

# Safety-III: A Systems Approach to Safety and Resilience in Healthcare and Other Complex, Adaptive Systems

Prof. Nancy Leveson  
Aeronautics and Astronautics  
Massachusetts Institute of Technology (MIT)

**Abstract:** A holistic safety approach, based on system theoretic concepts, is used in many industries, particularly aviation and air transportation, to significantly decrease accident rates. This paper describes how that approach could reduce adverse events in the complex, adaptive healthcare system and other sociotechnical systems. The difference is shown between this new systems approach and what is done or being proposed today as represented by Safety-II and High Reliability Organization (HRO) practices.

## Introduction

There is general agreement that healthcare has an unacceptable number of adverse events. Despite concerted efforts since the IOM report *To Err is Human* in 1999 [Kohn et al., 1999], and, of course, before it, progress in reducing preventable healthcare adverse events has been relatively limited beyond a few targeted areas [Bates and Singh, 2018].

The lack of progress of healthcare in significantly reducing adverse events is not due to a lack of good intentions and effort. There have been many suggestions for improvement (such as using timeouts and checklists and creating a blame free culture) along with attempts to implement such improvements widely. New technologies have been introduced. Numerous organizations have been created that are dedicated to improving healthcare safety. Yet, while adverse events in limited areas have been reduced, the uncoordinated, piecemeal attempts have not made the significant widespread improvements called for in the IOM report [Jha 2014].

Perhaps the explanation is that, like the proverbial blind men and the elephant, much effort has been focused on different parts of the “elephant,” while not adequately addressing the system as a whole. Most of the emphasis has been placed on individuals, such as physicians and nurses, who have only limited control over hospital operations and the larger healthcare system. The 1999 IOM report suggested that the problem is not bad people in healthcare—instead the problem is that good people are working in bad systems that need to be made safer. Yet most of the attempts to improve healthcare safety since that report have not sufficiently focused on changing the system as a whole.

Engineers, in contrast, have been able to reduce the occurrence of adverse events—called accidents or mishaps in fields other than healthcare—significantly over time by creating a total system that supports the goal of reducing accidents. The reduction has been spectacular in some industries, such as aviation. This paper describes how the holistic approach used in aviation and other industries could be applied to the U.S. healthcare system and contrasts this systems approach to what is done or being proposed today. In fact, a 2014 HHS report [U.S. HHS, 2014] found that significant improvements in healthcare safety since the IOM report have resulted from redesigns of the overall healthcare system structure, such as changing incentives and payment structures, improvements in information sharing, and so on, rather than in narrowly targeted fixes.

In aviation, physical aircraft are a small part of the overall air transportation system, which, like healthcare, is an extremely complex, information- and human-intensive system that is constantly adapting and changing. Many of the most effective safety improvements in aviation have been implemented in the social and organizational parts of the air transportation system and not just the

physical parts. Improvement efforts have been successful because they have focused on changing the system as a whole.

Is engineering applicable to healthcare? Healthcare focuses on controlling natural disease processes. The identification of medical interventions to prevent or control disease is not a subject for engineering, although increasingly engineered devices are being used to implement those medical interventions. However, the *process* used to provide medical care and treat disease *is* designed and is therefore engineered. System engineering can be defined as the process of designing, integrating, and managing complex systems over their lifetime. An engineered system is simply a combination of components (including humans) created to collectively perform a useful function—such as provide healthcare.

Anything that is designed can be thought of as being *engineered*, including healthcare procedures, protocols, policies, organizational structures, physical devices, and information systems. Some examples of engineered or designed artifacts in healthcare are procedures created to reduce infection, protocols such as order sets to reduce errors of omission, the use of RFID to reduce identification errors in providing treatment, pacemakers, and electronic health records. There are hundreds of such examples. The design and maintenance of such procedures, protocols, workflows, and practices is engineering, whether it is recognized as such or not.

Ensuring that those processes do not lead to harming patients, care providers, or bystanders is the natural province of safety engineering in healthcare. In addition, these healthcare components exist within a *system context* (the overall healthcare system) that itself needs to be designed or engineered to make them useful.

Radiation therapy is an example of a component of healthcare that is starting to use modern safety engineering procedures. Beyond the actual radiation emitting device, cancer therapy involves human diagnosis and development of treatment plans as well as the design and implementation of procedures and protocols for delivering radiation therapy. Safety engineering practices are being used successfully today for these safety-critical procedures [Ghorbani 2024; Ghorbani 2024; Pawlicki and Harry 2016; Pawlicki et.al 2016; Silvis-Cividjian 2020; Wong et.al 2021].

The opposite of the systems engineering approach to safety is represented by Safety-II [Hollnagel 2014, 2017] and High Reliability Organization (HRO) theory. These contrasting approaches are used in this paper to understand and explain the standard system engineering approach to designing safety into systems and more recent extensions, here called Safety-III for convenience, that are based on Systems Theory.

## Comparing Approaches to Preventing Adverse Events

One of the recent proposals for improving safety in healthcare and all industries is what Hollnagel has labeled as Safety-II. He contrasts Safety-II with what he suggests is the prevailing approach to safety in all fields, which he labels as Safety-I. In the past 150 years, however, the Safety-I approach, which he describes as primarily retroactive and focusing on after-the-fact accident analysis, has never been the prevailing approach to safety in any industry, including healthcare. The primary emphasis has, instead, *always* been on preventing accidents or losses [Leveson 2023].

Healthcare personnel do not, for example, wait until there is a crisis to institute infection control procedures in hospitals, to introduce personal protection equipment, to create protocols and policies for administering medications, and so on. It would be considered quite foolish, and even malpractice, in most industries not to use proactive measures to prevent known or easily knowable potential for losses. Healthcare uses few scientific or formal procedures for accomplishing this goal, which system engineering could provide, but that does not change the fact that prevention is the primary goal of most everyone. Of course, it would be foolish to ignore when adverse events occur and not learn from them,

but prevention and proactive actions have always been emphasized in preventing patient harm and in preventing accidents in nearly every industry.

After mischaracterizing standard practices as Safety-I, Hollnagel describes what he calls the alternative to Safety-I, which he labels Safety-II. He argues that Safety-II is superior. In fact, neither Safety-I nor Safety-II is practiced widely. HRO theory is more widely used in healthcare, particularly by government agencies, but it is based on many of the same incorrect assumptions.

This paper describes what is *actually* done in most fields to prevent accidents—a process called Safety Engineering—and contrasts that standard approach with Safety-I and Safety-II. Also described is a fourth and much newer approach to safety based on Systems Theory—informally often labeled “systems thinking” or a “systems approach.” For convenience, the systems-theoretic approach is labeled Safety-III in this paper to avoid confusion.<sup>1</sup> In fact, aspects of Safety-III have been used for the past 70 years, but only recently are they becoming the prevailing approach used for ensuring safety in the most complex, adaptive systems being created today.

Table 1 shows a table created by Hollnagel [Hollnagel 2014] summarizing what he calls Safety-I and Safety-II. The original three columns of the table have been augmented with two extra columns (shaded to differentiate columns added to the original Hollnagel table). The fourth column describes what is actually done in safety engineering today and has been for at least the past hundred years, which was omitted from Hollnagel’s original table. The fifth column shows a more modern alternative to all of these, here labeled Safety-III. Each row in the table is discussed in the rest of this paper.

	Safety-I	Safety-II	Standard System Safety Engineering	Safety-III
<b>Definition of Safety</b>	As few things as possible go wrong	As many things as possible go right	Safety is primarily defined as freedom from unacceptable losses as identified by the stakeholders, but may be defined in terms of acceptable risk of a specific loss in some fields or countries. The goal is to eliminate, mitigate, or control <u>hazards</u> , which are the states that can lead to these losses.	Safety is defined as freedom from unacceptable losses as identified by the system stakeholders. The goal is to eliminate, mitigate, or control <u>hazards</u> , which are the states that can lead to these losses.

<sup>1</sup> The term “systems approach” has been so widely misused that it has become somewhat meaningless. Sociologists often use a *sociotechnical* approach that includes consideration of the social and organizational aspects of systems. A *systems approach* or systems-theoretic approach as used here includes all aspects of systems (physical, social, organizational) but is based on *Systems Theory*, a scientific approach to complexity and complex systems started about a hundred years ago in the early parts of the 20<sup>th</sup> Century. The term Safety-III, to describe a *systems-theoretic approach* is used in this paper simply for convenience.

