December 2014

Observability-in-depth: An Essential Complement to the Defense-in-depth Safety Strategy in the Nuclear Industry

Francesca Favaro
Georgia Institute of Technology

Joseph Saleh
Georgia Institute of Technology

http://dx.doi.org/10.5516/NET.03.2014.021

OBSERVABILITY-IN-DEPTH: AN ESSENTIAL COMPLEMENT TO THE DEFENSE-IN-DEPTH SAFETY STRATEGY IN THE NUCLEAR INDUSTRY[1]

FRANCESCA M. FAVARÒ and JOSEPH H. SALEH*
Georgia Institute of Technology, Atlanta, GA 30332, USA
*Corresponding author. E-mail : jsaleh@gatech.edu

Received February 26, 2014
Accepted for Publication June 16, 2014

Defense-in-depth is a fundamental safety principle for the design and operation of nuclear power plants. Despite its general appeal, defense-in-depth is not without its drawbacks, which include its potential for concealing the occurrence of hazardous states in a system, and more generally rendering the latter more opaque for its operators and managers, thus resulting in safety blind spots. This in turn translates into a shrinking of the time window available for operators to identify an unfolding hazardous condition or situation and intervene to abate it. To prevent this drawback from materializing, we propose in this work a novel safety principle termed "observability-in-depth". We characterize it as the set of provisions technical, operational, and organizational designed to enable the monitoring and identification of emerging hazardous conditions and accident pathogens in real-time and over different time-scales. Observability-in-depth also requires the monitoring of conditions of all safety barriers that implement defense-in-depth; and in so doing it supports sensemaking of identified hazardous conditions, and the understanding of potential accident sequences that might follow (how they can propagate). Observability-in-depth is thus an information-centric principle, and its importance in accident prevention is in the value of the information it provides and actions or safety interventions it spurs.

We examine several "event reports" from the U.S. Nuclear Regulatory Commission database, which illustrate specific instances of violation of the observability-in-depth safety principle and the consequences that followed (e.g., unmonitored releases and loss of containments). We also revisit the Three Mile Island accident in light of the proposed principle, and identify causes and consequences of the lack of observability-in-depth related to this accident sequence. We illustrate both the benefits of adopting the observability-in-depth safety principle and the adverse consequences when this principle is violated or not implemented. This work constitutes a first step in the development of the observability-in-depth safety principle, and we hope this effort invites other researchers and safety professionals to further explore and develop this principle and its implementation.

KEYWORDS : Observability-in-depth, Accident Pathogen, Latent Failure, Defense-in-depth, Safety Blind Spot

1. INTRODUCTION

Defense-in-depth is a fundamental safety principle for the design and operation of nuclear power plants. It is the basis for risk-informed decisions by the U.S. Nuclear Regulatory Commission (NRC) [1–3]. In its bare essence, defense-in-depth consists in the design and implementation of multiple safety barriers, technical, procedural, and organizational, and whose objective is first to prevent accident initiating events from occurring, second to block accident sequences from escalating, and third to mitigate adverse consequences should the previous barriers fail. Accidents typically result from the absence, inadequacy, or breach of such defenses [4,5]. The purpose of defense-in-depth is to compensate for uncertainties, inadequacies, or incompleteness in risk analysis, and ultimately "to protect the plant, the plant operators, and the health and safety of the public" from adverse events [6]. Safety within the context of defense-in-depth should not be contingent on a single defensive element, hence the "depth" qualifier. Although there is yet no official definition of defense-in-depth by the NRC, whenever the term is used and a definition needed, one is created that is consistent with the intended use of the term [1]. There is however a general agreement on the need for defense-in-depth, its

[1] This work builds and expands on a paper presented at a PSAM conference.

* Corresponding author. Tel: + 1 404 385 6711; Email address: jsaleh@gatech.edu (J. H. Saleh)

objectives, the approach to achieve its goals, and the criteria to guide its implementation. A careful review of the different perspectives and uses of the term defense-in-depth is available in [6]. For our present purposes, the nuances in the different definitions of defense-in-depth are not relevant, and the (functional) definition provided above is sufficient for the discussion that follows, namely that defense-in-depth is defined by and embodied in the design of, and provisions for, diverse and multiple safety barriers, technical, procedural, and organizational, and whose objectives are the prevention of accident initiating events, the blocking of accident sequences, and the mitigation or containment of accident consequences[2]. It is worth clarifying that defense-in-depth can be implemented in many ways and it requires significant ingenuity—technical, operational, organizational, and regulatory—to conceive and implement in a variety of contexts and for dealing with different types of hazards and uncertainties. An attempt at formalizing defense-in-depth and quantifying its effects can be found in [7].

Despite its general appeal, defense-in-depth is not without its drawbacks [8,9]. For example, its successive lines of defense can (inadvertently) enhance mechanisms that conceal the transition of a system to an increasingly hazardous state, making "systems more […] opaque to the people who manage and operate them" [8]. As a result, system operators may be left blind to the possibility that hazard escalation is occurring, thus decreasing their situational awareness and shortening the time they have to intervene before an accident is released. **In other words, defense-in-depth may create safety blind spots and decrease situational awareness**, which in turn translate into a shrinking of the time window available for operators and decision-makers to identify an unfolding hazardous condition or situation and intervene to abate it. Several accident reports identified hidden failures and unobservable accidents pathogens as important contributing factors to the accidents, the Three Mile Island and the Texas City refinery accidents are such representative cases [10,11]. The NRC database for event reports contains about 90 cases of unmonitored release paths for contaminated air and more than 1400 cases of loss of containment[3]. This will be further examined in Section 3.

How can defense-in-depth enhance mechanisms that conceal the transition of a system to an increasingly hazardous state, and thus create safety blind spots? There are several pathways by which this can occur; we briefly note herein two intuitive modes of occurrence, a physical and a functional one:

- The physical pathway: the physical subset of safety barriers in defense-in-depth are placed between the energy source(s) and that which needs to be protected, to avoid uncontrolled release of energy, i.e., an accident, and to prevent harmful interaction between the energy source and that which is not meant to be a sink for said energy. This view is known as the energy model of accidents, and it has led to the development of several safety strategies (details can be found in [7]). The physical separation between energy source and individuals or resources to be protected can create obstacles to the observation and monitoring of the situation behind the (physical) safety barriers on the one hand, and it can obfuscate the status or condition of barriers on the other hand. For example, in underground coalmines, after a particular section is mined and abandoned (while the rest of the mine remains active), the section is sealed, and recommendations are provided for the design and strength of this safety barrier, i.e., the seal. If no additional precautions are taken, the barrier will obfuscate the conditions in the sealed area, and since ventilation is no longer available in that section, methane can accumulate and reach dangerous explosive levels[4]. In other words, the seal can create a safety blind spot and leave the miners unaware than an explosive mixture has built up behind the safety barrier. An ignition source would be the last remaining element to transform this condition into an accident. The Sago mine disaster in West Virginia (January 2006) in which 12 miners were killed was the result of such a situation [13].
- The functional pathway: defense-in-depth is intrinsically devised to slow down, minimize, or eliminate the effects of local faults or failure events on the overall safety and output of the system. Equivalently this means that defense-in-depth is meant among other things to decrease the sensitivity of the system output to local faults or failures and not carry, or highly attenuate, their "signature". This constitutes a degraded observability into the states of the system, and unless specifically addressed, it can conceal the transition of a system to an increasingly hazardous state and thus create safety blind spots.

Other or related pathways can also turn into hazard concealment mechanisms, for example when the state of

[2] The understanding of defense-in-depth should not be restricted to the physical barriers, e.g., fuel cladding and containment building.

