Archetypes for Organisational Safety

Karen Marais and Nancy G. Leveson; MIT Department of Aeronautics and Astronautics;
Cambridge, Massachusetts, U.S.A.

Keywords: organisational safety, system dynamics, archetypes

Abstract

We propose a framework using system dynamics to model the dynamic behaviour of organisations in accident analysis. Most current accident analysis techniques are event-based and do not adequately capture the dynamic complexity and non-linear interactions that characterize accidents in complex systems. In this paper we propose a set of system safety archetypes that model common safety culture flaws in organizations, i.e., the dynamic behaviour of organizations that often leads to accidents. As accident analysis and investigation tools, the archetypes can be used to develop dynamic models that describe the systemic and organizational factors contributing to the accident. The archetypes help clarify why safety-related decisions do not always result in the desired behaviour, and how independent decisions in different parts of the organisation can combine to impact safety.

Introduction

Modern socio-technical systems are becoming more complex and tightly coupled in response to increasing performance and cost requirements. Understanding these systems and analysing or accurately predicting their behaviour is often difficult. We are seeing a growing number of normal, or system, accidents, which are caused by dysfunctional interactions between components, rather than component failures. Such accidents are particularly difficult to predict or analyse [1]. Accident models focusing on direct relationships among component failure events or human errors are unable to capture these accident mechanisms adequately.

Systems and organizations continually experience change as adaptations are made in response to local pressures and short-term performance goals (e.g. productivity and cost). People adapt to their environment or they change their environment to better suit their purposes. Several decision makers at different times, in different parts of the company or organization, all striving locally to optimise performance may be preparing the stage for an accident, as illustrated by the 1987 Zeebrugge ferry disaster [2] and the Black Hawk friendly fire accident [3]. Safety defences therefore tend to degenerate systematically over time. When a larger view is taken, most accidents in complex systems can be seen to result from a migration to states of increasing risk over time. Once a system has migrated to an unsafe state, accidents are inevitable unless appropriate efforts are made to bring the system to a safe state. The Bhopal accident is a classic example.

One of the worst industrial accidents in history occurred in December 1984 at the Union Carbide chemical plant in Bhopal, India [4]. The accidental release of methyl isocyanate (MIC) resulted in at least 2000 fatalities, 10 000 permanent disabilities (including blindness), and 200 000 injuries. The Indian government blamed the accident on human error in the form of improperly performed maintenance activities. Using event-based accident models, numerous additional factors involved in the accident can be identified. But such models miss the fact that the plant had been moving over a period of many years toward a state of high-risk where almost any change in usual behaviour could lead to an accident.

If we wish to better understand past accidents and prevent future accidents, we need to look at how systems migrate towards states of increasing risk. Such understanding requires taking a long-term dynamic view of the system, and not just considering the proximate events, i.e., those events immediately preceding the actual loss event. System dynamics modelling is one way to describe dynamic change in systems. We have found it useful in understanding accidents, as argued in a companion paper, where we demonstrate its use in understanding the Walkerton *E. coli* outbreak [5]. But building system dynamics models is difficult for non-experts and usually achieved in an *ad hoc* and time-consuming manner. In developing the Walkerton system dynamics model, we found that a lot of time was needed to identify the variables of interest and to determine which relations between these variables were relevant to the accident. One way to accelerate and focus the modelling process is to start by applying archetypes that describe typical behaviour and flaws in the safety culture that have often been involved in accidents.

In this paper we propose a preliminary set of safety archetypes. The safety culture of an organization can be usefully described in terms of these safety archetypes. In accident analysis the archetypes can be used to identify and highlight change processes and the flawed decision-making that allowed the system to migrate towards an accident state. The archetypes will also form part of a new risk assessment method under development by the authors, where they will be used both as diagnostic and as prospective tools. As diagnostic or analytic tools they can be used to identify the structures underlying undesired behaviour. As synthesis tools they can be used to examine the potential undesired consequences of decisions.

This paper is organised as follows: We begin with a brief overview of system dynamics and its building blocks. Next, we propose a preliminary set of safety archetypes. In each case, illustrative examples of the archetype's application to safety are given.

System Dynamics

System dynamics is an approach to identifying, explaining, and eliminating problem behaviours in socio-economic systems, primarily by identifying feedback loops in the system. It provides a framework for dealing with dynamic complexity, where cause and effect are not obviously related. System dynamics is grounded in the theory of non-linear dynamics and feedback control, but also draws on cognitive and social psychology, organisation theory, economics, and other social sciences. For an extensive discussion, see [6].

Organisational Behaviour and the System Archetypes: System dynamics posits that the behaviour of a system arises from its structure. The structure is described in terms of feedback (causal) loops, stocks (levels) and flows (rates), and non-linearities created by interactions between system elements. In the system dynamics view, all dynamics (behaviour over time) can be explained by the interaction of the two basic types of feedback loops: positive and negative. Positive feedback loops are self-reinforcing, and are therefore referred to as *reinforcing* loops. Negative feedback loops tend to counteract change, and are therefore referred to as *balancing* loops. Engineers use negative feedback to stabilise systems in the presence of uncertainty.

System dynamics research has shown that in the case of socio-economic systems at least, many patterns of behaviour are generated by a small set of simplified 'generic structures'. Various classifications have been made [7]; see for example, generic infrastructures [8]. System archetypes are causal loop representations of generic patterns of behaviour over time, and are particularly useful for illustrating counter-intuitive behaviour [9]. Like all models, system archetypes are merely *approximations* of systems and their behaviour. Their value arises from the

compelling way in which they convey system insights [10]. We believe, in particular, that they provide important insights into accident causation in socio-technical systems.

Building Blocks: System dynamics models are built from three building blocks: the reinforcing loop, the balancing loop, and the delay.