<b>Safety Management Principle</b>	Reactive, respond when something happens, or is categorised [sic] as an unacceptable risk	Proactive, continuously trying to anticipate developments and events	Proactive, concentrates on preventing <u>hazards</u> and accidents but does learn from accidents, incidents, and audits of how system is performing.	Proactive, concentrates on preventing hazards and losses, but does learn from accidents, incidents, and audits of how system is performing.
<b>Explanations of accidents</b>	Accidents are caused by failures and malfunctions. The purpose of an investigation is to identify causes and contributory factors.	Things basically happen in the same way, regardless of the outcome. The purpose of an investigation is to understand how things usually go right as a basis for explaining how things go wrong.	Hazards are caused by linear chains of failure events. The purpose of investigation is to identify the chain of events and the root cause.	Hazards are caused by inadequate control over hazards. Linear causality is not assumed. There is no such thing as a root cause. The entire socio-technical system must be designed to prevent hazards. The goal of investigation is to identify why the safety control structure did not prevent the loss.
<b>Attitude to the human factor</b>	Humans are predominantly seen as a liability or a hazard.	Humans are seen as a resource necessary for system flexibility and resilience.	Humans are expected to prevent or respond to hazards and to be flexible and resourceful when they occur. Humans are left in systems even when automation is possible because they have advantages over totally automated systems. Human factors engineering deals with human error.	The system must be designed to allow humans to be flexible and resilient and to handle unexpected events.
<b>Role of performance variability</b>	Harmful, should be prevented as far as possible.	Inevitable but also useful. Should be monitored and managed.	Performance variability is only a safety concern when behavior moves outside safe boundaries. A goal is in system design to avoid hazards when performance (of operators, hardware, software, managers, etc.) varies outside safe boundaries.	Humans usually vary their performance for very good reasons. The goal is to design the system as a whole so that performance variability is not hazardous and conflicts between productivity and safety are eliminated or minimized. When performance (of operators, hardware, software, managers, etc.) varies outside safe

				boundaries, the goal is to design so that safety is still maintained.
--	--	--	--	---

### Definition of Safety

The first row of Prof. Hollnagel’s table is the following (augmented with the last two columns):

	Safety-I	Safety-II	Safety Engineering	Safety-III
<b>Definition of Safety</b>	As few things as possible go wrong	As many things as possible go right	Safety is primarily defined as freedom from unacceptable losses as identified by the stakeholders, but may be defined in terms of acceptable risk of a specific loss in some fields or countries. The goal is to eliminate, mitigate, or control hazards, which are the system states that can lead to losses.	Safety is defined as freedom from unacceptable losses as identified by the system stakeholders. The goal is to eliminate, mitigate, or control hazards, which are the states that can lead to losses. A probabilistic definition of risk is not emphasized.

### Safety-I and Safety-II

The major problem with Hollnagel’s definition of Safety-I and Safety-II is that the definitions and words used are too vague to be useful and, thus, are not used in standard safety engineering. Both science and engineering require precise definitions of the terminology used. Hollnagel writes:

*The focus of Safety-I is on things that go wrong and the corresponding efforts are to reduce the number of things that go wrong. The focus of Safety-II is on things that go right, and the corresponding efforts are to increase the number of things that go right [Hollnagel 2014, p. 177]*

An almost infinite number of things can go right or go wrong, most of which are of little importance to the system stakeholders or related at all to safety. An example of something going right is that patients enjoy their hospital meals and an example of going wrong is the opposite, that is, they do not enjoy them (perhaps because salt is reduced for cardiac patients). Another example of things going right is increasing profits by reducing the number of healthcare employees or reducing inventories. These actions are as likely (and probably more likely) to reduce safety than to improve it but they all involve “things going right.”

Every complex system has multiple goals. It is possible to increase the number of things that “go right” without having any impact on safety or even having a negative impact. It is also possible to increase the number of things that “go right” without reducing the number of things that “go wrong.”

Consider the following example where “go right” might be that a chemical plant produces its daily quota of chemicals while “go wrong” might be that toxic chemicals pollute the environment around the plant. Substituting these into the Hollnagel quote above produced the claim that “it is more important that daily quotas of chemicals are produced (goes right) than that the plant does not release toxic chemicals into the environment (does not go wrong).” They could both, in fact, happen, that is, both the number of chemicals produced and the amount of pollution may increase.

This point is critical. In Safety-II, there is an assumption that increasing the things that go right will decrease the things that go wrong and vice versa, i.e., that these are the only two results that can occur and they are a duality. Clearly, this assumption is not true. Vague, undefined terminology such as used in

Safety-II is not useful. Scientists and engineers avoid these semantic problems by using well-defined terminology.

The vagueness of the terminology used in Safety-II can make what appear to be logical arguments actually fallacious or untrue or even meaningless. Consider the statement by Hollnagel that

*It is more important—or should be more important—that things go right than that things do not go wrong* [Hollnagel 2014, p. 136].

Is it more important that passengers enjoy their flight (a thing that can go right) than that planes do not crash (a thing that can go wrong)? Is it more important that the plane arrives on time than that the passengers arrive safely?

If we use standard terminology, then it becomes apparent how incorrect or meaningless the arguments about Safety-II are. For example, let's say that in the context of safety, what "goes right" is that there is no adverse event and what goes wrong is that there *is* an adverse event. The statement above then, using simple substitution, becomes:

*"It is more important—or should be more important—that there is not an adverse event than that there is not an adverse event."*

Or more specifically, it is more important that your car brakes stop the car when necessary to avoid an accident ("go right") than that your brakes do not fail to stop the car when necessary to avoid an accident (do not "go wrong"). By using Hollnagel's vague and undefined terminology, it is possible to create nonsensical statements and truisms. Engineering and science use and define terminology more carefully.

Using vague, undefined terminology such as "goes right and goes wrong" also ignores all the important differences between various system properties, such as quality, profitability, efficiency, safety and reliability. Only by considering the complex relationships between desired system properties can we optimize the process of making necessary tradeoffs.

The most important confusion about different system properties is a common belief that safety and reliability are equivalent, particularly in the HRO (High Reliability Organization) literature. In fact, reliability and safety are very different system properties and sometimes conflict, that is, increasing reliability may lead to decreasing safety and vice versa [Dulac et.al. 2011]

As an example, taking a gun out to the desert, with nobody around for hundreds of miles, pointing the weapon away from oneself, and pulling the trigger, would be both reliable<sup>2</sup> and safe. Doing the same thing in a crowded mall would certainly not be considered safe, although the reliability of the weapon has not changed. A therapeutic medical intervention may reliably be used over years until conditions arise or something changes that makes it unsafe. The reliability in the provision of the intervention has not changed. Safety depends on context, while reliability does not. Increasing the reliability of implementing a potentially dangerous procedure can decrease safety. One of the most important changes in increasing patient safety is to eliminate the equating of reliability and safety. Most major accidents occur when the system is operating with high reliability.

Repeatability is also sometimes confused with reliability and safety. In quality control, such as on assembly lines in manufacturing, an important goal is repeatability and minimization of differences—that is, the goal is to reduce the differences between the products being produced. Repeatability, like reliability, however, does not ensure safety or prevent adverse events. Again, repeatability does not depend on context. The procedures used under normal conditions may lead to adverse events if used during a pandemic. Reliably and repeatedly performing potentially unsafe procedures on every patient does not lead to good and safe healthcare. Safety and high quality require flexibility and more than simply reliability and repeatability.

---

<sup>2</sup> Most guns today are very reliable in their operation. In fact, one problem often is that they are too easy to discharge accidentally.

## Standard System Safety Engineering

The focus in system safety engineering is not on things that go wrong or go right but instead on eliminating or reducing *hazards*. Hazards are states of the system that can lead to adverse events (losses). Examples of hazards in healthcare are hospital-acquired infections, wrong-site surgery, incorrect medication administration, and so on. These adverse events (hazards) are what we want to prevent. Anesthesiology, for example, has been very successful in applying safety engineering practices to reduce many common anesthesiology-related surgical hazards.

Consider the hazard in healthcare today of inappropriate medication being administered. The result may not always be bad, but there are cases in which it could be harmful so the hazard needs to be controlled. Simply focusing on the cases when the medication does not result in harm (goes right) will not solve any problems. Instead, safety engineering focuses on why inappropriate medication may be administered and then designs the system to prevent such events. Understanding how and why a given system design may lead to hazards allows iterative improvement of the design process. At the same time, we also need to learn from the losses that do occur.

In fact, most major accidents result from the overconfidence that usually comes with success [Leveson, 2023]. People often believe that risk is decreasing or has decreased when losses do not occur. In fact, the risk is the same as it always was: the risk has not changed—only our perception of the level of risk has changed. Ironically, not having losses may lead to an *increase* in risk due to changes in behavior resulting from overconfidence that a loss cannot occur.

This is the folly of trying to learn from “success” or the lack of bad consequences. Continuing success only gives us more confidence that what we are doing is the safe thing to do when it may not be safe under all conditions, only those we have already encountered. Hollnagel points out many times in his books that accidents are rare. That is exactly why we cannot learn only from success. We need to identify the hazards that can occur and act to prevent them before they do.

To augment those protection efforts, we need to learn from the losses or adverse events that do occur despite our best efforts. It is from failure that system designers learn how to be more successful in the future. While we always try to behave in ways that we believe are safe, we can never know in complex, adaptive systems whether those behaviors will always be safe when repeated over time. Procedures effective against most viruses turned out to be less effective against Covid-19. The world changes over time and therefore our protection against adverse events must also change.

Russell Ackoff, a professor of Management Science at the Wharton School, University of Pennsylvania and one of the great systems thinkers of our time wrote about human performance: *“All learning ultimately derives from mistakes. When we do something right, we already know how to do it; the most we get out of it is confirmation of our rightness”* [Ackoff 1974, 1999].

## Safety-III

Safety-III has the same goals as standard safety engineering, but it is based on a different theoretical foundation, i.e., Systems Theory. Safety-III, like standard safety engineering, also focuses on eliminating or controlling hazards, but the processes used to accomplish this goal, hazard analysis and prevention along with accident investigation, are different and are proving to be more powerful and effective.

General systems theory concepts date back to Aristotle, but Systems Theory as a scientific discipline began around the 1920's. At that time, the biologist Ludwig von Bertalanffy and scientists from different fields started to formulate a general theory of complexity [von Bertalanffy, 1969]. The movement was an attempt to go beyond the analytic decomposition, introduced by Rene Descartes, John Stuart Mill, and others, that was so successful in creating modern scientific enquiry during the eighteenth and nineteenth centuries [Weinberg 1975].

The traditional approach to analyzing the *design* of complex systems is to decompose them into separate components and observe the components individually. The *behavior* of complex system is studied by decomposing it into separate events, connected linearly over time. The introduction of analytic decomposition and linear models of causality made the scientific revolution and modern science possible.

Where analytical decomposition is not possible, statistical approaches treat the system as an undifferentiated collection of interchangeable parts. The Law of Large Numbers (as a sample size grows, its mean gets closer to the average of the whole population) is used to describe behavior in terms of averages.