[3] The search, executed on the Licensee Event Reports (LERs) database, for "unmonitored AND release AND path" returned 89 results, while "loss AND containment" returned 1477 results (keywords in the titles and abstracts of the reports). The database is available at https://lersearch.inl.gov/LERSearchCriteria.aspx, and was queried on 12/09/2013.

[4] Methane reaches an explosive range when its concentration in the atmosphere reaches between 5% and 15% [12].

the barrier itself is not monitored and no features are put in place to assess its condition, as in the case of the Davis Besse, which will be discussed in the next section.

To prevent this hazard-concealing potential of defense-in-depth from materializing, we propose in this work a safety principle termed "observability-in-depth". We provide the following preliminary definition of observability-in-depth as:

1. the set of provisions, technical, operational, and organizational designed to enable the monitoring and identification of emerging hazardous conditions and accident pathogens in real-time and over different time-scales;
2. the monitoring / reliable estimation of the conditions and status of all safety barriers that implement the defense-in-depth strategy (especially if they are degraded or breached);
3. the support of sensemaking of the previously identified hazardous conditions and accident pathogens, and the understanding of potential accident sequences that might follow (how they can propagate).

Observability-in-depth is fundamentally an information-centric principle, and its importance in accident prevention is in the value of the information it provides and actions or safety interventions it spurs, as we will discuss shortly. Observability-in-depth is intimately related to and supports situational awareness at the operational and organizational levels, and it allows us to conceive of a dynamic defense-in-depth safety strategy in which some defensive resources, safety barriers and others, are prioritized and re-allocated dynamically in response to emerging risks [9,14]. More details will follow in the next sections. It is worth clarifying that observability-in-depth, just like defense-in-depth, can be implemented in a variety of ways and it requires ingenuity—technical, operational, and organizational—to conceive and implement in different contexts and for dealing with different types of hazards, accident pathogens, and safety barriers.

Except in the nuclear industry where its use is more mature and dynamic than in other industries, Probabilistic Risk (or Safety) Assessment (PRA or PSA, here used interchangeably) has traditionally been performed offline and used as a static tool to help identify and prioritize various risks before system operation. The nuclear industry has two concepts that go beyond the traditional static risk analysis; they are the "Living Probabilistic Safety Assessment" or LPSA and the "risk monitors". Living PSA is defined as a safety assessment that is updated on a regular basis, and the updates are done to account for "changes in the design and operation of the plant, improvements in how the plant behaves in fault conditions, and improvements in PSA methods, models, and data" [15]. Although still an offline tool, LPSA offers an important methodological advance with respect to the traditional PSA in that its models and data are not static but exhibit some discreet dynamics and evolve by jerks (when updated); as such they are more useful and likely to reflect the actual risk status of a plant than static PSA. Risk monitors provide a more continuous monitoring of certain risks in a plant, and instead of the new aggregated averages that LPSA provide, risk monitors are meant to provide "point-in-time risk for each plant configuration, … [given] the current plant alignments, component outages, and activities being carried out that affect the risk and factors related to the plant operational state" [15]. Risk monitors are typically provided as software packages to nuclear power plants, and they are used among other things to schedule maintenance "to avoid peaks in risks [and] achieve greater flexibility in plant operation" [15]. Observability-in-depth introduces an online (real-time) mind-set into risk analysis and management, and it provides a broad strategic heading under which other tools and concepts such as LPSA, risk monitors, and prognostic and health management can be subsumed. Observability-in-depth supports the development of an online probabilistic safety assessment (not just a "living" one, as discussed previously), and this in turn can help dynamically re-order risk priorities based on emerging hazards, and re-allocate some defensive resources accordingly.

The objective of this work is to introduce the nuclear industry community to the observability-in-depth safety principle, and to make the case that it ought to be considered an important complement to the defense-in-depth safety strategy. We illustrate some of the adverse consequences when this principle is violated or not implemented using several recent "event reports" from the NRC database. By the same token, we identify the set of problems that fall within the scope of observability-in-depth and which this principle can help address or prevent. We also revisit the Three Mile Island accident and identify consequences of the lack of observability related to this accident sequence, and the failure to prevent it from unfolding in a timely manner.

This work constitutes a first step in the development of the observability-in-depth safety principle, and we hope this effort invites other researchers and safety professionals to further explore and develop this principle. The remainder of this work is organized as follows. Section 2 introduces the observability-in-depth safety principle and the many ways it can be implemented. Section 3 examines the role of observability-in-depth in the nuclear industry through detailed analyses of cases selected from the NRC database as well as the Three Mile Island accident. Section 4 concludes this work.

2. SAFETY DIAGNOSABILITY AND OBSERVABILITY-IN-DEPTH

Observability is a Control Theoretic concept, which roughly indicates how well the internal states of a system

can be inferred from the system's inputs and outputs[5]. More formally, a generic dynamical system given by Eq. (1)

$$\begin{cases} \dot{\boldsymbol{x}}(t) = F(\boldsymbol{x}(t), \boldsymbol{u}(t)) \\ \boldsymbol{y}(t) = G(\boldsymbol{x}(t), \boldsymbol{u}(t)) \end{cases} \quad (1)$$

is said to be observable if the knowledge of the set of inputs **u**(t) and the set of outputs **y**(t) – measured from some initial time t_0 – are sufficient to obtain a unique estimation of the system's state vector **x**(t) for all future instants following t_0. Equation 1 indicates a functional relationship between the evolution of the internal states of the system and the system's inputs and current states. In Control Theory, the term *state vector* has a precise formal definition and it constitutes the foundation for most analytical techniques in this field. Roughly speaking, the state vector of a system is the minimum set of variables that contain all the necessary information about the internal condition of a system at some time t_0; and that knowledge, along with the input(s) to the system (e.g., operators' inputs) is sufficient to determine the system's outputs or behavior.

Observability in Control Theory, as noted previously, is the ability to infer or estimate the internal system state from the output of the system without having to measure or "observe" that state directly (the distinction between state vector, **x**(t) and system output **y**(t) is important; see [14] and the references therein for details). Why or when is this feature relevant? It is sometimes the case in complex systems that the direct measurement and knowledge of the entire state vector is not possible due to a variety of reasons such as the lack of sensors, sensor hardware limitations, or inefficient information distribution among various subsystems of the plant. For example, in the case of an aircraft or drone tracking a nominal path, the angle of attack and the sideslip angle of the vehicle as well as the rate of change of both angles are parts of the system state (vector), since both the longitudinal and lateral dynamics of the vehicle are related to these angles and their rates of change. However, the accurate measurement of the two angles and their rates of change is not always possible. Their measurement would require, for example, accurate knowledge of the local wind conditions at every instant of time, which are often not available [14]. Observability and techniques of state estimations allow us to infer the vehicle's state without having to directly measure them.