A *Reinforcing Loop* is a structure that feeds on itself to produce growth or decline (positive feedback). An increase in *Variable 1* leads to an increase in *Variable 2*, as indicated by the '+' sign, which in turn leads to an increase in *Variable 1*, and so on. In the absence of external influences, both *Variable 1* and *Variable 2* will grow or decline exponentially. A characteristic of exponential growth is that the doubling time is constant. Because the initial growth is slow, it may be unnoticed until it becomes rapid, at which point it may be too late to control the growth. Reinforcing loops "generate growth, amplify deviations, and reinforce change" [6].

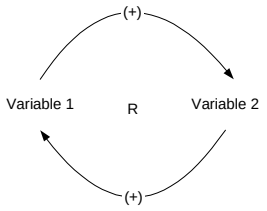

Figure 1 – Causal Loop Diagram of the Reinforcing Loop

A *Balancing Loop* is a structure that attempts to move a *variable* value to a *desired or reference value* through some *action* (negative feedback). The difference between the current state and the desired state is perceived as an *error*. An action proportional to the error is taken to decrease the error, so that, over time, the current state approaches the desired state. While the reinforcing loop tends to display exponential growth or decline, the balancing loop tends to settle down to the desired state. Because the size of the remedial action is proportional to the size of the error, the current state initially rapidly approaches the desired state. As the error decreases, the rate at which the current state approaches the desired state decreases.

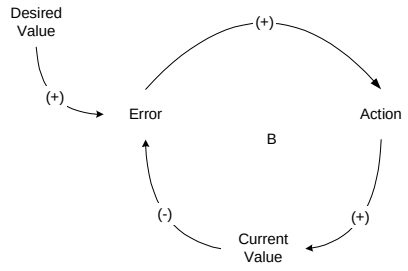

Figure 2 – Causal Loop Diagram of the Balancing Loop

Delays are used to model the time that elapses between cause and effect, and are indicated by a double line (Figure 3). Delays make it difficult to link cause and effect (dynamic complexity) and may result in unstable system behaviour. Consider the problem of navigating a ship down a narrow channel. Suppose that the ship is veering to one side of the channel, and the helmsman wishes to correct the course. Due to the ship's inertia, adjusting the rudder will not result in an immediate course change. There is a *delay* between a change in the rudder position and the resulting course change. In stressful situations, even experienced helmsmen may interpret a

delayed response as a complete lack of response, and accordingly make a larger change in the rudder position. When the ship's inertia is eventually overcome, the helmsman finds himself sailing towards the opposite side of the channel. If the helmsman continues to over-correct in this way, the ship will veer wildly from one side of the channel to the other, and may run aground.

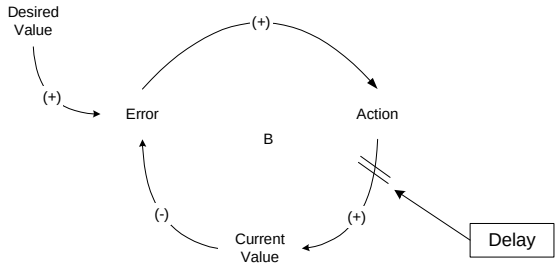

Figure 3 – Causal Loop Diagram of the Balancing Loop with Delay

Toward a Set of Safety Archetypes

In this section we propose a preliminary set of safety archetypes. General system behavioural archetypes have been described by Braun [11] and Wolstenholme [7]. While the general archetypes apply to all behaviour, our safety archetypes address specific behaviour related to flaws in an organization's safety culture. These safety archetypes can assist in representing and understanding the dynamic forces behind accidents and help accident investigators in their search for causal factors.

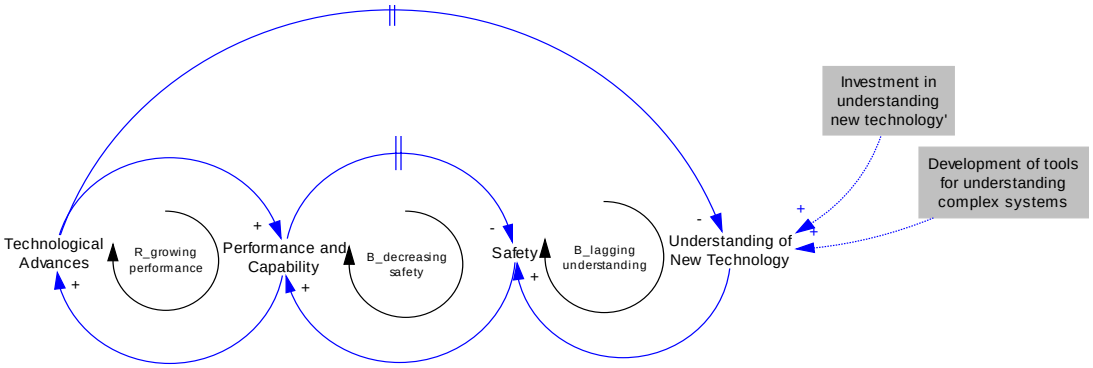

Figure 4 – Stagnant Safety Practices in the Face of Technological Advances

Stagnant Safety Practices in the Face of Technological Advances: This structure consists of a reinforcing loop ($R_{growing\ performance}$) and a balancing loop ($B_{decreasing\ safety}$). The reinforcing loop consists of some action in one part of an organisation, intended to achieve some outcome. Initially, the action is successful ($R_{growing\ performance}$), but after a time a constraint on performance is reached and the system reacts to limit the outcome ($B_{decreasing\ safety}$).

Here the constraint on safety is our understanding of new technology and the systems in which it is embedded. Technological advances result in an increase in performance in many areas, which in turn drives more advances ($R_{growing\ performance}$). As the speed of change accelerates, understanding of the safety implications lags further behind ($B_{lagging\ understanding}$). A characteristic feature of modern systems is that their complexity often exceeds our grasp. For example, it is alarmingly easy to write software whose behaviour cannot be predicted under all circumstances. This lack of

understanding translates into a decrease in safety ($B_{decreasing\ safety}$). We can ameliorate the problem by investing more resources in our understanding of new technologies, and by developing tools for understanding complex systems.