Because of its incredible success in the natural sciences, engineering and the social sciences have also traditionally used analytical decomposition and statistics to understand complex systems. But this approach has met with more limited success due to differences between natural vs. human-designed systems.

Traditional analytic decomposition is based on the assumption that the synthesis or combination of individual component analyses can be used to understand the properties of the system as a whole. This assumption in turn relies on a second assumption that the separation of the whole into separate components does not distort the phenomenon or system property of interest. A third critical assumption is that the interactions among the components are completely known and relatively simple. Statistical approaches, in addition, assume that the system components are sufficiently regular and random in their behavior that they can be studied statistically. These assumptions, alas, are seldom true for the types of tightly coupled, highly automated, and intricately interconnected sociotechnical systems common in healthcare and modern society today.

About a hundred years ago, recognition grew that the analytical decomposition and statistical approaches, which had revolutionized science, were not as effective when used for the study and design of complex, sociotechnical systems. Leaders in a variety of fields, such as von Bertalanffy in biology, Jerome Weiner in math, Jay Forrester and Russell Abbott in management, Margaret Mead in anthropology, James Gibson in psychology, and many others, created a new way of thinking about the systems they were studying, which were too complex and interdependent for complete analysis. At the same time, they were often too organized for the use of statistics, that is, too much underlying structure distorts the statistical results.

In these systems, the most important properties are *emergent*, which means that the properties of interest arise from the interactions and relationships among the parts of the system—that is, how the parts interact and fit together. Emergent properties cannot be identified by looking only at the individual components: Such properties can only be understood by looking at the system as a whole, taking into account all social and technical aspects. A common phrase to express the concept of emergence is that “the whole is greater than the sum of the parts.” This new approach to dealing with complex systems came to be known as Systems Theory, the term coined by von Bertalanffy for biology [1969].

Systems theory is not meant to replace the standard decompositional and statistical approaches, but instead to augment them with new scientific methods appropriate for ever more complex and interactive systems, in particular those created and designed by humans. It is still necessary to understand the behavior of the individual components of the observed universe and of human-designed systems. But there is also a need to understand complex interrelationships between the elements in a system in addition to the individual element behavior, for example, to understand the interplay of enzymes, the effect of pollution on ecosystems, the relationship between conscious and unconscious mental processes, and the structure and dynamics of social systems. It is usually not possible to examine the behavior of individual system components in isolation in order to completely understand or predict the effect of their behavior in the context of the whole.

Systems theory, therefore, focuses on systems taken as a whole. It is too easy to overlook system design flaws by examining only individual components and then trying to determine how the components will interact by assuming that the interactions are straightforward, occur as expected, and are easily understood. The most interesting and complex systems do not satisfy these assumptions.

Some important features of a systems approach are a focus on holism, interconnectedness and interdependence, context, dynamic complexity, and non-linearity, each of which is explained briefly here.

A systems-theoretic approach to problem solving deals with “wholes.” Systems are studied as an entity rather than a conglomeration of parts. Formal organizations (an army, a bureaucracy, a business enterprise, a healthcare organization) are treated as a system of mutually dependent components created with an overall purpose or goal. The system purpose is achieved through the operation of the separate components, subject to constraints on their individual behavior as well as on their interactions. System properties, such as quality, safety, and security emerge or are created through the operation of the system as a whole. For example, the behavior of two individuals in isolation may not lead to an adverse event and may appear safe in isolation. Putting them together, however, may not be safe.

In complex, adaptive systems today, it is no longer possible to deal separately with individual components; the complexity of the interactions among components, rather than simply failure of individual components, lead to most of the problems. For example, in hospital laboratory data, adverse events can be caused by inadequate laboratory analysis procedures. The most difficult adverse events to prevent, however, are those stemming from flawed communication between the physician and the laboratory, perhaps via an electronic health record system.

Using Systems Theory, the paradigm for preventing accidents and losses changes from simply making each component highly reliable, which is no longer very effective in complex systems, to controlling both the behavior of the individual components and the interactions among them. In the laboratory data case, both the analysis of specimens in the laboratory *and* the communication of the results to other parts of the healthcare system needs to be controlled in order to prevent hazardous behavior and adverse events [Leveson et.al., 2023]. Controlling unsafe *system* behavior becomes the primary goal in preventing losses rather than simply enhancing reliability by preventing individual component failure.

One result of interconnectedness and interdependence in complex systems is that changing one part of the system often creates unintended consequences and events in other parts of the system or in the system as a whole. This phenomenon is sometimes called the *Law of Unintended Consequences*. It results from the fact that the elements in systems are in mutual interaction. If we want to fix something or effectively change the behavior of a complex system, we must first understand the system as a whole and how the parts work together to achieve the system goals. Reducing queuing in the emergency department may have negative consequences on the intake of patients to the other parts of the hospitals. Changes need to be evaluated for their impact not only on one part of the system, but on the system as a whole.

Context is another important concept in a systems-theoretic approach to problem solving. All behavior is affected by the context in which it occurs. Human behavior, for example, is driven in part by our mental models or beliefs about the context and environment in which we are working. Tools based on systems theory can identify the contextual factors that lead to incorrect mental models or that make it difficult to notice when mental models are incorrect. In the worst case, the system design sometimes creates incentives for unsafe behavior. Carefully designing the context in which humans work can play an important role in reducing human error or in increasing it [Dekker 2011, Koppel and Gordon2012].

Two other properties of complex systems that create the need for a system-theoretic approach are dynamic complexity and non-linearity. Dynamic complexity is related to changes over time, where cause and effect are not related in a simple way. Dynamic complexity makes understanding and changing systems more challenging. For example, delays between cause and effect can lead to instability and unsafe decision making. Non-linearity is discussed later in this paper.

## Safety Management “Principle”

	Safety-I	Safety-II	System Safety Engineering	Safety-III
<b>Safety Management Principle</b>	Reactive, respond when something happens, or is categorized as an unacceptable risk	Proactive, continuously trying to anticipate developments and events	Proactive, concentrates on preventing hazards and accidents but does learn from accidents, incidents, and audits of how system is performing	Proactive, concentrates on preventing hazards and losses, but does learn from accidents, incidents, and audits of how the system is performing. Uses more powerful analysis tools. Emphasizes changing context to change human behavior and building safety into the system design.

### Safety-I

Purely reactive safety effort simply does not exist and has never existed. A few instances can be found in particularly backward companies, often in third world countries where there are lots of accidents and human life is not valued highly. Even then, purely reactive behavior is usually found only in the most primitive workplace environments.

### Safety-II

While Hollnagel suggests that proactively anticipating hazards and trying to prevent them is a new feature of Safety-II, it is universal standard practice and has been for a long time. Nearly everyone today—and for at least the last hundred years but probably longer—tries to anticipate hazardous and unwanted events and prevent them. There is even mention of preventing losses in the Bible. Healthcare is included here, although the specific proactive approaches used in healthcare tend to be less systematic than those in most other industries. Medical professionals do not wait until hospital-based infections occur, for example, before implementing infection controls such as sterilization procedures.

Surprisingly, no way of actually accomplishing proactive anticipation of events is mentioned in the Safety-II literature nor is any mention of the standard techniques and tools to accomplish this goal used in virtually all industries. Such efforts usually require specialists in using the proactive tools. The frontline workers (nurses, doctors, technicians) cannot be held responsible for anticipating adverse events and preventing them and for making systems resilient (as is suggested in some Safety-II literature). In most industries, these preventive activities are performed by safety professionals who work with frontline workers and management to create safe policies and protocols for the specific environment involved.

### Standard System Safety Engineering

Safety engineering has always focused on identifying and preventing hazards. Once the important losses or hazards (determined by the system stakeholders) have been identified, then the system safety engineering process begins. The specific tools and techniques used may differ among industries because of the different nature of the hazards involved, but there are general activities common to most safety engineering efforts. These activities include:

**Hazard Analysis:** Hazard analysis is used to identify how hazards could occur, that is, the scenarios and events leading to hazards, for example, why an incorrect medication might be given to a patient. Worst case conditions are considered: if a car is travelling at 100 mph, a wet road may lead to a hazard, in this case the inability to stop in a required distance and time, even if the brakes are effective at lower speeds and under different conditions. Usually multiple scenarios are identified, some of which may be quite complex.

Many different techniques have been developed for performing hazard analysis and identifying causal scenarios. Their efficacy and completeness depend on a match between the tools' assumptions about why hazardous behavior occurs and the characteristics of the system in which they are used. Surprisingly, forms of Failure Modes and Effects Analysis (FMEA) are being introduced in healthcare at the same time that FMEA is being phased out of other industries because of its inability to handle complex systems.

**Design for Safety:** The goal here is to use the information obtained from the hazard analysis to engineer or "design out" accidents before they occur, that is, to prevent hazards. The actual design approaches are often unique to various industries and the characteristics of the hazards in those industries.

The highest priority is to eliminate hazards from the system design, for example, to design systems that cannot get into hazardous states. In aviation, where an aircraft can have miles of wiring, female and male connectors were introduced to prevent incorrect connections. An example in healthcare is the hazard of incorrectly connecting the tubing that delivers medication, oxygen, and nutrition therapy. A design solution is to provide functionally dissimilar tubes or catheters that cannot be connected incorrectly. Using such design solutions, however, may require tradeoffs with the overall system goals, such as efficiency and profitability: It is more efficient and cheaper to use universal connectors.

If elimination is not possible or the tradeoffs are unacceptable, the secondary goal is to reduce the occurrence of hazards, but reducing occurrence is clearly less desirable than the elimination because it does not guarantee that hazards will not occur. For the tubing misconnection hazard, administrative controls such as policies and practices can reduce the number of misconnections, but probably will not eliminate them completely. The costs of any remaining adverse events may, in the long run, be greater than the cost of providing dissimilar connectors, i.e., eliminating the hazard.