Given this brief explanation of Observability in Control theory and our previous definition of the observability-in-depth safety principle, the reader may have noticed that the two terms present some differences. Indeed while observability-in-depth adopts its first term from Control Theory, and it is inspired by the general quest of Observability – namely figuring out or estimating the internal states of a system for (better) decision-making purposes – the safety principle is significantly broader than its counterpart in Control Theory, and it is more constraining in terms of requirements for compliance. Three important extensions are worth making explicit:

1. A system may be observable in a control theoretic sense, but if no provisions are taken to establish an "observer" of its states[6], its observability is meaningless since it is not acted upon, and as such it cannot support better decision-making. Observability-in-depth specifically requires among other things an active scanning for potential hazardous conditions and accident pathogens in a system, and that the states of all safety barriers be observed (whether directly or through estimation);
2. Another important extension is in relation to accident pathogens: accident pathogens can be thought of as adverse latent conditions or failures, which compounded with other factors can further advance an accident sequence, precipitate an accident, or aggravate its consequences. An accident pathogen is thus a distinctive causal factor in an accident sequence, and it is inactive or lurking until triggered by other factors. Accident pathogens by definition have no visible effect on the system's output under nominal operating conditions, and as such they are not observable in a control theoretic sense. Yet these are specifically what observability-in-depth is meant to scan for and identify (among other things). Since state estimation (observability in a control theoretic sense) is not possible for these conditions, direct observation and monitoring is required for accident pathogens[7].
3. Finally, observability in a control-theoretic sense deals with system's inputs and outputs to estimate the quasi-current state of the system[8]. It is thus quasi-retrospective in nature. Observability-in-depth, on

[5] The terms observability and diagnosability are used in a related manner. While there are some differences between them (in their domain of applicability and the nature of the underlying mathematical models they apply to, time-driven dynamical systems in the first case and discrete event systems in the other), these differences are not relevant for our purposes, and we will occasionally use these two terms interchangeably. Observability-in-depth remains the overarching category under which both observability and diagnosability will be subsumed.

[6] An *observer* in Control Theory is typically an algorithmic feature put in place to perform the state estimation of a system.

[7] Accident pathogens may be in the degraded states of the defense-in-depth provisions, lessening their *defensive* potential. It is important to ensure that such pathogens are continuously scanned for and monitored so that operators and decision-makers do not rely on a misleading estimation of the efficiency of available defensive resources for accident prevention.

[8] There is typically a time delay (even if it is very short) associated with and required for performing the state estimation.

the other hand, requires additional information (e.g., triggering thresholds of accidents on certain state variables), and it includes an important aspect of predicting the propagation of current states to assess potential accident sequences that might follow (sensemaking). In this sense, observability-in-depth has a significant prospective dimension.

In short, it is proper to acknowledge that observability-in-depth borrows the concept of observability from Control Theory with its emphasis on the knowledge of the internal states of a system toward an improved decision-making and control (in our case, accident prevention). However, beyond this general similarity, the two terms are different in their scope and implications: the safety principle is significantly broader than its counterpart in Control Theory; it is also more constraining in terms of requirements for compliance; and it is meant to address specific unsafe conditions (states) that are beyond the "field of view" of observability in a control theoretic sense[9].

The ability to observe or diagnose the transition of a system to a hazardous state or the occurrence of a safety-degrading event is critical for the continued safety of operations. Roughly speaking, operators make decisions during system operation, which are both based on and affect the internal conditions/states of the system [16]. If the internal system states are not reliably observed or estimated (and reported), there is a distinct possibility that operators will make flawed decisions, which in turn can compromise the safe operation of the system or fail to check the escalation of an accident sequence (e.g., no decision when an intervention is warranted).

To better illustrate the importance of the notion of observability of hazardous states, consider the following illustrative example[10]. A system departs from nominal operating conditions and begins drifting toward an increasingly hazardous state as shown in Figure 1. Various safety barriers can be interposed between the nominal operating conditions (states) and the accident release (for some specifics about this point, in ref. [11] for example, the system is a splitting tower at an oil refinery, which is filling up with hydrocarbon. The barriers are safety pressure valves and specific design features designed to contain any overflow before the accident, namely loss of containment, occurs). We represent the accident trajectory by plotting the evolution over time of the hazard level of the system, here considered loosely speaking as the closeness of the accident to being released. Assume that safety barriers are implemented to prevent the system from reaching hazard level H_0 in Figure 1, and that additional barriers are in place to block further escalation past H_1 and H_2 should the previous barriers fail or prove inadequate.

The solid line in Figure 1 represents the actual hazard level of the current state of the system, hereafter noted as *H(t)*, while the dashed line represents the operators' assumed hazard level, $\hat{H}(t)$, estimated from available information or through direct sensor observation.

The gap between these two quantities, the actual and the estimated hazard levels can be noted as:

$$\Delta H = \| H(t) - \hat{H}(t) \| \tag{2}$$

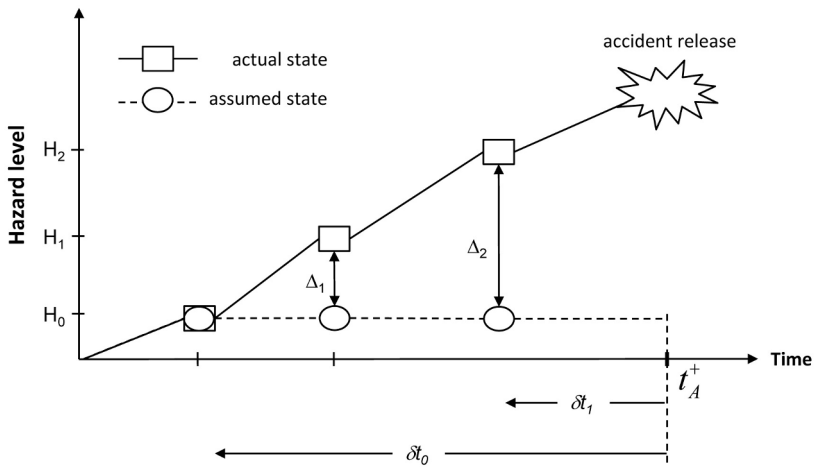

Fig. 1. Schematic Representation of System Diagnosability/Observability, Adapted from [11]

[9] The used word "observed" hereafter is not meant to be understood in its restricted control theoretic sense but its broader sense of directly observed, monitored, or reliably estimated.

[10] This is based on ref. [11] by the authors and it is included here for convenience and illustrative purposes.

This gap can result for example from the absence of observability into hazardous conditions in the system (e.g., missing sensors), or degraded observability (e.g., miscalibrated sensors or covered gauges), which can mislead the operators about the actual state of the system. Examples of these situations will be examined shortly.

In a previous work [11], we argued that all safety-degrading events or hazardous states that defense-in-depth is meant to protect against be diagnosable, that is, the failure or breach of any element in the implementation of defense-in-depth be observable—directly monitored or reliably estimated. This constitutes one aspect of the observability-in-depth safety principle. This principle implies among other things, and as a first step, that safety-critical elements in a system be properly instrumented to reflect their actual state, the extent of their degradation if any, and their breach if or when that occurs. Many examples of accidents occurred, or were not prevented in a timely manner, because of a lack of implementation of this principle. We will examine such cases in Section 3.

In light of Figure 1, the purpose of observability-in-depth is (i) to minimize the gap between the actual and the estimated hazard levels (ΔH), and (ii) to ensure that at the hazards levels associated with various safety barriers, H_0, H_1, and H_2 in the figure, the two curves coincide if these hazard levels are reached (e.g., $\Delta H = 0$ if H_0 is reached—the safety barriers designed to prevent the system from reaching H_0 is breached). The end-objective is to provide sufficient time for the operators and decision-makers to understand an unfolding hazardous situation and intervene in a timely manner to abate it. By contrast, a persistent gap between the actual and the estimated hazard levels, as shown in Figure 1, leaves the operators and decision-makers blind to the developing hazardous situation, and it shrinks the time window, and options, available to intervene.