Decreasing Safety Consciousness: The success of a safety program may be limited by the characteristics of the system to which the program is applied, or by the nature of the program itself. A strategy, policy, or process that initially promotes improved safety may eventually reach a point where its continued application may cause a decline in safety.

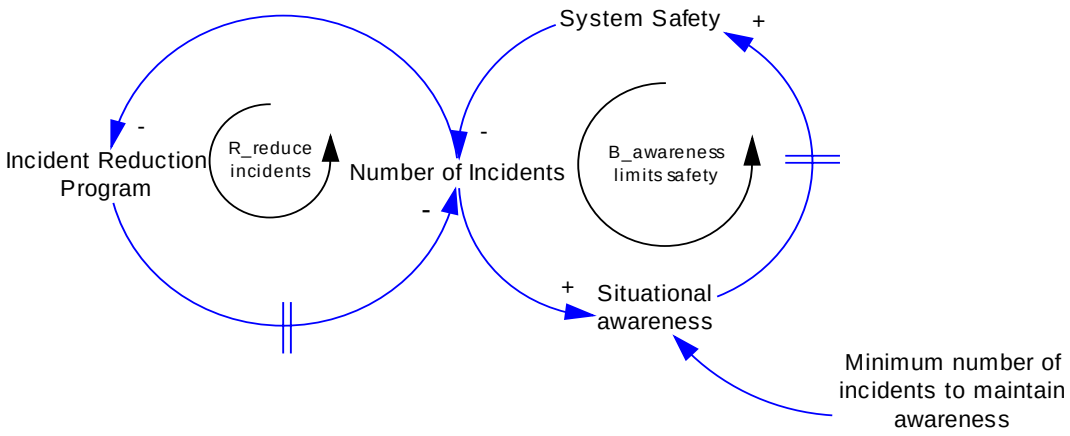

Figure 5 – *Decreasing Safety Consciousness*—Incident reduction measures initially improve system safety ($R_{reduce\ incidents}$). But the absence of incidents renders the system mute, and situational awareness of safety is decreased. The result is a decrease in system safety ($B_{awareness\ limits\ safety}$).

Consider the case of ultra-safe systems[1]. Common sense tells us that in order to increase safety, errors, incidents and breakdowns must be reduced or eliminated. This is true for systems where the rate of incidents and accidents is high. In the case of ultra-safe systems, continued elimination of errors, incidents, and breakdowns may paradoxically decrease safety [12].

Continued optimisation of a given set of safety measures, does not necessarily increase safety further. Over-optimisation numbs the adaptive capabilities of human and technical systems, while covering up minor system failures. In the case of error reduction, for example, it has been found that error plays an ecological role in the control of performance, and that detected errors are necessary to maintain situational awareness. Similarly, programs to reduce the number of incidents and breakdowns may also perversely decrease safety. As the perceived level of safety increases, investments are redirected from safety measures to improving system performance. Over-stretched system performance leads to new risks, which may materialise in the form of disastrous accidents [2]. Beyond a certain incident reduction quota, the absence of incidents, as opposed to the presence of a minimum number of incidents, does not prevent accidents from occurring. It may be necessary to tolerate a certain level of errors, incidents, breakdowns, and even accidents to protect the system against disastrous accidents. Figure 5 illustrates how incident reduction programs can result in a decrease in system safety.

[1] Amalberti defines ultra-safe systems as those where the risk of disaster is below one accident per 10^7 events [12]. For this discussion, it is sufficient to define high-risk systems as those that are not ultra-safe.

Amalberti argues that the combination of a system with a given set of safety measures bears within itself a maximum safety potential, which cannot be exceeded by continued optimisation of those safety measures. Continued optimisation of a particular safety measure 'mutes' some system aspects, thereby decreasing system awareness and adversely affecting safety. To obtain further increases in safety beyond this limit, additional, new safety measures are necessary. Therefore, to maintain safety, safety measures must be aggregated, but no single safety measure should be overly optimised.

Consider the strong emphasis on redundancy as a safety and reliability measure in many systems. Some degree of redundancy is useful in increasing reliability, and possibly safety. But more redundancy is not necessarily better, and may be worse. While redundancy may increase reliability, it does not necessarily increase and may decrease safety. First, a reliance on redundancy may lead to decreased emphasis on other safety engineering techniques. If system designers believe that redundancy will limit the effect of design errors they may be less motivated to find and eliminate these errors. In practice, redundancy may 'cover up', or mute, design errors and prevent them from becoming visible until something catastrophic occurs. Second, increasing redundancy increases system complexity. More complex systems are less amenable to testing and maintenance, and their properties and behaviour are difficult to predict accurately [13].

For example, an Air Force system included a relief valve to be opened by the operator to protect against over-pressurisation [4]. A secondary valve was installed as backup in case the primary relief valve failed. The operator had to know when the primary valve had not opened in order to determine that the secondary valve had to be opened. One day, the operator issued a command to open the primary valve. The position indicator and open indicator lights both illuminated although the primary relief valve had not opened. The operator, thinking that the primary valve had opened, did not activate the secondary valve and an explosion occurred. A post-accident investigation discovered that the indicator light circuit was wired to indicate only the presence of power at the valve, and not the actual valve position. The indicator showed only that the activation button had been pushed, not that the valve had opened. Redundancy could not provide protection against the underlying design error. Worse, the overconfidence provided by the redundancy convinced the engineers that an examination of the wiring design was not needed and the design error was therefore not found.

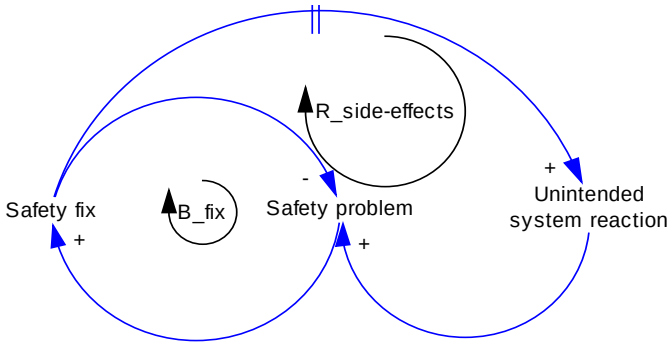

Figure 6 – Unintended Side Effects of Safety Fixes

Unintended Side Effects of Safety Fixes: The unintended consequences of poorly designed responses to safety problems, whether they are symptomatic treatments or supposed fundamental solutions, can worsen the problem.