If hazards do occur, then the final recourse is to try to reduce the impact, but, again, reducing impact is clearly less desirable than eliminating or preventing hazards. In the tubing example, it is less desirable to provide a way to detect that an incorrect connection has been made and to reverse it before serious harm occurs than to prevent a tubing misconnection in the first place. The safest solution usually is to provide both of these, that is, to try to prevent or reduce the occurrence of hazards but also to provide a way to avoid a loss by reversing the hazard (if it does occur) before a loss results.

Healthcare has the additional complication that there exist situations where tradeoffs are necessary among equally dangerous alternatives. Not treating a patient may be equally or more dangerous than providing a high-risk treatment. In those cases, solutions tailored to such situations may be necessary, such as providing as much information and assistance as possible to caregivers in their decision making.

More details on how to design for safety, including the design of human-automation interaction, can be found in Leveson (2023).

**Human factors engineering and human-centered design:** In human factors engineering, psychological concepts are applied to engineering designs to prevent human errors and to provide human operators with the ability to safely control the system. Sophisticated human-centered design concepts started to be developed in the 1980s and were first widely applied in the aviation community [Billings, 1997]. In this approach to engineering design, the role of humans in the control of systems is the

focus from the beginning of system design [Norman, 2013]. Given the intensive human nature of healthcare, human-centered design should be of highest priority but many times is not

**Operations:** Systems must not only be designed to be safe, but they must also be operated safely. Operational safety almost always includes oversight of system operation to ensure that the assumptions made during system analysis, design, and certification hold in the operational environment and that changes over time are not leading to increasing levels of risk. If dangerous conditions or behavior are caught in time, accidents can be prevented.

Operational safety involves managing operations to ensure proper training, identifying and detecting leading and lagging indicators of risk, instituting procedures to manage change and adaptation, performing maintenance, etc. Extensive data collection and analysis during operations has played an important role in improving design and operational safety in aviation and in nuclear power.

Systems also will change and adapt over time and operational safety efforts must accordingly adapt and change. One of the most important causes of decreases in operational safety is simply continuing to do the same things that were successful in the past, which appears to be promoted in Safety-II. The world is continually changing and the way we prevent losses must change in accordance. Success in the past does not guarantee success in the future.

**Management and Policy:** Emphasis on the design of Safety Management Systems within safety engineering dates back to the middle of the last century. These efforts involve creating more effective organizational structures, cultures, and information systems to prevent adverse events.

**Accident investigation and analysis:** Every industry investigates accidents, but it is usually a small part of the safety engineering effort. Most people recognize that waiting for accidents or adverse events, investigating them, and then preventing the reoccurrence of those particular events is a poor way to improve safety. At the least, identifying one cause at a time is extremely inefficient. Investigating adverse events, however, is an important and useful way to identify false assumptions about the real operation of the system as opposed to the assumed system operation during design. If a certain procedure is designed to prevent human errors, for example, the occurrence of multiple instances of human error in following that procedure is an indication that the underlying assumptions about typical human behavior were incorrect. The National Transportation Safety Board provides extensive investigation and learning from the few aviation accidents that do occur in order to understand why the existing controls were not effective.

**Regulation and Licensing:** Regulation may involve rules enforced by an oversight agency, voluntary standards, or certification/licensing authority. Regulation usually involves some type of approval of new systems before they are allowed to be used or certification of individuals before they are allowed to practice in a field. It also almost always includes oversight into the operation of the systems to ensure that assumptions about operation made during analysis, design, and certification still hold and that changes over time are not leading to increasing levels of risk. If such dangerous conditions are caught in time, accidents can be prevented. Examples of ways that oversight agencies collect information during operations include licensee event reports in nuclear power plants, aviation safety reporting systems, and auditing of airline, airport, and air traffic control operations. Such feedback exists in healthcare, but it tends to be less complete and less rigorous, which can lead to inadequate control over hazards.

**Stakeholder Groups:** Critical industries usually have stakeholder groups, such as the Airline Pilots Association, who are very active in providing input about perceived safety issues. Healthcare has many such groups.

This process of identifying hazards and then designing systems, including the operations and management aspects, to eliminate, prevent, or minimize hazard occurrence and impact is what safety engineering is all about. It is not investigating accidents, but of course that is done in order to learn

where proactive efforts went wrong. The goal certainly is NOT just to try to make the system operate safety by doing things “right.” There is no “right” or “wrong,” but only hazards or adverse events related to a particular system context and tradeoffs in assuring various system properties.

HRO theory defines an HRO organization or industry as one where accident rates are low compared to similar organizations or other industries [Roberts, 1990]. But that does not mean that those organizations or industries follow the principles prescribed in HRO theory [Dulac et.al. 2011]. In fact, I have found that organizations that follow HRO principles religiously have been involved in some of the worst accidents of the past 30 years (for example, [Baker Panel (on the Texas City Refinery Explosion) 2007; National Commission on the Deepwater Horizon Oil Spill 2011]. Organizations and industries with low accident rates use the system safety engineering practices described here [Leveson 2012 and 2023].

In some industries and countries, the risk or probability of a loss or a hazard may be calculated and used in decision making. The accuracy of such numbers, particularly for today’s complex, adaptive systems, is controversial [Leveson, 2023].

**Safety-III**

Safety-III has the same goals and activities as standard system safety engineering, but the specific techniques and tools used are designed to be more effective for today’s complex systems than the older techniques used in the past. The complexity of today’s systems has introduced new causes of losses, as described in the next section. New tools are needed. For example, FMEA and its variants (such as Functional FMEA) were designed for a very different world that is not as relevant for systems today and thus have limited effectiveness.

In addition, Safety-III emphasizes human-centered design to create systems that reduce human and other errors rather than focusing on trying to change the humans themselves.

Finally, emphasis in Safety-III is on designing safety into systems from very early system concept formation. Designing a system that is inherently safe is much easier and usually more effective than trying to make it safe later. For example, negative and positive pressure rooms in hospitals to prevent the spread of pathogens may be more effective than simply asking personnel to follow protocols and wear PPE as prevention measures. Of course, both may be needed to account for different causes of the hazards.

In summary, the primary management principle for safety for almost all fields is on the “engineering” or design of systems to prevent accidents. Most every industry and company also investigate accidents—it would be irresponsible not to use experience to update our assumptions and current knowledge. No industry that we know of, however, focuses on “things going right” or depends on investigation alone rather than prioritizing proactive prevention, even healthcare.

**Accident Causality and Causality Models**

	Safety-I	Safety-II	Standard Safety Engineering	Safety-III
<b>Explanations of accidents</b>	Accidents are caused by failures and malfunctions. The purpose of an investigation is to identify causes and	Things basically happen in the same way, regardless of the outcome. The purpose of an investigation is to understand how things usually go	Accidents are caused by linear chains of failure events. The purpose of investigation is to identify the linear chain of events and the root cause at the start of the chain.	Accidents are caused by inadequate control over hazards. Linear causality is not assumed. There is no such thing as a root cause. The entire socio-technical system must be designed to prevent

	contributory factors.	right as a basis for explaining how things go wrong.		hazards. The goal of investigation is to identify why the safety control structure (the designed controls) did not prevent the loss.
--	-----------------------	--	--	--

Causality is too complex a subject to go beyond a brief introduction here. For a more complete explanation of differing concepts of causality and the cause of accidents, see Leveson [2012 and 2023].

Safety-I

For several hundred years, accidents have traditionally been assumed to be caused by linear chains of failure events. Both accident investigation and prevention are based on this assumption. Newer concepts have been introduced in Safety-III. And, yes, the purpose of *all* accident investigation, as stated in Hollnagel’s table, is to identify causes and contributory factors in order to determine what needs to be changed. That is, and should be, the goal of all approaches to improving safety.

Safety-II

It is difficult to understand what Hollnagel [Hollnagel 2014] could mean by “things basically happen in the same way, regardless of outcome.” He does not provide any concrete examples in his books to explain the statement. The problem may be again in his use of vague and undefined terminology like “things,” “happen, and “same way.” What “things”? Everything happens in the same way? What could “happen in the same way” possibly mean? If I forget to turn off a faucet, it can lead to a flood in my kitchen but it will not lead to my car having a flat tire or to me being hit by lightning. The things that go wrong during surgery are very different than the things that go wrong in medication administration.

Hollnagel might mean that the general set of causes of things “going wrong” is the same as the general set of causes for things “going right.” However, all this statement says is that there is a set of causes and they lead to effects. Clearly, the outcome of an action depends on what has happened prior to the outcome and on the current conditions. Without any examples, it is difficult to understand this seemingly odd statement. And Hollnagel’s books contain almost no examples.

Having participated in a large number of accident investigations, I cannot imagine why investigators would spend their time understanding how things usually “go right.” We know that before the accident occurred. What we don’t know is why it “went wrong.” We know, for example, why pilots usually behave correctly, why the airplane maintains lift and stays in the air, and why brakes usually work to stop a car when the driver presses on the brake pedal. We trained people or designed devices to do those things.

When an accident occurs, the system (including the human operators) did not work the way the designer expected it to work. Why did the pilot behave unsafely, the aircraft lose lift, the brakes fail, or the physician prescribe a medication that triggers an allergic reaction? That’s what we need to understand to prevent a reoccurrence. Hollnagel recommends that

*[Using Safety-II for investigation] is done by constructing an account of everyday, successful performance and then looking for how performance variability, alone or in combination, could lead to loss of control. ... Asking for what went wrong leads to a search for errors, failures, and malfunctions. Asking for what did not go right creates a need to understand how work normally takes place and how things go right [Hollnagel 2014, p. 163].*

The problem here again is the vague definitions of “right” and “wrong.” Asking “what did not go right” is equivalent to asking “what went wrong.” Using the braking example again, asking what went wrong is to ask why the driver did not press the brakes in time to prevent the collision. Asking “what did not go right”

(as Hollnagel suggests above) involves asking why the driver did not press the brakes in time to prevent the collision. They are identical.