A gap between the actual and the estimated hazard reflects degraded situational awareness. In [11], we noted that

> *the concept of situation awareness involves an operator's comprehension of a dynamic situation that he/she is monitoring or controlling [17; 18]. It is an important construct in cognitive engineering and is meant to capture, among other things, the operator's "understanding of the state of the environment, including relevant parameters of the system" [19].*

As such, observability-in-depth is intimately related to situational awareness, and it supports one important subset of the latter, namely the awareness of the occurrence of hazardous states in the system, and the understanding of the potential accident sequences that might follow.

Consider the following example, which highlights one potentially catastrophic consequence of the lack of observability-in-depth with respect to a particular safety element in a nuclear power plant. The incident occurred at the Davis-Besse Nuclear Power Station. In March 2002 a cavity of about 20-30 square inches was discovered in the reactor lid during an inspection targeted for reactor pressure vessel (RPV) head penetration (VHP) nozzle cracking due to primary water stress corrosion [20]. The discovery of the cavity was in a sense fortuitous:

> *"During these inspections, the licensee discovered cracks in several VHP nozzles. Subsequent to the machining process to repair VHP Nozzle 3, the nozzle was observed to displace, or tip in the downhill direction as the machining apparatus was withdrawn. The displacement led DBNPS personnel to examine the region adjacent to VHP Nozzle 3" [20].*

The cavity "extended completely through the 6.63 inch thick carbon steel reactor pressure vessel head down to a thin internal liner of stainless steel cladding" [20]. The degradation and breach of the reactor lid developed over an extended period of time unbeknown to the operators and plant managers. It was due to corrosion from a leak of boric acid. This lack of observability of the state or degradation of the reactor pressure vessel head barrier could have resulted in a massive loss of coolant with potential meltdown of the reactor [20]. This was a serious near miss, and the only element that prevented an accident from occurring was the internal cladding, which withstood the primary system pressure over the cavity during system operation and was neither designed for nor qualified to perform such function [20].

There are a number of lessons to be learned from this near miss at the Davis-Besse power plant, and many recommendations were provided in the NRC report [20], including for example heightened regulatory oversight of the plant. In addition to the specific recommendations provided, we propose that this and many other similar near misses support a more general recommendation, namely the adoption of the observability-in-depth safety principle, which was violated in this case, and whose implementation could have identified the degradation of this RPV safety barrier in a more timely manner.

Observability-in-depth can be implemented in many ways, and it requires creativity and technical ingenuity to design and implement in a variety of contexts and for monitoring different types of hazards and states of safety barriers. Regulations cannot be prescriptive in this regard, but a safety case can be required from the designers/operators to demonstrate compliance with this principle.

The "depth" qualifier in observability-in-depth serves two purposes, and it has both a temporal and causal dimension, as explained next.

First "depth" is used to distinguish the safety principle from the control theoretic concept of observability, and without which some confusion might arise as to the use and meaning of the term, especially among readers who are Control-literate (see earlier discussion on the similarities and differences between these two terms). Second "depth" in observability is meant to provide a parallel to "depth" in defense-in-depth in the following sense: whatever

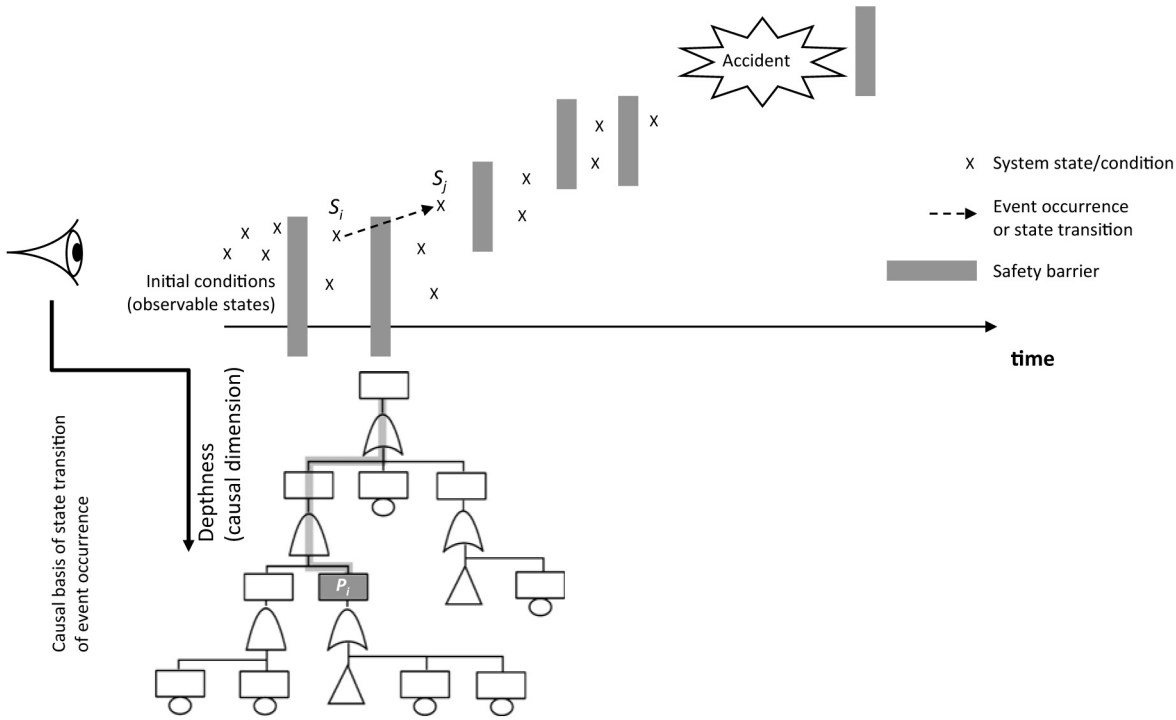

Fig. 2. Schematic Illustration of an Accident Sequence, Defense-in-depth, and the Causal Dimension of "Depth" in Observability-in-depth

"level" of depth safety barriers are placed at (in a potential accident sequence), observability-in-depth is meant to chaperone each safety barrier to monitor its condition and status.

As noted previously, observability-in-depth characterizes the ability to identify adverse states and conditions far upstream (early) in an accident sequence. It also reflects the ability to observe emerging accident pathogens and latent failures before their effect becomes manifest on the system's output, or before an increasingly hazardous transition occurs in an accident sequence. "Depth", as a result, has both a temporal and a causal dimension.

To appreciate its temporal dimension, consider Figure 1 and assume that an initiating event triggers an accident sequence. During the accident sequence, the time of occurrence of the accident is unknown and preferably right-censored (i.e., it will be averted). We noted this time in Figure 1 as t_A^+. Looking back form the vantage point of the time of occurrence of the accident (t_A^+ as the origin of the new clock), the further away we can identify hazardous states or transitions, the more depth we have in the implementation of our safety principle. In other words, the temporal dimension of observability-in-depth is reflected in the δt shown in Figure 1, with δt_0 reflecting more depth of observability than δt_1.

To appreciate the causal dimension of "depth" in observability-in-depth, consider Figure 2, which represents a set of safety barriers and various hazardous states. Each hazardous transition/escalation in an accident sequence has a set of underlying causes, and Figure 2 includes the underlying causes of a transition from S_i to S_j in the form of a Fault Tree.

The condition P_i in the fault tree is a latent failure or accident pathogen [21]; it does not have a visible effect on the system behaviour or operation, until the second condition in its AND gate occurs. If the system reaches state S_i, the hazardous transition to S_j will occur, thus further advancing the accident sequence. The ability to observe such latent causal factors or accident pathogens in an accident sequence before they have a visible effect on the system operation is another aspect of the *depthness* of observability. In other words, the "further down" a fault tree are adverse conditions identified, the more depth there is to the observability-in-depth principle.