This structure consists of a balancing loop (B_{fix}) and a reinforcing loop ($R_{side\text{-}effects}$). The loops interact so that the desired result initially produced by the safety fix in the balancing loop is, after some delay, offset by the undesired side effects in the reinforcing loop. Initially, the *Safety fix* ameliorates the *Safety problem* (B_{fix}). After a delay, the *Unintended system reaction* becomes visible ($R_{side\text{-}effects}$). Undesired aspects of the system reaction worsen the problem, and accordingly the safety fix is applied more strongly ($R_{side\text{-}effects}$). The safety fix ironically contributes to the worsening of the problem.

Well-intentioned, commonplace solutions to safety problems often fail to help, have unintended side effects, or exacerbate problems. The example below illustrates how disciplining workers and writing more detailed procedures may fail to reduce the number of equipment breakdowns.

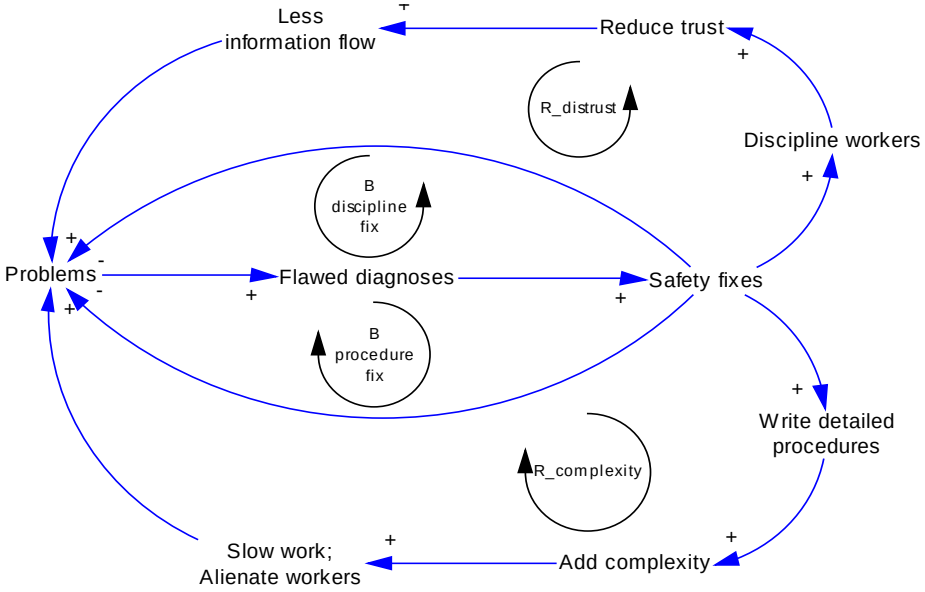

Figure 7 – The common response to incidents or accidents of disciplining workers or writing more detailed procedures is intended to solve the problems (balancing loops $B_{discipline\ fix}$ and $B_{procedure\ fix}$). But these fixes often result in reinforcing loops ($R_{distrust}$ and $R_{complexity}$) that eventually make the problems worse.

Consider a plant that is experiencing increasing equipment breakdowns, which are attributed to poor maintenance. A typical 'fix' for maintenance-related problems is to write more detailed maintenance procedures and to monitor compliance with these procedures more closely. More detailed procedures can translate to fewer errors in a particular task. But workers also tend to view more detailed procedures and closer supervision as mistrust and regimentation, causing them to lose motivation, or comply blindly or maliciously with procedures that may be incomplete or incorrect. Skilled workers may find the new regime intrusive and look for more interesting work elsewhere. Excessive restrictions on behaviour discourages problem solving and encourages blind adherence to procedures, even when such compliance is not optimal in terms of safety or productivity. Blaming or disciplining individual workers, designed to create an atmosphere of accountability, encourages all workers to hide problems. For example, when the Federal Aviation Administration provided immunity from prosecution to pilots who reported near-collisions, the number of reports tripled; when immunity was later retracted, the number of reports decreased six-fold [14]. When incidents are deliberately concealed, the underlying problems do not become visible, often worsen, and may lead to more problems (Figure 7).

Unintended Side Effects behaviour occurs when the fundamental problem is not understood, or when the solutions to the fundamental problem are not appropriate or are improperly implemented. We can avoid or escape this behaviour by correctly identifying the fundamental problem and designing appropriate solution strategies. Identifying the fundamental problem is often difficult, and designing and implementing solution strategies can be challenging. An awareness of the long-term negative implications that fixes often have can provide the impetus to search for fundamental solutions instead.

Fixing Symptoms Rather Than Root Causes: In this archetype, *Symptomatic solutions* are implemented in response to *Problem Symptoms* ($B_{symptoms}$), temporarily decreasing the symptoms (Figure 8). If the *Fundamental Solution* is known, *Side Effects* of the symptomatic solutions may either decrease the desire to implement the fundamental solution, or act to decrease the effectiveness of the fundamental solution ($R_{side\ effects}$). Alternatively, if the fundamental solution is not known, the symptomatic solutions may decrease the ability to find the fundamental solution, for example by masking the problem symptoms.