We already know why drivers normally press the brakes and stop the car before a collision occurs. We need to know why that did not occur in this case, i.e., why the behavior varied from the usual behavior or, in other words, varied beyond the safe boundaries. Things that “go right” here happen because the driver presses the brakes in time to prevent a collision. Things that “go wrong” here happen because the driver does not press the brakes in time to prevent a collision. The safe and unsafe results do not happen in the same way, as claimed. Asking “how performance variability, alone or in combination could lead to loss of control” is exactly what is done now and in the past. We know that unsafe medication was administered when a patient suffers an adverse reaction to it. That’s obvious. We need to know why unsafe medication was administered in order to prevent a reoccurrence of that hazard.

While it is possible to learn from “things going right,” namely, when nothing untoward happens, we can also learn the wrong things at this time unless we have complete understanding of how the system works and the current system state. That is rarely the case. We simply know, for example, that when a particular medication was administered the patient suffered no adverse consequence. Such information does not allow us to conclude that administering that medication in other instances will result in the same positive outcome.

Because accidents are rare, as are the conditions that lead to them, people may assume that something is safe because nothing untoward happens for the first hundred times they do something. If the conditions change, such as the underlying patient medical conditions, the next time doing ostensibly the same thing can lead to an adverse event. The Therac-25 radiation therapy machine was used hundreds of times before a patient was massively overdosed due to a design flaw and changes in the operator’s behavior over time [Leveson 2023]. In today’s highly complex systems, it is not possible for human personnel always to understand the system design in enough detail and depth to assure that their experience in the past applies to the present state.

As stated earlier, we learn only—or at least the most—from non-success rather than from success. In the search for cures, physicians may use approved medications off-label. If the result is successful, does that prove that the off-label use is safe? Only carefully controlled studies (“evidence-based medicine”) can provide high assurance. That does not mean that off-label use should not be allowed until such studies are conducted. Decision making always involves tradeoffs. The point is that learning from success is quite weak compared to learning from failure or from carefully designed experimentation.

Jens Rasmussen has suggested that, in the search for efficiency and productivity, we tend to migrate to the boundaries of safe behavior. We don’t know where those boundaries are, however, until we go beyond them and have incidents or accidents [Rasmussen, 1997]. Humans do not learn where these boundaries are or even that they have crossed them when everything “goes right” but only when things “go wrong.”

In fact, usually the only way to identify where the boundaries of safe operation exist *is* to go past them and see what happens. An accident does not always result every time people cross the boundaries, but how does one know a safe boundary has been passed when no adverse event occurs? When we skip a procedure or steps in a procedure in the interest of efficiency and saving time and nothing bad occurs, then we assume that skipping those steps is safe. When boundaries are passed without unsafe consequences, people learn to do unsafe things. At least, that is, until an adverse event *does* occur. Just examining how people have behaved when “things go right” (no loss occurs) is not adequate for reducing adverse events.

### Standard Safety Engineering

The same model of causality as a linear chain of failure events over time (described under Safety-I) exists in traditional system safety engineering: Events A and B together lead to event C which leads to event D.

From this standard perspective, root causes are the events starting the chains of cause and effect, such as A and B. The goal is to predict why the root cause and other subsequent failures might occur *before* there is an accident or to investigate why they occurred after an accident in case our proactive activities are not able to prevent all adverse events.

### Safety-III

Safety III uses more sophisticated models of causality than simple linear chains of failure events. The latter do not adequately explain the causes of losses in our complex, adaptive systems today. In complex, human-intensive systems, the most important relationships between components and events include the potential for feedback and other types of communication leading to non-sequential or circular causality. A physician prescribes a treatment and gathers feedback about whether that treatment was successful. The feedback helps to determine what next steps should be taken. Goal-seeking behavior includes feedback and monitoring of information about the state of the system and the components in it as actions are taken.

Thus, in causality models based on systems theory, a system's behavior is not viewed in terms of a linear sequence of events but instead emerges from a circular structure comprised of feedback loops. Decisions are often made on the basis of the feedback about the impact of that or other decisions.

In a systems approach, finding someone to blame is not the goal. Instead the goal is to understand why the system and each component in the overall system behaved the way it did given the context in which the behavior occurred. As an example, a typical chain of events might involve a nurse not administering a required medication to a patient before surgery, leading to an adverse patient outcome. In traditional causal analysis, the nurse's failure to give the medication is seen as the root cause of the adverse event.

In Systems Theory, a system's behavior is not viewed in terms of a linear sequence of events but instead emerges from more complex causality. Finding someone or some event in the chain of events to explain why the unwanted result occurred or to assign blame (including to management) is not the goal but rather to understand why the system as a whole, as well as each component within that system, behaved the way they did and to understand the complex, perhaps circular, causality involved. That information can be used to change the system design to prevent the sequence of events leading to a loss.

In the case of the medication not being administered described above (a true case), the nurse had never administered the medication in similar cases, and that omission had not led to adverse results. The surgical procedure involved was new, however, and in this case the medication was needed. Nobody informed the nurse of this new requirement. The standard direct communication between the physician and nurse (the *handoff*) before surgery did not involve any discussion about that medication or whether it had been administered and nobody thought there was a need to change the handoff procedure to include that question as it had not been needed in the past. In addition, the electronic health record did not specify (and was never intended to specify, for a variety of good reasons) who was actually responsible for administering the required medication. The nurse was not told of her new responsibility, and the doctor was confused also. Everyone involved thought that someone else was administering the medication. Nobody checked to determine whether the medication was actually given, but there were no procedures in place to perform this check (again, this type of surgery was new to this hospital). In addition, there was no easy way to get the information needed from the electronic health record to determine that the medication had not been administered.

The examination of the contributions of the entire system design to the events provides more information about how to change the system to prevent similar adverse events than identifying simple behavioral causal chains. This example points to needed changes in procedures, communication, assignment of responsibility, design of electronic information systems, management of change, and management oversight. While much has been written about eliminating the culture of blame in hospitals,

few practical ways to do this have been suggested. Simply not assigning blame does not prevent the same events from happening again. Changing the process for investigating adverse events, using an accident causality model based on System Theory can provide this missing piece.<sup>3</sup>

The differences between standard linear causality models and system theoretic ones are described in more depth in books on Systems Theory [Meadows 2008; Sterman 2000] and in Leveson [2012, 2023]. Basically, more aspects of causality are included. Causality can be much more complex than is assumed in most adverse event investigation procedures.

As another example, government agencies or payor organizations in some industries have attempted to promote safety by rewarding groups that have the fewest accidents and incidents [Levitt and Parker, 1976]. As a result, fewer incidents were reported, but serious accidents actually increased. The original goal of increasing safety was subverted into one of reporting the fewest incidents. The lower reporting levels resulted from the less serious incidents being hidden and thus never being investigated and the causes eliminated. The learning that could have been achieved from investigating the less serious events was inhibited. One way of avoiding this phenomenon is to encourage employees to report safety-related incidents, rather than rewarding employees or groups with the best safety records or establishing goals such as zero accidents.

A similar type of behavior can be observed in organizations that reward rigorously following standard procedures. In this case, the benefits of the procedures are often not visible, and employees view the requirements as impeding their normal working processes. Employees may then obey the letter of the standards, but not comply with the underlying intentions [Marais et. al., 2006].

The goal of accident investigation in Safety-III is not to assign blame. Most people are trying to do the right thing; we need to understand, as emphasized by Dekker [1994], why it made sense to them to behave in a hazardous way.

### Attitude Toward Human Factors

	Safety-I	Safety-II	System Safety Engineering	Safety-III
Attitude to the human factor	Humans are predominantly seen as a liability or a hazard.	Humans are seen as a resource necessary for system flexibility and resilience.	Humans are expected to prevent or respond to hazards and to be flexible and resourceful when they occur.	The system must be designed to allow humans to be flexible and resilient and to handle unexpected events.

#### Safety-I

Once again, the description of Safety-I, characterized by Hollnagel—without any citations or examples—as what people do today, is incorrect. Humans are *not* predominantly seen as a liability or hazard in safety engineering or in engineering in general. In fact, they are and have always been seen as necessary for system flexibility and resilience, just as Prof. Hollnagel suggests they should be. Most attempts to replace humans with machines are driven by the desire to save money and increase efficiency, not as a way to make things safer or less hazardous. For example, the primary impetus behind removing pilots from aircraft cockpits is cost and a shortage of pilots, not to increase safety. In fact,

<sup>3</sup> This example is real. After the third occurrence, it was investigated using Safety-III techniques, and the causal factors noted here were identified [Samost 2015; Leveson, et.al. 2020].

safety is the primary reason that the FAA is pushing back on the desire of airlines to allow cockpits with only one pilot.

In healthcare, automated devices are often employed because doing everything by hand today would be too slow and require more trained people than could be recruited: Scaling up to larger systems, increasing throughput, and reducing time delays is difficult for jobs depending solely on human workers. Switchboard operators had to be replaced by automation to make the communication required in our complex society feasible. Most automation in complex, high-skill environments today is introduced to assist humans in doing their jobs, not to replace them. The speed and efficiency usually demanded today and the scale of our operations is infeasible without automated assistance.

It is also true, however, that users and operators have been a convenient scapegoat after accidents to deflect attention away from mistakes made in system design (including the design of automation) and management.