3. EXAMPLES OF ADVERSE CONSEQUENCES WHEN OBSERVABILITY-IN-DEPTH IS COMPROMISED OR NOT IMPLEMENTED

In this section we provide a few examples that illustrate some of the adverse consequences that can follow from the lack of, or degraded, observability into hazardous conditions. The purpose is to show both the importance

of observability-in-depth by examining cases when it is not implemented, and by the same token to highlight the set of problems that it can help address or prevent. We begin with the well-known Three Mile Island accident and examine it from this perspective of observability-in-depth (or deficiencies in). Then we discuss several "event reports" from the U.S. Nuclear Regulatory Commission database, which reflect potential concealment of accident pathogens in the lines of defense.

3.1 The Archetype Case Study: the Three Mile Island Accident

The Three Mile Island (TMI) accident of March 1979 is perhaps the most famous incidents in the history of nuclear power plants in the United States. A complex sequence of events led to the loss of the water-coolant, which resulted in a partial core meltdown [10] and caused over $2 billion in damages [22].

The chain of causality leading to the accident has been widely discussed, see for example [10, 23, 24], and the accident became the subject of numerous debates for the complexity of the sequence of events starting from a leaky valve and emergency pump shutdown and leading to the reactor partial meltdown. The accident resulted from a combination of factors, including four separate malfunctions in the internal and external cooling circuits, overall sloppy maintenance and organizational deficiencies[11] [10, 23], and operators' errors. Our purpose here is not to revisit the accident sequence, but to examine it from one particular perspective, namely that of observability-in-depth, and to highlight how deficiencies in the implementation of this principle contributed to the accident sequence or failed to prevent its escalation. Some brief technical knowledge is required for our discussion. A schematic representation of the reactor core with the cooling system circuits is shown in Figure 3.

The heat generated by the reactor core at the TMI plant was removed by a heat exchanger at the intersection of two cooling circuits: a primary internal circuit directly connected to the reactor core, and a secondary external circuit connected to steam turbines (see Figure 3). Main pumps as well as emergency backups and pressure relief valves existed for both the internal and external circuits. Steam downstream of the heat exchanger drove the turbines (the power generation elements). This particular design, as well as the main pumps and emergency backups, and the pressure relief valves are specific elements in the implementation of defense-in-depth. And while they were particularly important for the safe operation of the plant, the fact that observability-in-depth was lacking or compromised in their design, as we will discuss shortly, rendered this a defense-blind strategy. Moreover, the inability to observe and assess the states of some of these safety barriers not only failed to prevent the escalation of the accident sequence, but also directly contributed to its advancement. In short, we argue that the Three Mile Island accident was to a large extent the result of a violation of the observability-in-depth safety principle, and while its proper implementation would

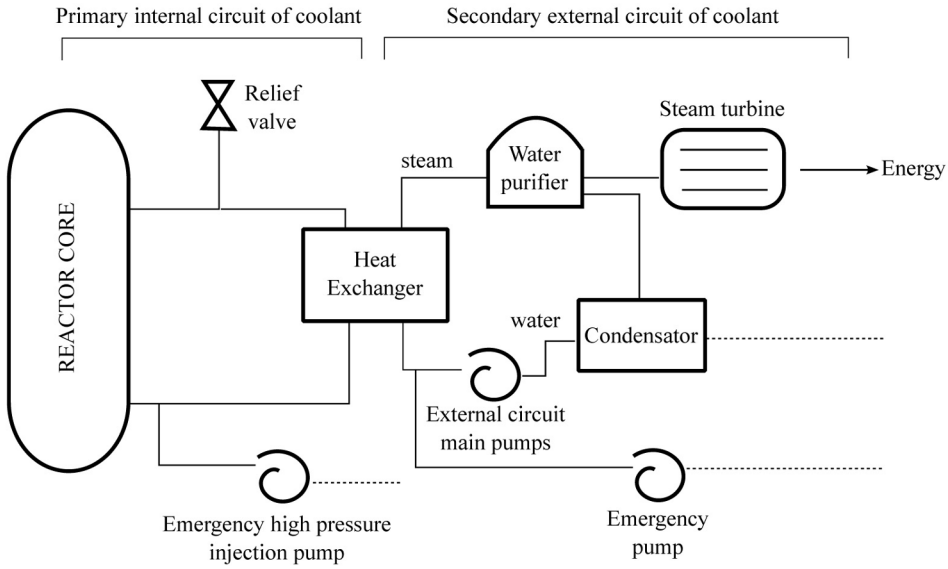

Fig. 3. Simplified Schematic of the Reactor Core and the Cooling Circuits, Adapted from [26]

[11] Gorinson et al. [25] and Hopkins [10] highlight how events similar to those indicated in Figure 4 had occurred in an incident 18 months earlier at the Davis-Besse reactor. Also previous failures of the relief valves had been witnessed in reactors manufactured by the same firm of the TMI plant. These and other near misses and warning signs apparently went unnoticed by the management of the TMI nuclear reactor.

not have prevented the initiating events from occurring—some of the factors noted previously were directly responsible for this, namely technical failures, sloppy maintenance, and organizational deficiencies—it would have ensured that the accident sequence was terminated in a timely manner before core meltdown.

The accident sequence mainly concerned the primary and secondary cooling circuits of the reactor core. A simplified overview of the events that led to the reactor core partial meltdown in shown in Figure 4.