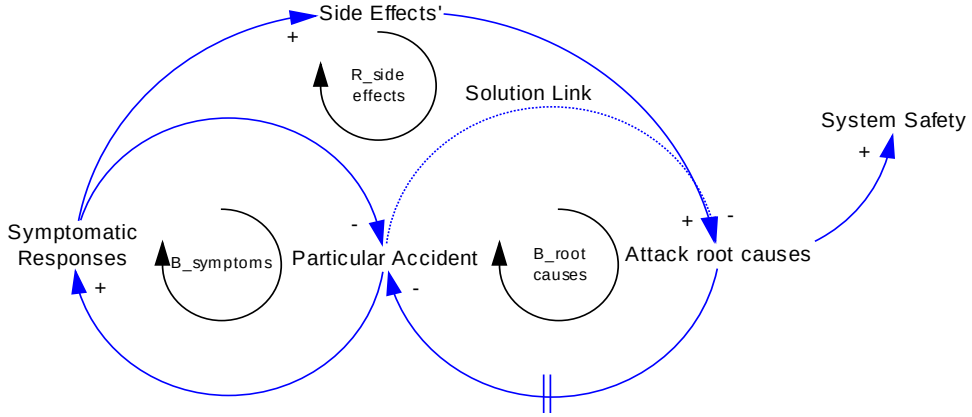

Figure 8 – Fixing Symptoms: Symptomatic solutions decrease the likelihood of recurrence of the same accident ($B_{symptoms}$), but do not address the underlying conditions that allowed the accident to occur in the first place. A side effect is that the impetus to find fundamental solutions is decreased ($R_{side\ effects}$). Organisations should instead perform root cause analysis and use the resulting insights to formulate fundamental solutions that address the underlying systemic causal factors ($B_{root\ causes}$).

Fixing Symptoms illustrates the tension between the appeal of short-term, symptomatic solutions, and the long-term impact of fundamental solutions. Symptomatic solutions are usually easier, faster, and cheaper to implement than long-term fundamental solutions. Initially, positive results to symptomatic solutions are seen immediately, as the visible symptoms are eliminated. Once a symptomatic solution has been successfully applied, the pressure to find and implement a fundamental solution tends to decrease. Over time, the solutions may become less effective, or different symptoms of the underlying problem may arise; in response new symptomatic solutions are devised. The underlying problem remains. If the fundamental problem is not dealt with, symptoms can be expected to continue surfacing in various forms. Long-term, fundamental solutions, on the other hand, may be more difficult to devise, more difficult to implement, take longer to show results, and are often initially more costly. At the same time, external pressures often demand a 'quick-fix' to the problem.

The reactive focus of many safety programs results in placing primary emphasis on investigating previous incidents and accidents in an attempt to prevent future accidents. These efforts are not

always fruitful. Excessive focus is placed on preventing recurrence of exactly the same accident, without taking sufficient account of the underlying systemic factors that allowed safety to deteriorate [4]. Attempts to identify the deeper factors or conditions that allowed the accident to occur (i.e., root cause analysis) are often insufficient.

For example, Carroll et al. have identified instances of inadequate root cause analysis at nuclear plants [15]. In the nuclear and chemical industries, problem investigation teams are assigned to examine serious incidents and troubling trends. These investigations are part of corrective action programs to improve safety and performance. Although considerable resources are devoted to these programs, the investigations do not always result in effective learning. The investigations studied by the authors tended to focus on only a few proximal causes. These causes were typically technical or involved human error, and their solutions were obvious, easily implemented, and acceptable to powerful stakeholders. Little effort was made to uncover root causes or devise fundamental solutions.

Symptomatic solutions to accidents often only decrease the likelihood of that particular accident recurring. They do not eliminate the deeper structural deficiencies that led to the accident in the first place. Once a symptomatic solution has been successfully applied, the perceived need to solve the underlying structural problem may disappear, reducing the pressure to find a fundamental solution. To improve safety in the long term the fundamental problem or structural deficiency that is causing the symptoms must be identified.

For example, if an aircraft rudder failure is shown to be the result of insufficient or poor maintenance, the recommended action may be to improve the rudder maintenance procedures. But deeper problems, such as subtle management pressure to increase maintenance throughput, may have caused the maintenance to be poorly performed in the first place.

Identifying the root causes of incidents and accidents is not always easy to do. Symptomatic solutions may be suppressing the symptoms, creating the illusion that no problem exists. These solutions may be consciously or unconsciously formulated and applied. Unconsciously applied solutions (e.g. unconsciously correcting for misaligned steering on a motor vehicle) may so successfully mask the underlying problem that operators are not aware of the problem symptoms, let alone the fundamental problem. In order to understand the symptoms of the problem, it is necessary to identify the conscious and unconscious symptomatic solutions. Because any individual only has a limited view of the system, obtaining different viewpoints of the symptoms, the problem, and the system can help in identifying the fundamental problem.

Eliminating root causes is likely to be more difficult, time-consuming, and costly to implement than implementing symptomatic solutions. It is essential to obtain commitment from all parties involved with the implementation of the proposed solution. Without such commitment, the solution is unlikely to be successfully applied. Side effects of the solution must be identified as far as possible. Of course it may not be possible to foresee all the side effects. Awareness of the potential for side effects makes it easier to identify and deal with them if they do occur. Where side effects of symptomatic solutions may undermine the fundamental solution, it is necessary to stop applying these solutions before applying the fundamental solution.

Eroding Safety: This archetype illustrates how safety goals may erode or become subverted over time. We can expect to observe *Eroding Safety* behaviour in systems where an accident was preceded by a declining emphasis on safety, such as decreasing safety goals. This decline is an example of migration towards unsafe behaviour. *Eroding Safety* is often difficult to observe while it is occurring because change tends to happen gradually. At short time scales, changes may be

imperceptible. It is only after an accident has occurred that the extent of change is noticed, if at all. The first example illustrates how complacency can grow in an organisation with a history of safe operation. The second example illustrates why well-designed safety programs do not always achieve their goals.

Complacency: A history of safe operations often results in growing complacency. Figure 9 illustrates how complacency can arise. Consider a system that initially operates with a high accident rate. In order to bring the accident rate down, the system is closely monitored, possibly both internally (company rules and procedures) and externally (government regulation). Close oversight eventually decreases the accident rate, and may bring it to the point where people do not believe that accidents can or will occur. In the apparent absence of a threat to safety, oversight may seem draconian and unnecessarily costly. Coupled with budgetary pressures, this anti-regulation sentiment creates pressure to decrease oversight. Decreased oversight is manifested on the one hand by less training and fewer or less strict certification requirements, and on the other hand by decreased inspection and monitoring. A decrease in these activities eventually leads to an increase in the risk of accidents, and so the accident rate increases.

