand how and why human actions vary, not just to explain how accidents can occur but also to understand how safety and efficiency come about. Karl Weick made the point that “safety is a dynamic non-event”. This means that safety is the absence of incidents and accidents, which in turn depends on the dynamic characteristics of the system. It is the way in which the system behaves, including the people in the system, which creates safety and accidents alike. The classification of the variety of human actions must therefore be supplemented by a model or description of how they occur. This is outlined below.

B. The Efficiency-Thoroughness Trade-Off (ETTO) Principle

Human performance must always satisfy multiple, changing, and often conflicting criteria. Humans are usually able to cope with this imposed complexity because they can adjust what they do and how they do it to match the current conditions. One way of looking at this is by noting how people constantly try to optimise their performance by making a trade-off between efficiency and thoroughness. On the one hand people genuinely try to meet

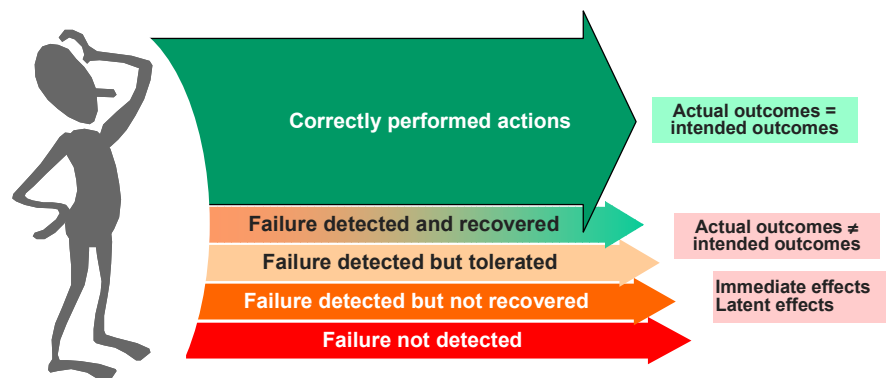


Figure 3: Varieties of human actions.

task demands and to be as thorough as they believe is necessary. On the other hand they try to do this as efficiently as possible, which means that they avoid spending unnecessary effort or wasting time.

In making this trade-off people are greatly helped by the regularity or stability of their work environment and, indeed, the regularity of the world at large. If work environments were continually changing they would lack the predictability that makes it possible to take shortcuts and learn how things can be done in a more efficient manner. Yet it is precisely because work environments – by design or by nature – have some measure of regularity or stability that they are predictable, and therefore allow performance to be optimised.

The benefits of making shortcuts are obvious: instead of checking every possible condition or prerequisite of an action, efforts can be saved to check conditions that are known to vary across situations, or conditions that are seen as being more important. In the case of RO-RO ferries, for instance, if the bow port always is closed when the ferry leaves harbour, then there is no need explicitly to verify this condition. And the bow port is always closed because regulations say that it should be so. Or, to take another example, if a hospital laboratory has routines to ensure that the right type of blood is issued, then it is only necessary to check that the identification of the patient is correct. The nurse has to bring the blood to the right patient, but need not check whether the blood is of the right type.

Human performance is efficient because people quickly learn to disregard those aspects or conditions that normally are insignificant. This adjustment is furthermore not only a convenient ploy for the individual, but also a necessary condition for the joint system (i.e., people and technology seen together) as a whole. Just as individuals adjust their performance to avoid unnecessary efforts, so does the joint system. This creates a functional entanglement, which is essential for understanding why failures occur. The performance adjustment on the joint system level cannot be effective unless the aggregated effects of what individuals do are relatively stable, since this constitutes an important part of the joint system's environment. On the other hand, the efficient performance of the joint system contributes in a significant manner to the regularity of the work environment for the individuals, which is a pre-condition for their performance adjustment.

As far as the level of individual human performance is concerned, the local optimisation – through shortcuts, heuristics, and expectation-driven actions – is the norm rather than the exception. Indeed, normal performance is not that which is prescribed by rules and regulation but rather that which takes place as a result of the adjustments, i.e., the equilibrium that reflects the regularity of the work environment. This means that we cannot find the cause of failures in the normal actions since they, by definition, are not wrong. This is consistent with the view of complexity theory according to which some properties of the system cannot be attributed to individual components but rather emerge from the whole system.

The conclusion is that both normal performance and failures are emergent phenomena, hence that neither can be attributed to or explained by specific components or parts. For the humans in the system this means in particular that the reason why they sometimes fail, in the sense that the outcome of their actions differ from what was intended or required, is due to the variability of the context and conditions rather than to action failures. The adaptability and flexibility of human work is the reason for its efficiency. At the same time it is also the reason for the failures that occur, although it is never the cause of the failures. Herein lies the paradox of optimal performance at the individual level. If anything is unreasonable, it is the requirement to be both efficient and thorough at the same time – or rather to be thorough when with hindsight it was wrong to be efficient.

III. REFERENCES

- Amalberti, R. (1996). *La conduite des systèmes à risques*, Paris: PUF.
- Barnes, V. E., Haagensen, B. C. & O'Hara, J. M. (2002). The human performance evaluation process: A resource for reviewing the identification and resolution of human performance problems (NUREG/CR-6751). Washington, DC: U.S. Nuclear Regulatory Commission Office of Nuclear Regulatory Research
- Dörner, D. (1980). On the difficulties people have in dealing with complexity. *Simulation & Games*, 11(1), 87-106.
- Green, A. E. (1988). Human factors in industrial risk assessment - some early work. In L. P. Goodstein, H. B. Andersen & S. E. Olsen (Eds.), *Task, errors and mental models*. London: Taylor & Francis.
- Heinrich, H. W., Petersen, D. & Roos, N. (1980). *Industrial Accident Prevention (Fifth Edition)*. New York: McGraw-Hill Book Company.
- Kanse, L. & van der Schaaf, T. W. (2000). Recovery from failures – Understanding the positive role of human operators during incidents. In: D. De Waard, C. Weikert, J. Hoonhout, & J. Ramaekers (Eds.): *Proceedings Human Factors and Ergonomics Society Europe Chapter Annual Meeting 2000*, Maastricht, Netherlands, November 1-3, 2000, p. 367-379.
- Perrow, C. (1986). *Complex organizations: A critical essay* (3rd ed.). New York: Random House.
- Reason, J. T. (1987). The Chernobyl errors. *Bulletin of the British Psychological Society*, 40, 210-216.
- Reason, J. T. (1990). *Human error*. Cambridge, U.K.: Cambridge University Press.
- Reason, J. T. (1997). *Managing the risks of organizational accidents*. Aldershot, UK: Ashgate.
- Sheridan, T. B. (1992). *Telerobotics, automation, and human supervisory control*. Cambridge, MA: MIT Press.
- Svenson, O. (1991). The accident evolution and barrier function (AEB) model applied to incident analysis in the processing industries. *Risk Analysis*, 11(3), 499-507.
- Svenson, O. (2001). Accident and Incident Analysis Based on the Accident Evolution and Barrier Function (AEB) Model. *Cognition, Technology & Work*, 3(1), 42-52.
- Woods, D. D., Johannesen, L. J., Cook, R. I. & Sarter, N. B. (1994). *Behind human error: Cognitive systems, computers and hindsight*. Columbus, Ohio: CSERIAC.