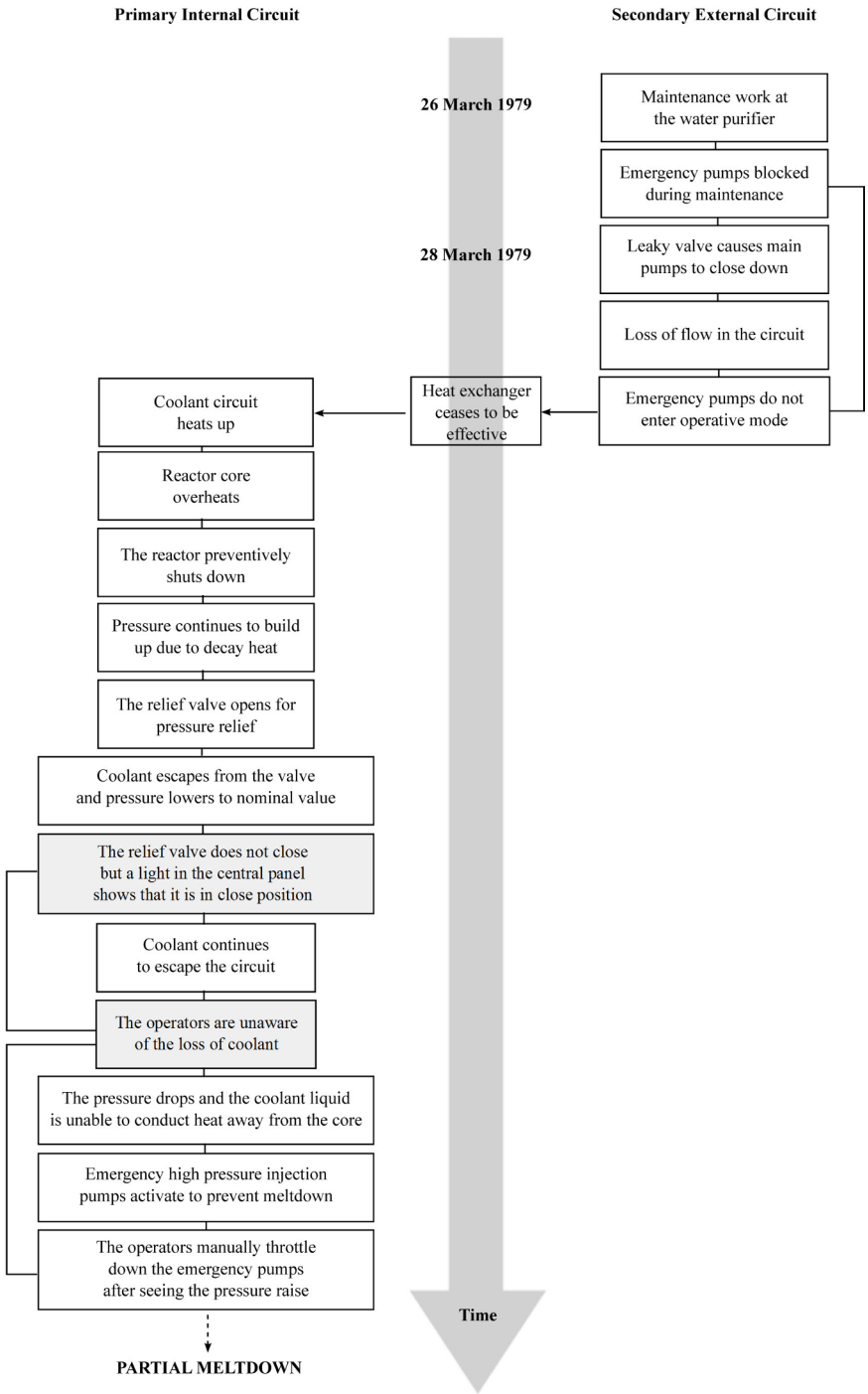

Fig. 4. Simplified Accident Sequence

The accident sequence was triggered by a leak in the external cooling circuit, which caused the main pumps to shut down [10]. Two days prior to this event, the emergency pumps of the external circuit had been shut down for maintenance work and were still inoperable. This condition, apparently unknown to the operators at the time of the accident (first unobserved latent adverse condition), led to the impossibility to dissipate the heat from the internal circuit. As a result, the reactor core began to overheat, leading to its preventive shutdown. However, the pressure in the internal circuit kept increasing due to "decay heat" from the reactor core [10]. At this point, the emergency relief valve of the internal circuit opened, letting the coolant escape and lowering the pressure to the nominal value.

The relief valve should have closed when the pressure fell to proper levels, but it became stuck open. Instruments in the control room however indicated that the valve was closed [26]. The decrease in pressure activated the emergency high-pressure injection pumps in the internal circuit to prevent the core meltdown [10]. After noticing the pressure rise in the internal circuit, the operators were unaware of the loss of coolant from the internal circuit. This is the second major unobserved condition in this accident sequence, and it was due not only to the flawed sensor that was monitoring the status of the relief valve, but also to the absence of provisions to directly monitor and estimate the coolant flow in the primary internal circuit[12]. **We conceive of this situation as a gross violation of the observability-in-depth safety principle—two major elements in the implementation of defense-in-depth were not properly monitored and their status not directly observable.**

The operators, still unaware of the loss of coolant from the internal circuit, manually throttled down the emergency pumps. This was considered in hindsight as a significant operator error and it led directly to the accident—the reactor's (partial) core meltdown. However, as shown here, this decision was the result of flawed or missing information that degraded the operators situational awareness and failed to convey the hazardous conditions of various safety barriers. It took them about 2 hours and 20 minutes to understand that a loss of coolant accident (LOCA) was ongoing. The total meltdown was then prevented by flooding the reactor core with cold water. While this extreme measure prevented the release of radioactive material, major irreversible damage had already been done [10].

[12] We thank a reviewer for providing the following clarification, that for the TMI power plant, indicators of the auxiliary the feedwater block valves position existed. However, tags covered those indicators – an inadvertent outcome of bad practice. The impossibility for the operators to observe the correct position of the valves, whether stemming from design flaws, bad practice, or other mechanisms, constitutes a violation of the OID principle. These mechanisms contribute to the incorrect estimation of the plant condition, hence deteriorating the operator situational awareness.

Different authors have debated at length the controversial issue of "operator error" in the early termination of the high-pressure injection pumps [10, 23]. Hopkins points out that, "had the pumps been allowed to continue operating, the accident could have been avoided" [10]. The "design flaws" of the relief valve and its monitoring system caused the control room to receive an incorrect signal of its position. The operators then acted on this incorrect understanding of the plant condition. To the best of our knowledge, there is no explanation in the literature as to why this condition unfolded.

The impossibility to monitor and diagnose an ongoing LOCA from the relief valve status is not the only violation of observability-in-depth. For instance, there was no instrument that allowed the operator to understand how much water covered the reactor core [26]. The time history of the water level could have improved the situational awareness of the operators and their understanding of the actual hazard level reached by the reactor.

Perrow choose this accident as the archetype of his "normal accident theory", where an accident "is termed *normal* because it is inherent in the characteristics of tightly coupled, complex systems and cannot be avoided" [23]. The normal accident argument, and specifically its applicability to the TMI accident, was criticized by Hopkins [10]. In his work, Hopkins provides a careful analysis of Perrow's point of view and notes that "Perrow claim[ed] that the information available to the operators [was] so flawed that there was no way they could have been expected to understand what was going on and react in an effective manner" [10]. Perrow's conclusion based on this observation is that the accident was indeed a "normal" occurrence, in the sense that its incomprehensible nature made prevention impossible. We agree that the flawed and missing information about the status of critical safety elements at TMI degraded the operators' situational awareness and hampered their ability to safely operate the plant. However instead of conceiving and accepting this and similar accidents as "normal", we trace back one important element in their causal chain, namely the lack of observability of emerging hazardous states, and we conceive of a safety principle, observability-in-depth, whose implementation can help prevent similar occurrences.

3.2 Observability-in-depth and the NRC Database of Licensee Event Reports

In this subsection we discuss several event reports from the NRC's Licensee Event Reports (LER) database, which illustrate more situations that can results from the lack of, or degraded, observability into hazardous conditions. By the same token, theses examples highlight an additional set of problems that fall within the scope of observability-in-depth and which this safety principle can help address or prevent.

Caveat: It is important to note that this subsection

does not constitute nor should it be considered a basis for statistical analysis of the problem of lack of observability of adverse conditions in the LER database—although this would be an interesting topic and a fruitful venue for future research. This subsection is merely for illustrative purposes and to better delineate the scope and extent of observability-in-depth.

The NRC has required nuclear power plants to submit LER since 1980, and more than 51,000[13] of these reports have since been submitted. Commercial nuclear reactor licensees are required to report certain event information when adverse conditions occur in a nuclear power plant, which are beyond its technical specifications [27, 28]. For example, the malfunction of a required safety barrier or the discovery of a potential design flaw would trigger the need for an LER. Once an LER is submitted, NRC staff review it to understand and confirm the licensee's assessment of the situation. NRC staff experts also determine whether the licensee's resolution of the issue continues to maintain adequate levels of safety and protection of the public [29]. The NRC provides public online access to the LER database. Each report consists of an abstract, a description of the events sequence, the event significance and implications, the identified causes, the implemented corrective actions, and additional information (e.g. information on similar previous occurrences).

Compared with the case study approach in Subsection 3.1, event reports (LER) allow the discussion of a broader set of situations, as the events reported are usually less serious in terms of their consequences and their occurrence relatively straightforward (or not as involved as in the TMI accident).