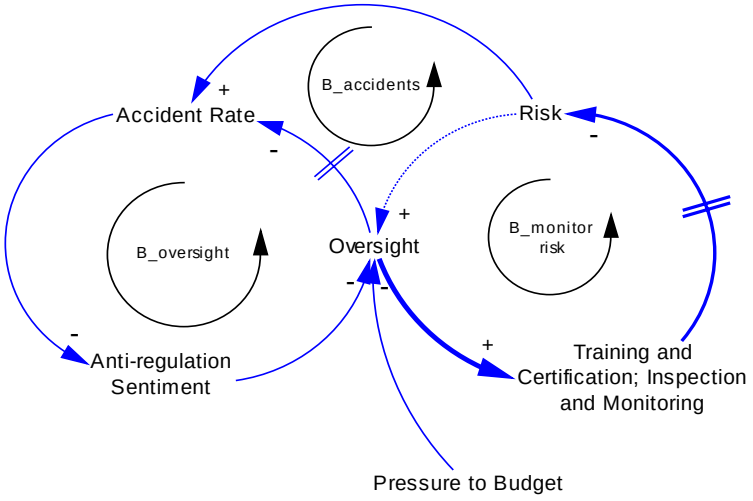

Figure 9 – Complacency occurs when low accident rates encourage anti-regulation sentiment, which coupled with budget pressures leads to less oversight ($B_{oversight}$). But decreasing oversight means decreased training and certification, and decreased inspection and monitoring, which in turn increases the risk of accidents and hence the accident rate ($B_{accidents}$). One way to avoid the complacency trap is to continuously monitor risk and set the level of oversight accordingly ($B_{monitor\ risk}$).

Following the Apollo launch pad fire in 1967, NASA established one of the best system safety programs of the time [16]. But nearly two decades later the Rogers Commission report on the Challenger accident referred to a 'Silent Safety Program' that had lost some of its effectiveness since Apollo. In particular, the report cited growing complacency at the agency, as the perception grew that Shuttle operations were routine (emphasis added) [17]:

> Following successful completion of the orbital flight test phase of the Shuttle program, the system was declared to be operational. Subsequently, several safety, reliability and quality assurance organizations found themselves with reduced and/or reorganized functional capability… The apparent reason for such actions was *a perception that less safety, reliability and quality assurance activity would be required during 'routine' Shuttle operations*. This reasoning was faulty. The machinery is highly complex, and the requirements are exacting… As the system

matures and the experience changes, careful tracking will be required to prevent premature failures… *Complacency and failures in supervision and reporting seriously aggravate these risks.*

The problem with complacency is twofold. First, it is difficult not to become complacent when success follows upon success. Second, it is difficult for an organisation to realise that it is becoming complacent, and often a serious accident is required to shake the complacency.

Organisations can avoid sinking into complacency by continuously monitoring risk and setting the level of oversight accordingly, as shown by the dotted line in Figure 9. Complacency arises because the accident rate usually does not immediately increase following a decrease in oversight. Inertia in the system temporarily keeps the accident risk at a low level, creating the impression that oversight is set at the appropriate level. All the while, the system is migrating towards the boundary of safe behaviour [2]. When accidents start occurring, the link to decreased oversight is not immediately obvious. When making the connection between risk and the level of oversight, the long-term trend in the risk level must be considered, rather than short-term fluctuations.

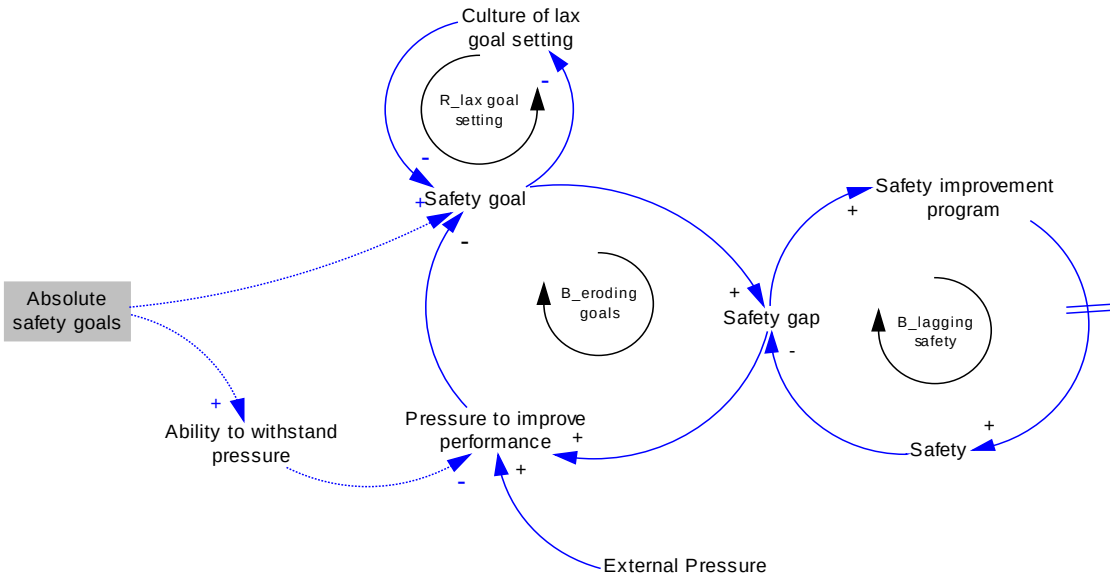

Figure 10 – Safety improvement programs typically do not show immediate results ($B_{lagging\ safety}$). External pressure results in a decreasing emphasis on safety and a lowering of the safety goals ($B_{safety\ emphasis}$ and $B_{eroding\ goals}$). These balancing loops interact to repeatedly lower the safety goal. Repeated lowering of the goal results in a reinforcing dynamic ($R_{lax\ goal\ setting}$) that encourages lax goal setting in the future. The problem can be addressed by setting absolute safety goals, perhaps based on some external standard.