The term “humans” is used a lot by Hollnagel, but it appears that he is mostly referring to frontline personnel in relatively simple workplaces. There are lots of humans in our systems: humans design systems, manage them, certify them, and maintain them, as well as operate them. They are not predominantly seen as hazardous or a liability but as necessary for successful operation. To improve their performance, sophisticated human factors research and human-centered design is viewed as important in creating today’s complex systems. These ideas are not new, but have existed for a long time, with various levels of adoption in different industries. Those with the best safety records also have the most sophisticated human factors practices.

There are a few cases where our technology has outpaced the ability of humans to operate the systems, even with automated assistance. For example, unstable aircraft are being designed for the military because of advantages in maneuverability, etc. They must be flown by automation because humans cannot adjust the control surfaces at the speed needed to control the aircraft. But note that even then, human pilots are still retained to provide higher-level control, e.g., to manage the automation. Pilotless drones have surprisingly poor safety performance, but they are used because the tradeoffs are very different in the defense industry than in commercial aviation.

While unstable aircraft could provide benefits in the commercial aviation community, such features are not being introduced quickly because designers are not willing to trust safety provided only by automation. In healthcare, the goal of automation is usually not to eliminate error-prone humans but to do things that humans are not capable of doing or to do things for which there are not enough trained humans. Automation may provide information to physicians, but it is still almost always the physicians who make final decisions about how to treat patients.

## Safety-II

Once again, what Hollnagel describes as the attitude toward human factors in Safety-II is simply what has always been true in safety engineering and in engineering in general. Considering humans apart from the design of the systems in which they perform their duties, however, is the opposite of a systems approach. Human factors must start by looking at the design of the overall system, not the behavior of the humans who are in that system.

## System Safety Engineering

System safety engineering leaves humans in systems not because they are seen as a liability or hazard but because they are needed to provide flexibility and deal with unexpected events. We depend on humans to do what cannot be definitively defined as the safe thing to do in all situations. An important role for humans in automated systems is to deal with the events that result from the system not working flawlessly. The fact that people can adjust work to the actual conditions is exactly why we do not automate them out of system designs.

However, there are limitations that need to be recognized. Hollnagel claims that:

*People can detect and correct when something goes wrong or when it is about to go wrong, and hence intervene before the situation seriously weakens [Hollnagel 2014, p. 120].*

This statement is true only if the system design allows the flexibility required to do these things. The important principle when taking a system's view of safety is that the system must be *designed to allow successful resilience by human operators*. To accomplish this goal, the system design must satisfy three requirements [Rasmussen, 1987a, 1987b, 1990]:

1. The system has been designed to provide people with the information they need to detect when something goes wrong. Stated in a more scientific way, the humans in the system, including operators, managers, government overseers, etc. must be aware that a hazardous state or a precursor to that hazardous state has occurred. The same is true for software or any type of automated system controller that we expect to respond to hazards. Without knowing that a hazardous state exists, then it is not possible to respond in order to minimize damage.
2. Accurate information about the current state of the system must be available in a timely manner. The hazardous state must be observable within the time period necessary to prevent a loss. For example, the physician or nurse must get feedback about a patient having a bad reaction to prescribed medication while they can still reverse the effects. The responsible people must have the information necessary to solve problems, that is, the information necessary to respond effectively to a potential hazard, in time to use this information to solve the problem. In healthcare, the information provided in an EHR may not be easily located or it may be incorrect or misleading or the information may not get to the EHR in time to prevent an adverse event [Fry and Schulte 2019; Koppel 2016; Moy et.al. 27; Nathalie and Kwon 2019].
3. The system must have been designed to provide people with the ability to intervene. If there are no actions that the controller (human operator or manager, software, social structure) has available to recover from a hazard, then people can be resilient—in terms of knowing what to do—but not be able to respond in an effective way. A very simple example of a design feature that allows intervention is the “undo” button, which most of us depend on in new software applications. More specifically, if a hazard does occur, perhaps because of errors on the part of the frontline personnel themselves, either they must have a means to reverse the errors or the ability to move to a non-hazardous state. Alternatively, other parts of the system must have these capabilities. If the personnel are not provided with effective tools, such as direct controls and overrides in an automation-intensive system, the humans may be adaptable and resilient but will not have the ability to prevent losses. As a simple example, if a physician knows that the standard medication for treating a patient is worse than an alternative in a particular case, the physician must have the ability to obtain the alternative medication.

All three of these conditions involve system design and are not simply a function of the adaptability of the people involved. During the Fukushima Daiichi nuclear power plant accident in Japan, the ability of the operators to shut down the nuclear reaction was thwarted by the fact that the electrical supply equipment needed to do so was located in the basement, which was immediately flooded by the tsunami and became inoperable. All the information systems that could provide information about what was happening in the reactors stopped operating because of the lack of electricity. The operators were highly trained and potentially resilient and they tried to respond, but the design prohibited them from taking effective action. Modern Safety-III hazard analysis and system design techniques can detect and prevent such system design flaws.

As noted, human resilience requires that the system be designed to allow the human to be effective in handling unforeseen events. That sounds impossible, i.e., how does one design for something that is not foreseen? In fact, it may not be possible in all cases. But we can in many cases provide humans with the information about the current state and with ways to modify the state. Clearly, however, we cannot provide the controller—whether automated or human—with unlimited information or the means to control everything. What that means is that resilience will always be limited, but improved hazard analysis and system design techniques can assist in designing to stretch those limits.

Sometimes designers are so sure that their systems can correct themselves that they do not include appropriate means for human intervention. Clearly, systems can (and do) fail or, more generally, behave in ways that the designers did not anticipate. But if the designers do not provide appropriate means for the operator to intervene, losses can occur.

These are all system design problems, not simply human behavior problems

### Safety-III

*Resilience* in Safety-III is defined as the ability of the *system* (as a whole, including human, physical, and social components) to maintain the safety constraints in the face of unplanned or inadequately controlled behavior or hazards. In a resilient system, losses are prevented or minimized in the face of unexpected events. To accomplish this goal, the entire sociotechnical system must be designed to be resilient, not just part of it. The most critical need is to learn how to design systems in which humans are *able* to be resilient.

Modern human factors and systems thinking recognizes that human behavior is greatly influenced by the environment in which it occurs. Human behavior, in fact, can only be fully understood by examining the context in which it occurs. As an example, healthcare workers may want to be flexible but are often hindered by the design of electronic health systems.

Safety-III takes a modern view of human error, which focuses on system design to prevent or reduce human error [Norman 2013; Zsombok and Klein 2014]. A related focus is on how to design the system to enhance human performance.

Starting from the viewpoint that human behavior cannot be understood apart from the environment (context) in which it occurs, then performance degradation, errors, and adverse events are due to mismatches between capabilities and the demands of the task environment [Dierks et.al 2008; Koppel and Gordon 2012; Ovretveit 2011; Tarik et.al. 2021]. For example, delivering healthcare in an increasingly distributed and complex environment may lead to incorrect sequencing, duplication, or omission of important interventions. Intense productivity pressures and heavy workload may lead to delayed treatment or response to patient physical changes.

In addition, providing large volumes of data in the face of conflicting or inadequate decision support may paradoxically increase uncertainty or ambiguity and may lead to faulty decision-making or delayed recognition of critical clinical trajectories [Koppel 2016]. In many cases, the environment or clinical system—representing everything from policies and priorities to the social and technical components—does not support the provider in executing relatively straightforward patient care tasks. In some cases, the demands of the task have been increased by organizational policies and by the physical, functional and social features of the environment [Koppel and Gordon 2012] In other words, relatively straightforward tasks, such as titrating a pressor to maintain blood pressure within a specific range, have been made more demanding by policies or practices that constrain medical practice such as a cumbersome pharmaceutical procurement process.

The larger system in which the healthcare personnel are embedded has factors that strongly impact safety. For example, with the current cost-containment pressures and desire to deliver care more efficiently, providers are being pushed to perform increasingly complex procedures in an ambulatory (outpatient) setting [Stringfellow et.al., 2009]. New safety concerns are emerging as a result: Outpatient

surgical units may be less fully equipped for contingencies and often are remote from inpatient crisis management teams. Pre-procedure patient preparation may be considerably less formal, including little or no pre-planned backup by specialists for management of complications.

In contrast to inpatient units, which are capable of processing patients round the clock, outpatient units have fixed operational hours leading to “sundown” issues—shut-down of these units at a specific time—which leads to increasing potential for premature patient discharge or rushed execution of procedures towards the end of the day. Even when patients are discharged using appropriate readiness criteria, there can be heavy reliance on patient and family to manage post-procedure recovery and an increased need to coordinate care across several physical transitions. Perhaps the most important concern is that operational decisions and policies governing staffing, resource allocation, information exchange, space allocation and prompt access to specialists, can favor productivity goals over safety.

The tension between productivity, efficiency and safety is particularly strong in large, urban medical centers, where high demand for traditional as well as complex and rescue procedures has generated tremendous throughput pressure. Because services can be delivered more economically in an outpatient setting, the envelope is being pushed to perform increasingly complex cases in this manner. To counter the increased complexity and risk, a large number of safety controls have been introduced to safeguard patients, such as mandatory staffing ratios, pre-procedure time outs, required engagement of specialists to assist with certain types of procedures, and a large number of redundancy-based checklists and cross-referencing protocols. However, because these safety controls have the potential to add to the overall cost and slow or delay the process of care —thus undermining the dominant economic objective for this particular delivery model—they often are waived by physicians and nurses in the interest of maximizing the productivity and efficiency objectives [Baker 2022; Dierks et.al. 2008; Koppel and Gordon 2012; Stringfellow et.al. 2009]

Preventing human “errors” and unsafe behavior using Safety-III focuses on these contextual influences on healthcare providers rather than focusing on the humans removed from the context in which they find themselves.