- *CASE I – Inoperable emergency diesel generator due to low fuel oil in storage tank [30]*: this case resulted from incorrect readings of the level of fuel oil contained in the storage tank of an emergency diesel generator. A low level of fuel oil (below the required minimum) was discovered during an inspection and investigation revealed that incorrect readings had been going on for more than a month. According to the report "the primary cause was a challenging method for determining tank level" [30]. The level indicator reading was also susceptible to exogenous disturbances, becoming "more unreliable under adverse conditions (e.g., poor weather, low light conditions)" [30]. Contributing factors included also a malfunctioning tank level indicator and the corresponding alarm. The investigation highlighted the "inadequate instrument design" of the fuel oil tank level alarm and the indicator.

  While this situation did not pose a considerable threat to the safety of the plant, it constituted a latent failure or adverse pre-existing condition, which when compounded with other factors, could have further advanced an accident sequence, for example if the emergency generators were called upon. As such, this condition constitutes a non-negligible accident pathogen [21]. The fact that it was not observable or its observability compromised is an instance of failed implementation of the observability-in-depth principle (specifically in this case a redundant safety barrier was inoperable and its breach was not monitored or reliably observable).

- *CASE II – Unmonitored Flowpath in Safety Injection Cooling Pumps [31]*: In this case a review of the pump testing surveillances showed that unmonitored flowpaths existed for different pumps, including the safety injection pump of the cooling system at the Millstone Nuclear Power Station. According to the report, the regulations current at the time of the system design, did not "explicitly require" [31] to monitor the total pump flow. The unmonitored flowpaths diverted flow from the pump discharge prior to the point at which the flow measurement was taken, hence resulting in a condition of compromised observability of the pump flow.

  At the time of the instrumentation design, the potential impact of unmonitored flowpaths on the ability to test in order to detect pump degradation was not fully realized [31]. Indeed, unmonitored flowpaths have the potential to mask the detection of pump degradation by altering the expected measurements of flow and differential pressure. As with the previous case, this condition compromises the observability of the state of the pump, thus allowing an accident pathogen to emerge and go unnoticed.

- *CASE III – Design Deficiency - Potential for an Unmonitored Release Path [32]*: In this case an unmonitored release path of contaminated air was identified during an engineering evaluation of the station service water system circuit. The identified condition would allow "contaminated air to enter the service water piping [...] and to subsequently be released to the outside environment" in case of a Loss of Power/Loss of Coolant Accident event, thus resulting in a loss of secondary containment [32].

  The failure to identify this release path was considered in the report as non-compliance with General Design Criteria. Moreover, it shows the inability to observe a potential accident sequence. In other words, in case of loss of secondary containment through this particular path, the operators would not be able to identify the release of the contaminated air to the outside environment. Observability-in-depth was in this case violated by this particular release path.

[13] Query executed on 12/10/2013.

Table 1. Selected LER – search Keywords and Scenarios Summary

Case ID and Event Report #	Keyword	Highlighted Scenario
I - 3521996022	Malfunctioning Indicator	Compromised/degraded observation
II - 4231998027	Unmonitored	Compromised observation
III - 3541997025	Unmonitored	Unobserved
IV - 2451997037	Unmonitored	Unobserved

- *CASE IV – Unmonitored Release Path Due to Radioactive Ash [33]*: In this case an unmonitored release path of contaminated ash was identified during the preparations to put a heating boiler into service for the winter season. The event report established that "if the ash on the fire side of the boiler contains radioactive constituents, some of the particulate matter could be discharged through the boiler exhaust" [33].

  This event may appear less severe and unremarkable compared with the previous ones. But the interesting point here is that the ashes in the boiler resulted from an original contamination and leak that occurred 25 years before the discovery of the unmonitored release path. This constitutes an interesting example not only of the unobserved accident pathogen, but also of the lack of a defense barrier against the release of the contaminated ashes.

These cases only represent the tip of the iceberg of instances of adverse conditions that can be gleaned from the LER database, and which can be considered in some ways instances of violation of observability-in-depth. Further examination of this database for events that include unobserved adverse conditions and breaches of safety barriers would be a fruitful venue for future research.

4. CONCLUSION

To prevent the hazard-concealing potential of defense-in-depth from materializing, and more generally to introduce a real-time mind-set into risk analysis and management, we proposed in this work a safety principle termed observability-in-depth, which helps focus attention on these issues. We defined it as:

1. the set of provisions, technical, operational, and organizational designed to enable the monitoring and identification of emerging hazardous conditions and accident pathogens in real-time and over different time-scales;
2. the monitoring / reliable estimation of the conditions and status of all safety barriers that implement the defense-in-depth strategy (especially if they are degraded or breached);
3. the sensemaking of the emerging hazardous conditions and the understanding of potential accident sequences that might follow (and how they can propagate).

In this sense, observability-in-depth should be thought of as a complement to the well-established defense-in-depth safety strategy, without which the latter can devolve into a defense-blind safety strategy. Observability-in-depth is thus fundamentally an information-centric principle, and its importance in accident prevention is in the value of the information it provides and actions or safety interventions it spurs.

One objective of observability-in-depth is to minimize the gap between the actual and the estimated hazard levels in a system in real-time, and in so doing to provide sufficient time for the operators and decision-makers to understand an unfolding hazardous situation and intervene in a timely manner to abate it. As such, we proposed that observability-in-depth is intimately related to situational awareness, and it supports one important aspect of the latter, namely the awareness of the occurrence of hazardous states in the system in real time, and the understanding of the potential accident sequences that might follow (sensemaking of the emerging hazardous conditions). The hazardous states can be technical, operational, or organizational. We explained that the *depth* qualifier in our principle has both a causal and a temporal dimension, and it is meant to characterize the ability to identify adverse states and conditions far upstream in an accident sequence.

Changing mind-sets: Probability Risk Assessment (PRA) has traditionally been performed offline and used as a static tool to help identify and prioritize various risks before system operation (see caveat in the Introduction about the status of PRA in the nuclear industry). Similarly, defense-in-depth has to some extent an implicit static connotation. Observability-in-depth introduces a real-time mind-set into risk analysis and management, and it supports the development of an online probabilistic risk assessment, which in turn can help dynamically re-order risk priorities based on emerging hazards, and re-allocate some defensive resources accordingly. As such, observability-in-depth can help conceive of a **dynamic defense-in-depth safety strategy** in which some defensive resources, safety barriers and others, are prioritized and allocated dynamically in response to emerging risks.

Observability-in-depth extends beyond specific techniques, alarms, and instrumentations, and it provides a broad strategic heading under which tools and techniques such as living probabilistic safety assessment (LPSA), risk monitors, and prognostic and health management can be subsumed.

This work constitutes a first step in the development of the observability-in-depth safety principle, and we hope this effort invites other researchers and safety professionals to further explore and build on this principle. We believe some fruitful venues for further research include examining:

i. the implementation of observability-in-depth in specific contexts and at multiple scales, at the technical and operational plant levels, and at the industry level (different time scales will also have to be accounted for);
ii. the extent to which observability-in-depth is required in specific contexts, and what changes (additions, deletions, modifications, etc.) would be required to achieve appropriate levels;
iii. the integration of observability-in-depth and its outputs into existing (living) probabilistic safety assessments;
iv. the policy implications of observability-in-depth, and the planning of a test case or pilot project for its implementation;
v. (on a more theoretical level) the relationship / interplay between the three constructs: observability-in-depth, situational awareness, and sensemaking.

REFERENCES

[1] Sørensen, J. N., Apostolakis, G. E., Kress, T. S., and Powers, D. A. *"On the Role of Defense in Depth in Risk-Informed Regulation*". In: Proceedings of the PSA '99, 1999.