Disappointing Safety Programs: *Eroding Safety* illustrates why safety programs do not always live up to their expectations (Figure 10). Safety improvement programs can be expensive and often do not show immediate results. While the eventual costs of not improving safety can be high, the immediate cost of a safety program is subject to external pressures (e.g. management pressure for performance improvement). The combination of seeming ineffectiveness and external pressures makes it tempting to adjust the goals of the safety program. This adjustment is not necessarily seen as a failure, and may even be viewed as an improvement.

For example, a common response to failed programs is to restructure parts of, or the entire organisation in question. After the Challenger accident NASA responded by reorganising the

safety and quality programs at NASA Headquarters and the field centres. A new office of Safety, Reliability, Maintainability and Quality Assurance (SRM&QA) was established and overall management of the safety function was elevated to the level of associate administrator, in an attempt to increase awareness of significant safety and quality issues at the highest levels of NASA management. This reorganisation was presented as one of the most significant improvements following the Challenger accident [18]. In fact, this reorganisation failed to achieve its goal over the long term, and many of the same 'silent safety program' characteristics have become evident following the Space Shuttle Columbia accident.

While restructuring and reorganisation is sometimes necessary, it does not always address the underlying problem. Another, more subtle form of downward goal adjustment is the eternally receding deadline. In this case, the goals remain the same, but the deadline for meeting the goals is continually shifted back, effectively lowering the goals.

Pressure for increased performance (e.g. delivery times, profit) can make it difficult to remain focussed on safety goals. *Eroding Safety* illustrates how these pressures can contribute to safety improvement goals not being met. The challenge is to resist external pressures that work against safety improvement programs, whether overtly or in a less obvious manner. Anchoring the safety goals to externally generated and enforced standards or deadlines can make adjustments in goals more visible or more difficult to make. For example, government regulators impose certain minimum safety standards on some industries, such as the nuclear power industry.

The safety program must provide a clear plan and a realistic timeframe for improving safety. It must provide concrete steps towards achieving the safety goal, as well as interim measures of progress. If a safety program is seen as working against performance (e.g. preventing on-time delivery of goods), there will be a reciprocal tendency to work against the program, thereby decreasing its effectiveness. Managers who pay lip service to safety programs but simultaneously demand increased performance encourage a lax attitude to safety at lower organisational levels. Only when there is buy-in at all levels of the organisation can a safety program succeed.

Incident Reporting Schemes: Consider what often happens when incident reporting schemes are implemented (Figure 11). The primary purpose of these schemes is to encourage workers to be more careful on a day-to-day basis, thus reducing the number of incidents. As an incentive to reduce the number of incidents, workers with the best safety records (as measured by fewest reported incidents) are rewarded. Rewarding workers who report the fewest number of incidents is an incentive to withhold information about small accidents and near misses. Underreporting of incidents creates the illusion that the system is becoming safer, when, in fact, it has merely been muted. Management becomes less aware of the behaviour of the system, and safety may therefore decrease. At the worker level, the original goal of increasing safety is subverted into one of reporting the fewest incidents. Ironically, the introduction of an incident reporting scheme can decrease safety, as found in a study of the California construction industry [19].

The *Eroding Safety* archetype illustrates how unforeseen side effects of safety programs can work against the success of the programs. In implementing safety programs it is essential to consider carefully what incentives or rewards will be used to ensure compliance. If symptomatic behaviour is rewarded (e.g. fewest reported incidents), it is likely that workers will find other ways to generate the same symptoms (e.g. underreporting incidents). If incentives are inappropriately formulated, compliance with the intent of the program may be lower than if no incentives were offered. This behaviour can also be observed in organisations that operate according to process certification standards. In this case the purported rewards are often not visible and employees view the requirements as impeding their normal working processes. Employees therefore obey

the letter of the process and documentation standards, but do not comply with the underlying intentions.

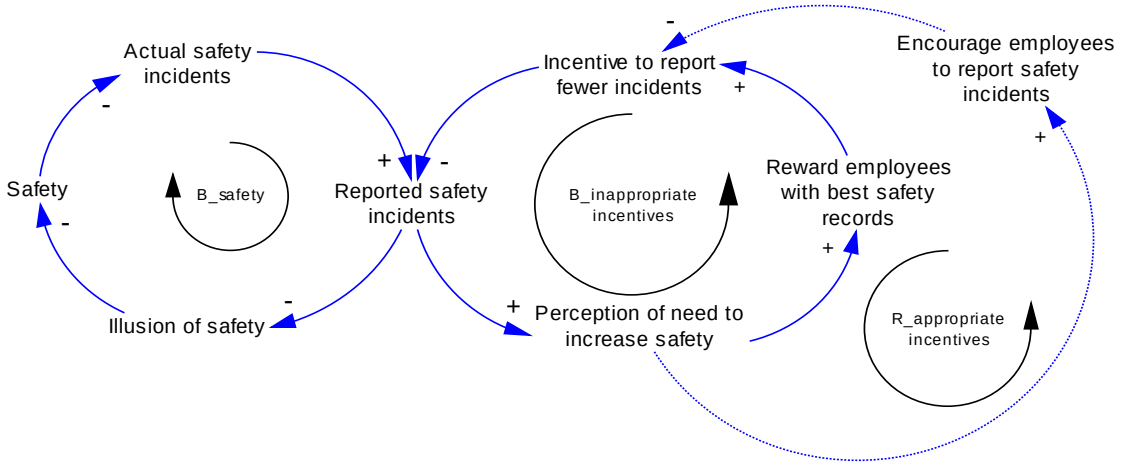

Figure 11 – An incident-reporting scheme is implemented to encourage workers to be more careful on a day-to-day basis. Workers with the best safety records are rewarded. The reward acts as an incentive to underreport incidents ($B_{inappropriate\ incentives}$). Underreporting of incidents decreases awareness of the system and creates the illusion of safety. The result is an unnoticed decrease in the system's safety (B_{safety}). One way of avoiding this type of behaviour is to encourage employees to report safety incidents, rather than rewarding employees with the best safety records ($R_{appropriate\ incentives}$).