**Role of Performance Variability**

	Safety-I	Safety-II	System Safety Engineering	Safety-III
<b>Role of Performance Variability</b>	Harmful, should be prevented as far as possible.	Inevitable but also useful. Should be monitored and managed.	Performance variability is only a safety concern when behavior moves outside safe boundaries. A goal is in system design to avoid hazards when performance (of operators, hardware, software, managers, etc.) varies outside those safe boundaries.	Humans usually vary their performance for very good reasons. The goal is to design the system as a whole so that performance variability is not hazardous and conflicts between productivity and safety are eliminated or minimized.

### Safety-I

Again, the description of Safety-I is very puzzling. Most safety engineers do not think of performance variability by humans as something that should be prevented.

Reliability and repeatability are a theme in the HRO literature, but as argued earlier, reliability and safety are different properties [Dulac et.al. 2011]. Reliability and repeatability are not required safety nor do they ensure it: Adverse events can result from unsafe processes that are reliably repeated. Safe processes may not involve repetition at all, only the elimination or control of hazards.

Humans are retained in almost all systems today *because* they are able to vary their performance to match current conditions. Engineers rely on human performance variability in the form of adaptability in order to achieve the system goals in the face of changes in the system and the environment.

### Safety-II

Hollnagel suggests that human variability is inevitable, but also useful and that it should be monitored and managed. It is doubtful that anyone would disagree, but the relationship to safety is not clear. Performance variability is more a concern of quality control and efficiency in assembly line processes where the goal is to produce identical widgets, not safety in the design and operation of complex systems. HRO theory seems to promote the same confusion.

### Standard Safety Engineering

Standard safety engineering does not, in general, try to eliminate human performance variation or consider it to be harmful. To prevent accidents, in fact, operators are usually counted on not to blindly follow procedures when they are unsafe.

There are some situations when performance variation can be unsafe. One involves the behavior of physical devices. We do not want our car brakes to vary in performance. Healthcare safety relies on the performance of medical equipment, such as anesthesia and diagnostic equipment or heart rate monitors, not varying randomly in their behavior. But it appears that Hollnagel was not referring to hardware here as Safety-II seems to focus almost totally on human behavior.

There are circumstances where human procedures are useful and compliance important. Pilots, for example, cannot be expected to be experts on and understand everything about the design of the aircraft. Even many engineers are only experts on parts of the design of very complex systems. In addition, pilots in an emergency situation usually do not have the time to start reasoning from first principles. Therefore, they are trained on how to respond and are given procedures to apply based on the current conditions when they do not have time to read through a manual to find a solution to a problem in the twenty seconds before crashing.

If the procedures do not work, then something else has to be done, and we try to train pilots (and we depend on them) to make good decisions in these cases. But that does not mean that we should not provide pilots or healthcare personnel with procedures and treatment protocols and train them about when it is appropriate to use them and when it is not. Flying and healthcare would be a lot more dangerous if we did not.

In more everyday examples, we also depend on skills or learned procedures when time is of the essence. If I had to figure out what to do every time I have to stop my car to avoid an accident and not just automatically step on the brakes (a learned procedure), I would be involved in a lot of accidents. I do not want to encourage performance variability in this type of situation.

### Safety-III

Humans usually vary their performance for very good reasons, and, in fact, often must do so to accomplish the system goals. The goal is to design the system as a whole so that performance variability is not hazardous and conflicts between productivity and safety are eliminated or minimized. When

performance (of operators, hardware, software, managers, etc.) varies outside safe boundaries, the goal is to design so that safety is still maintained. This design goal is called fault tolerant and fail-safe design.

There are situations, however, when constraining performance outside safe boundaries is necessary. Hollnagel states that:

*Unacceptable outcomes or failures cannot be prevented by eliminating or constraining performance variability since that would also affect the desired acceptable outcomes [Hollnagel 2014, p. 138].*

This is another black and white statement and overgeneralization. There are times when constraining performance variability is exactly the right thing to do even if it could affect desired acceptable outcomes. We require that pilots lock the cockpit door in the aircraft in order to keep out those wanting to do harm to the aircraft and passengers. There are probably some odd cases where it might be better not to have cockpit doors locked, as in the case of the pilot suicide in a Germanwings aircraft. The copilot waited until the Captain had left the cockpit temporarily and then locked the captain out and committed suicide by flying into a mountain. But those cases are much less likely than the need for the design to prevent entry to bad guys.<sup>4</sup>

The most difficult decisions in system design involve determining how much and what kind of protection needs to be provided in order to prevent unintentional or intentional but unsafe behavior. There are *always* tradeoffs. Sometimes performance needs to be constrained to protect against specific types of behaviors. We do not allow houses to be built from highly flammable materials nor allow the use of potentially faulty electrical wiring practices. Often, however, we allow some types of deviation from safety constraints and standards when it is necessary or desirable but provide controls over when these types of deviations are acceptable.

In Safety-III, the goal is to *design the system* in which people are working so that performance variability is safe and conflicts between productivity, achieving system goals, and safety are eliminated or minimized as much as possible.

In summary, nobody would argue against allowing human performance to vary. The problem is that doing so is not as simple as monitoring and managing human performance variation. The system design must allow people to respond effectively, which requires preplanning and appropriate design. As a simple example, backup power generators are provided in hospitals in case power failures occur. Simply depending on the healthcare personnel to adapt without power is unrealistic.

## Conclusions

A great deal of effort has been devoted to improving patient safety, but progress has been slow and limited. System safety engineering provides a way forward. The practice of system safety engineering has made it possible for the aviation industry to drive down accident rates to extremely low levels while ensuring that transportation and business goals are achieved. Safety in aviation does not rely on perfect behavior of people and equipment for the tens of millions of commercial flights that occur each year. It also does not depend on a few independent activities, such as checklists and extensive accident investigation, but on a total system design where each part of the system reinforces the safety-related activities of the other parts and provides information for proper decision-making.

The industries that have been most successful in preventing accidents have employed this type of comprehensive safety engineering approach that includes modeling and hazard analysis, design for safety and fault-tolerance, safety management systems, human-factors engineering and human-centered design, safety programs during operations, regulation and licensing, event reporting systems, and, yes, accident investigation and analysis. It is based on system-theoretic concepts that consider the system as a whole and not just parts of it. The least successful fields in reducing accidents, such as workplace or

---

<sup>4</sup> Override codes to unlock the door are now provided to aircraft crew to avoid such situations.

occupational safety, have focused on the behavior of the workers and not on designing (engineering) the entire system for safety.

Safety-III takes this approach even further and augments it for the complex systems we are creating and using today. Some initial attempts to apply Safety-III to healthcare have been very promising, for example [Baker 2023; Balgos 2012; Bargal et.al 2018; Brooks and Ochieng 2021; Canham et.al. 2018; Dierks et al. 2008; Ghorbani et.al 2024b; Grimmatt 2021, 2022; Harari 2024; Kadupokotla 2017; Leveson et.al 2020; Leveson et.al 2023; O'Neil 2014; Pawlicki et.al. 2016; Raman et.al. 2016; Proctor et.al 2015; Raman et.al 2016; Samost 2015; Samost-Williams and Nanji 2020; Tang et.al 2015; Thomas et.al 2016; Trantz et.al 2018; Wong 2020; Wong et al. 2020; Wong et.al. 2021; Yamaguchi and Thomas 2018].

Health care professionals are working hard to reduce adverse events. Safety-III provides a promising and proven way to accomplish this goal.

## References

1. Ackoff, R.L. A Lifetime of Systems Thinking, *The Systems Thinker*, vol. 10(5): 1-4, Pegasus Communications, 1999.
2. Ackoff, R.L. *Repackaging the Future: A Systems Approach to Social Problems*, Wiley, 1974.
3. Baker, E.W. *Safety in Hospital Medication Administration Applying STAMP Processes*, Master's Thesis, SDM, MIT, January 2022
4. Baker Panel, *The Report of the BP U.S. Refineries Independent Safety Review Panel*, January 2007.
5. Balgos, V. A Systems Theoretic Application to Design for the Safety of Medical Diagnostic Devices, M.S. Thesis, MIT, 2012.
6. Bargal, B., Benneyan, J. and Eisner, J. Use of STPA to Design Safer Opioid Prescribing Processes, 2018, *IISE Transactions on Occupational Ergonomics and Human Factors*, 6(3-4):1-16
7. Brooks, A. and Ochieng, W. Safety-Guided Design: Integrating STPA into the Systems Engineering Process for the Safety of Remote Health Workers, *2021 STAMP Workshop*, <http://psas.scripts.mit.edu/home/mit-stamp-workshop-presentations/>
8. Bates, D.W. and H. Singh. Two Decades Since 'To Err Is Human': An Assessment of Progress and Emerging Priorities In Patient Safety. *Health Affairs* 37:1736–1743, 2018.
9. Billings, C.E. *Aviation Automation: The Search for a Human-Centered Approach*, Mahwah, New Jersey: Lawrence Erlbaum Associates, 1997.
10. Canham, A., T. Jun, P. Waterson and S. Khalid. Integrating systemic accident analysis into patient safety incident investigation practices. *Applied Ergonomics* 72, 1–9, 2018.
11. Dekker, S., *The Field Guide to Understanding Human Error*, Farnham, U.K.: Ashgate, 1994
12. Dierks, M.M. N. Dulac, and N.G. Leveson. "System Dynamics Approach to Model Risk in Complex Healthcare Settings: Time Constraints, Production Pressures, and Compliance with Safety Controls," International System Dynamics Conference, Athens, July 2008.
13. Dulac, N., Marais, K., and Leveson, N. Beyond Normal Accidents and High Reliability Organizations: The Need for a Systems Approach to Safety in Complex Systems, Research Report, MIT, 2011.
14. Fry E and Schulte, F. Death by a Thousand Clicks: Where Electronic Health Records Went Wrong. *Fortune Magazine*, March 2019.