[2] NRC, US. "*Causes and Significance of Design Basis Issues at US Nuclear Power Plants*". Draft Report, Washington, DC: US Nuclear Regulatory Commission, Office of Nuclear Regulatory Research, 2000.

[3] Saleh, J. H., Marais, K. B., Bakolas, E. and Cowlagi, R. V. "*Highlights from the literature on accident causation and system safety: Review of major ideas, recent contributions, and challenges*." Reliability Engineering & System Safety, Volume 95, Issue 11, pp. 1105-1116, 2010.

[4] Rasmussen, J. "*Risk management in a dynamic society: a modeling problem*". Safety Science, Volume 27, Issues 2-3, pp. 183–213, 1997.

[5] Svedung, I., and Rasmussen, J. "*Graphic representation of accident scenarios: mapping system structure and the causation of accidents*". Safety Science, Volume 40, Issue 5, pp. 397–417, 2002.

[6] NRC-ML13277A421, Enclosure 3 (2013). Defense-in-depth observations and detailed history. Nuclear Regulatory Commission, Washington, D.C. *Available at http://pbadupws.nrc.gov/docs/ML1327/ML13277A425.pdf [Accessed May 13, 2014]*

[7] Saleh, J.H., Marais, K. B., Favarò, F.M. "System safety principles: A multidisciplinary engineering perspective". Journal of Loss Prevention in the Process Industry, vol. 29, 2014, pp. 283–294.

[8] Reason, J. T. "*Managing the risks of organizational accidents*". Aldershot, Hants, England; Brookfield, Vt., USA: Ashgate, 1997.

[9] Favarò, F. M., and Saleh, J. H. "*Observability in Depth: novel safety strategy to complement defense-in-depth for dynamic real-time allocation of defensive resources*". Presented at the ESREL Conference September 29 – October 2 2013, Amsterdam, 2013.

[10] Hopkins, A. "*Was Three Mile Island a 'Normal Accident'*?". Journal of Contingencies and Crisis Management, Volume 9, Issue 2, pp. 65-72, 2001.

[11] Saleh, J. H., Haga, R. A., Favarò, F. M., Bakolas, E. (2014a) "*Texas City Refinery Accident: Case Study in Breakdown of Defense-In-Depth and Violation of the Safety-Diagnosability Principle*". Engineering Failure Analysis, Volume 36, pp. 121-133, 2014.

[12] Saleh, J. H., Cummings, A. M. "*Safety in the Mining Industry and the Unfinished Legacy of Mining Accidents: Safety Levers and the Principle of Defense-in-Depth for Addressing Mining Hazards*." Safety Science, Vol. 49, Issue 6, 2011, pp. 764–777.

[13] Gates, R.A., Phillips, R.L., Urosek, J.E., et al. "*Report of investigation: Fatal underground coal mine explosion, January 2, 2006. Sago Mine*". Mine Safety and Health Administration, 2007. Available at http://www.msha.gov/sagomine/sagomine.asp. [Accessed May 13, 2014]

[14] Bakolas, E., and Saleh, J. H. "*Augmenting defense-in-depth with the concepts of observability and diagnosability from Control Theory and Discrete Event Systems*". Reliability Engineering & System Safety, Volume 96, Issue 1, pp. 184-193, 2011.

[15] OECD/NEA "CSNI Technical Opinion Papers: (7) *Living PSA and its Use in the Nuclear Safety Decision-making Process; (8) Development and Use of Risk Monitors at Nuclear Power Plants*." Nuclear Energy Agency, Paris, France, 2005 Available at http://www.oecd-nea.org/nsd/reports/2005/nea_4411-PSA-risk-monitors.pdf [Accessed May 12, 2014]

[16] Le Bot, P. "*Human reliability data, human error and accident models- illustration through the Three Mile Island accident analysis*." Reliability Engineering and System Safety. Volume 83, No. 2, pp. 153-167, 2004.

[17] Endsley, M. R. "*Toward a theory of situation awareness in dynamic systems*." Human Factors: The Journal of the Human Factors and Ergonomics Society, Vol. 37, No. 1, pp. 32-64, 1995a.

[18] Durso, F. T., and Sethumadhavan, A. (2008). "*Situation awareness: Understanding dynamic environments*." Human Factors: The Journal of the Human Factors and Ergonomics Society, Vol. 50, No. 3, pp. 442-448, 2008.

[19] Endsley, M. R. "*Measurement of situation awareness in dynamic systems*". Human Factors: The Journal of the Human Factors and Ergonomics Society, Vol. 37, No. 1, pp. 65-84, 1995b.

[20] NRC, US "*DAVIS-BESSE REACTOR VESSEL HEAD DEGRADATION LESSONS-LEARNED TASK FORCE REPORT*" available at www.nrc.gov/reactors/operating/ops-experience/vessel-head-degradation/lessons-learned/lessons-learned-files/lltf-rpt-ml022760172.pdf

[21] Saleh, J. H., Saltmarsh, E., Favarò, F. M., Brevault, L. "*Accident precursors, near misses, and warning signs: critical review and formal definition within the framework of Discrete Event Systems*". Reliability Engineering and System Safety, Volume 114, pp.148-154, 2013.

[22] Sovacool, B. K. "*The costs of failure: a preliminary*

assessment of major energy accidents, 1907–2007." Energy Policy, Volume 36, No. 5, pp. 1802-1820, 2008.
[23] Perrow, C. "*The President's Commission and the normal accident*." Accident at Three Mile Island: The Human Dimensions pp. 173-84, 1982.
[24] Rogovin, M. "*Three Mile Island: A report to the Commissioners and to the public*". No. NUREG/CR-1250 (Vol. 1). Nuclear Regulatory Commission, Washington, DC (USA), 1979.
[25] Gorinson, Stanley, M., and Kane, K. P. "*Report of the Office of Chief Counsel on the role of the managing utility and its suppliers*". No. NP-25106. President's Commission on the Accident at Three Mile Island, Washington, DC (USA), 1979.
[26] NRC, US "*Backgrounder on the Three Mile Island Accident*", available at http://www.nrc.gov/reading-rm/doc-collections/fact-sheets/3mile-isle.html
[27] NRC, US – LERSearch database website, available at https://lersearch.inl.gov/Entry.aspx
[28] NRC, US "10CFR50.73 *Licensee Event Report System*" available at http://www.nrc.gov/reading-rm/doc-collections/cfr/part050/part050-0073.html
[29] NRC, US public blog "*Easy Searching for Licensee Event Reports*" available at http://public-blog.nrc-gateway.gov/ 2011 /03/04/easy-searching-for-licensee-event-reports/
[30] NRC, US "*Licensee Event Report 96-022-00 Emergency Diesel Generator Inoperable Due to Low Fuel Oil in Storage Tank*" Limerick Geberating Station, Unit 1, December 31st 1996.
[31] NRC, US "*Licensee Event Report 98-027-00 Unmonitored Flowpath in Safety Injection Cooling Pumps May Prevent Detection of Pump Degradation*" Millstone Power Station Unit 3, April 16th 1998.
[32] NRC, US "*Licensee Event Report 97-025-00 Design Deficiency - Potential for an Unmonitored Release Path Through the Station Service Water System*" Hope Creek Generating Station, October 4th 1997.
[33] NRC, US "*Licensee Event Report 97-037-00 Unmonitored Release Path Due to Radioactive Ash in the House Heating Boiler*" Millstone Power Station Unit 1, September 10th 1997.