The intent of safety programs must be communicated at all levels of the organisation. Employees must be provided with the necessary resources to perform their part in the programs. They must be empowered to make safety-based choices in cases where such decisions might adversely affect productivity. If employees understand the intent of, and are therefore committed to the program, they are more likely to comply with the intent than with the letter of the law.

Conclusions and Future Work

We have proposed a preliminary set of safety archetypes by specializing general system archetypes developed in system dynamics. The archetypes can be used to describe flaws in an organization's safety culture. In accident analysis, the archetypes can be used to model the dynamic aspects of safety-related behaviour at the organisational level. They are also useful in structuring post-investigation recommendations, by highlighting the mechanisms or root causes that led to the accident. The archetypes explain why safety-related decisions do not always result in the intended outcomes, and how independent decisions in different parts of an organisation can inadvertently interact to decrease safety.

In future work we will further develop this preliminary set of archetypes and demonstrate the application of the safety archetypes to accident analysis, by using them in the modelling of socio-technical accidents. In related work, they will be used in the creation of new approaches to risk assessment and management.

Acknowledgement

This research was supported in part by NSR ITR Grant CCR-0085829 and NASA Ames (Engineering for Complex Systems) Grant NAG2-1543.

References

[1] Perrow, Charles, Normal Accidents: Living with High-Risk Technologies, Princeton University Press, New Jersey, 1999.

[2] Rasmussen, Jens, "Risk Management in a Dynamic Society: A Modelling Problem," Safety Science, Vol. 27, No. 2, pp. 183-213, 1997.

[3] Leveson, Nancy G., Allen, Polly, Storey, Margaret-Anne, "The Analysis of a Friendly Fire Accident using a Systems Model of Accidents," *Proceedings of the 20th International System Safety Conference*, Denver Colorado, 5-9 August 2002.

[4] Leveson, Nancy G., *A New Approach to System Safety Engineering*, 2003. Available online at http://sunnyday.mit.edu.

[5] Leveson Nancy G., Daouk, Mirna, Dulac, Nicolas, and Marais, Karen, "Applying STAMP in Accident Analysis," Submitted to the *2nd Workshop on the Investigation and Reporting of Accidents*, September 2003.

[6] Sterman, John D., "System Dynamics: Systems Thinking and Modelling for a Complex World," *Proceedings of the ESD Internal Symposium*, MIT, Cambridge, MA, May 2002.

[7] Wolstenholme, Eric F., "Toward the Definition and Use of a Core Set of Archetypal Structures in System Dynamics," *System Dynamics Review*, Vol. 19, No. 1, Spring 2003, pp. 7-26.

[8] Paich, M., "Generic Structures," System Dynamics Review, Vol. 1, pp. 126-132, 1985.

[9] Senge, P. M., *The Fifth Discipline: The Art and Practice of the Learning Organisation*, Doubleday Currency, New York, 1990.

[10] Lane, David C., "Reinterpreting 'Generic Structure': Evolution, Application and Limitations of a Concept," *System Dynamics Review*, Vol. 12, pp. 87-120, 1996.

[11] Braun, William, *The System Archetypes*, 2002. Available online at: http://www.uni-klu.ac.at/~gossimit/pap/sd/wb_sysarch.pdf.

[12] Amalberti, R., "The Paradoxes of Almost Totally Safe Transportation Systems," *Safety Science*, Vol. 37, pp. 109-126, 2001.

[13] Graham, John, *Fast Reactor Safety*, Academic Press, New York, 1971.

[14] Tamuz, M., "Developing Organizational Safety Information Systems." In Apostolakis, George E., and Wu J.S. (Eds.), *Proceedings of PSAM II, Vol. 2, Los Angeles*, University of California, pp. 71: 7-12.

[15] Carroll, John S., Rudolph, Jenny W. and Hatakenaka, Sachi, "Organizational Learning from Experience in High-Hazard Industries: Problem Investigations as Off-line Reflective Practice", MIT Sloan Working Paper No. 4359-02, April 2002. Available online at: http://ssrn.com/abstract=305718

[16] Commission on Engineering and Technical Systems, *An Assessment of Space Shuttle Flight Software Development Processes*, Committee for Review of Oversight Mechanisms for Space Shuttle Flight Software Processes, Aeronautics and Space Engineering Board, National Research Council, National Academies Press, 1993. Available online at: http://www.nap.edu/

[17] Rogers, William P., Chairman, *Report of the Presidential Commission on the Space Shuttle Challenger Accident*, Government Printing Office, Washington DC, 1986.

[18] Press Release 88-01-05. Available online at: http://spacelink.nasa.gov/NASA.News/NASA.News.Releases/.index.html

[19] Levitt, Raymond E., and Henry W. Parker, "Reducing Construction Accidents - Top Management's Role," *ASCE Journal of the Construction Division*, Vol. 102, No. CO3, September 1976, pp. 465-478.

Biography

Karen Marais, MIT Department of Aeronautics and Astronautics; Cambridge, Massachusetts, U.S.A.; telephone +1.617.258.5046; e-mail karen.marais@alum.mit.edu.

Ms. Marais is a doctoral candidate at MIT in the Department of Aeronautics and Astronautics. Her research interests include systems engineering, safety, and risk assessment.

Nancy Leveson, MIT Department of Aeronautics and Astronautics; Cambridge, Massachusetts, U.S.A., telephone +1.617.258.0505; email leveson@mit.edu.

Dr. Leveson has dual appointments as a professor of Aeronautics and Astronautics and a professor of Engineering Systems at MIT. She is a member of the National Academy of Engineering. Her research interests include system engineering, system safety, human-computer interaction, and software engineering.