15. Ghorbani, J., Marquez, M. and Samedy, P. Application of CAST in Site Identification Safety in Interventional Radiology (IR), 2024 STAMP Workshop, <http://psas.scripts.mit.edu/home/mit-stamp-workshop-presentations/>
16. Ghorbani, J., Samedy, P., Marquez, M., Chu, B., Bellamy, M. A Prospective Systems Safety Analysis: System-Theoretic Process Analysis (STPA) Applied to an Interventional Radiology Procedure, HFES, Int. Symposium on Human Factors and Ergonomics in Health Care, Chicago, 2024b
17. Grimmett, W. Introducing STAMP to a Major Health Organization, 2021 STAMP Workshop, <http://psas.scripts.mit.edu/home/mit-stamp-workshop-presentations/>
18. Grimmett, W. Clinical Governance Hazard Analysis: Introducing STAMP to a Major Health Organisation, 2022 STAMP Workshop, <http://psas.scripts.mit.edu/home/mit-stamp-workshop-presentations/>
19. Harari, R. Augmented reality for crisis management in the operating room: A System-Theoretic Process Analysis Approach, 2024 STAMP Workshop, <http://psas.scripts.mit.edu/home/mit-stamp-workshop-presentations/>
20. Hollnagel, E. Safety-I and Safety-II: The Past and Future of Safety Management, Routledge, 2014.
21. Hollnagel, E. Safety-II in Practice: Developing The Resilience Potentials, Routledge, 2017.
22. Jha AK. Testimony of Ashish K. Jah to the U.S. Senate Committee on Health, Education, Labor, and Pensions, July 17, 2014, Washington D.C, July 9, 2014
23. Kadupokotla, K. STPA Analysis for Clinical Programming Software of Cochlear Implant System for Profoundly Deaf People. 2017 STAMP Workshop, <http://psas.scripts.mit.edu/home/mit-stamp-workshop-presentations/>
24. Koppel R. "Great Promises of Healthcare Information Technology Deliver Less." Chapter in, C.A. Weaver (ed.), *Healthcare Information Management Systems: Cases, Strategies, and Solutions, Health Informatics*, Springer International Publishing Switzerland, 2016
25. Kohn, K.T., Corrigan J.M., Donaldson M.S. (eds), *To Err is Human: Building a Safer Health System*, Committee on Quality Health Care in America, Institute of Medicine: National Academy Press, Washington, D.C. 1999.
26. Koppel R and S. Gordon. *First, Do Less Harm: Confronting the Inconvenient Problems of Patient Safety*. Ithaca, NY: Cornell University Press. 2012
27. Leveson, N. *Engineering a Safer World*, MIT Press, 2012.
28. Leveson, An Introduction to System Safety Engineering, MIT Press, 2023.
29. Leveson, N., Samost, A., Dekker S., Finkelstein, S, Raman, J. A systems Approach to analyzing and preventing hospital adverse events, *Journal of Patient Safety*, Vol. 16, No. 2., 2020.
30. Leveson, N., Thomas J., Harrington, P., Rose, R, Powell, S. and Keller, A. System Safety within Laboratory Data Exchanges Report, Research Report, MIT. Sept. 2023..
31. Levitt, R.E. and Parker, H.W. "Reducing construction accidents—top management's role," *ASCE Journal of the Construction Division*, Vol. 102, No. CO3, Sept. 1976, pp. 465–478.
32. Marais, K., Saleh, J., and Leveson, N. Archetypes for organizational safety, *Safety Science*, 44(7):565-582, 2006.
33. Meadows, D. *Thinking in Systems: A Primer*, Chelsea Green Publishing, 2008.

34. Moy AJ, Schwartz JM, Chen R, Sadri S, Lucas E, Cato KD. Measurement of clinical documentation burden among physicians and nurses using electronic health records: a scoping review. *J Am Med Inform Assoc*; 28 (6): 998-1008, May 2021
35. Nathalie J.N and C.S. Kwon. Electronic Health Records—A System Only as Beneficial as Its Data. *JAMA Netw Open*, 2(9), 2019.
36. National Commission on the BP Deepwater Horizon Oil Spill and Offshore Drilling, , *The Gulf Oil Disaster and the Future of Offshore Drilling Report to the President*, GPO, Washington, D.C. January 2011.
37. Norman, D. *The Design of Everyday Things*, New York, NY: Basic Books, 2013.
38. O'Neil, M., "Application of CAST to Hospital Adverse Events," Master's Thesis, MIT, June 2014
39. Øvretveit, J., 2011. Understanding the conditions for improvement: Research to discover which context influences affect improvement success. *BMJ Quality & Safety* 20, i18–i23, 2011
40. Pawlicki, T. and Harry, T. An Approach to Determine Safety Indicators in Radiation Therapy. 2016 STAMP Workshop, <http://psas.scripts.mit.edu/home/mit-stamp-workshop-presentations/>
41. Pawlicki, T., A. Samost, D.W. Brown, R.P. Manger, G-Y. Kim. Application of systems and control theory-based hazard analysis to radiation oncology. *Medical Physics* 43, 1514–1530, 2016.
42. Proctor, S., Hatcliff, J., Fernando, A., Weininger, S. Using STPA to Support Risk Management for Interoperable Medical Systems, 2015 STAMP Workshop, <http://psas.scripts.mit.edu/home/mit-stamp-workshop-presentations>
43. Raman, J., N. Leveson, A. Samost, N. Dobrilovic, J. Oldham, S. Dekker, S. Finkelstein. When a checklist is not enough: How to improve them and what else is needed, *Journal of Thoracic and Cardiovascular Surgery*, 2016, Aug: 152(2):585-592, 2016.
44. Rasmussen, J. The Definition of Human Error and a Taxonomy for Technical System Design, in *New Technology and Human Error*, edited by Jens Rasmussen, Keith Duncan, and Jacques Leplat, p. 293-301, New York: John Wiley & Sons, 1987a
45. Rasmussen, J. Approaches to the control of human error on chemical plant safety, In *Proceedings of the International Symposium on Preventing Major Chemical Accidents*, American Institute of Chemical Engineers, 1987b.
46. Rasmussen, J. Human Error and the Problem of Causality in Analysis of Accidents, in *Human Factors in Hazardous Situations*, edited by D.E. Broadbent, J. Reason, and A. Baddeley, p. 1-12, Oxford: Clarendon Press, 1990
47. Rasmussen, J. Risk Management in a Dynamic Society: A Modeling Problem, *Safety Science*, 27(2/3):183-210, 1997.
48. Roberts, K.H. Managing high reliability organizations. *California Management Review*, 32(4), pp. 101–114, 1990.
49. Samost, A. A systems approach to patient safety: preventing and predicting medical accidents using systems theory (Thesis). Massachusetts Institute of Technology, 2015.
50. Aubrey Samost-Williams and Karen C. Nanji, A Systems Theoretic Process Analysis of the Medication Use Process in the Operating Room, *Anesthesiology*, Vol. 133, 332-341 August 2020

51. Silvis-Cividjian, N. Using a systems-theoretic approach to analyze safety in radiation therapy—first steps and lessons learned. *Safety Science*, 122(2-3),Feb. 2020
52. Sterman, J., *Business Dynamics: Systems Thinking and Modeling for a Complex World*, McGraw-Hill, 2000.
53. Stringfellow, M.V. N.G. Leveson, and M. Dierks, The Impact of Healthcare Incentive Structures on Risk, *Int. System Dynamics Conference*, Albuquerque, July 2009.
54. Tang, A., Samost, A., Viswanathan, A., Cormack, R., and Damato, A.. Analyzing the Safety Implications of a Brachytherapy Process Improvement Project Utilizing a Novel System-Theory-Based Hazard-Analysis Technique, *Medical Physics*, June 2015.
55. Tariq, R.A, Vashisht, R.; Sinha, A., and Scherba, Y. Medication Dispensing Errors and Prevention, 2021.
56. Thomas, J., Ang, Y.H., Chung, K., and Gao, O.Q. STPA Analysis of Intravenous Patient-Controlled Analgesia, 2016 STAMP Workshop, <http://psas.scripts.mit.edu/home/mit-stamp-workshop-presentations/>
57. Trantza, S., Edwards, B., Dokas, I., and Sherael. W. CAST Analysis of Pregnancy Events in the U.K. Due to Isotretinoin Administration, 2018 STAMP Workshop, <http://psas.scripts.mit.edu/home/mit-stamp-workshop-presentations/>
58. US HHS, New HHS Data Shows Major Strides Made in Patient Safety, Leading to Improved Care and Savings, May 7, 2014
59. Von Bertalanffy, L. *General Systems Theory*, New York: George Braziller, 1969
60. Weinberg, G.M. *An Introduction to General Systems Theory*, New York: Dorset House Publishing, 1975.
61. Wong, L. Enabling Effective Safety Learning in Healthcare, Ph.D. Dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2021
62. Wong, L., Huynh, E., Mak, R.H., Leveson, N., and Singer, L. "STAMPing out MRI Simulation Hazards with a System-Theoretic Accident Model and Processes Approach to Proactive Hazard Assessment" *International Journal of Radiation Oncology Biology Physics* [03603016] 108.3: e204-e205. Nov. 2020
63. Wong, L., Pawlicki, T., and Leveson, N. Radiotherapy Application of Causal Analysis based on Systems Theory (CAST), *63rd Annual Meeting of the American Association of Physicists in Medicine*, June 2021.
64. Yamaguchi, S. and Thomas, J. A system safety approach for tomographic treatment, *Safety Science*, Vol 118, Oct. 2019, p. 772-782.
65. Zsombok, C.E. and Klein, G. *Naturalistic Decision Making*, Taylor and Francis, 2014.