A method of identifying and analyzing irrational system behavior in a system of systems

Van Bossuyt, Douglas L.; O'Halloran, Bryan M.; Arlitt, Ryan M.

Published in:
Systems Engineering

Link to article, DOI:
[10.1002/sys.21520](https://doi.org/10.1002/sys.21520)

Publication date:
2019

Document Version
Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):
Van Bossuyt, D. L., O'Halloran, B. M., & Arlitt, R. M. (2019). A method of identifying and analyzing irrational system behavior in a system of systems. *Systems Engineering*, 22(6), 519-537.
<https://doi.org/10.1002/sys.21520>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

A Method of Identifying and Analyzing Irrational System Behavior in a System of Systems

Dr. Douglas L. Van Bossuyt¹ | Dr. Bryan M. O'Halloran¹ |
Dr. Ryan M. Arlitt²

¹Department of Systems Engineering,
Naval Postgraduate School, 777 Dyer Road,
Monterey, California, 93943, USA. Phone:
+01 831-656-7572, +01 831-656-2430.

²Department of Mechanical Engineering,
Technical University of Denmark, Lyngby,
Denmark. Phone: +45 256250

Correspondence

Dr. Douglas L. Van Bossuyt, Department of
Systems Engineering, Naval Postgraduate
School, Monterey, California, 93943, USA
Email: douglas.vanbossuyt@nps.edu

Funding information

System of interest (Sol) failures can sometimes be traced to an unexpected behavior occurring within another system that is a member of the SoS with the Sol. This article presents a method for use when designing a Sol that helps to analyze a SoS for unexpected behaviors from existing SoS members during the Sol's conceptual functional modeling phase of system architecture. The concept of irrationality initiators – unanticipated or unexpected failure flows emitted from one system that adversely impact a Sol which appear to be impossible or irrational to engineers developing the new system – is introduced and implemented in a quantitative risk analysis method. The method is implemented in the Failure Flow Identification and Propagation (FFIP) framework to yield a probability distribution of failure paths through a Sol in the SoS. An example of a network of autonomous vehicles operating in a partially denied environment is presented to demonstrate the method. The method presented in this paper allows practitioners to more easily identify potential failure paths and prioritize fixing vulnerabilities in a Sol during functional modeling when significant changes can still be made with minimal impact to cost and schedule.

KEYWORDS

System of Systems, Systems Engineering, Risk Analysis, Failure Analysis, System Modeling, Irrational System Behavior, Irrationality

1 | INTRODUCTION

In spite of extensive efforts undertaken during the design of systems, system failures continue to occur regularly. This is demonstrated by a multitude of system failure examples making headline news. Over a period of 52 years (1957-2009), there were over 400 publicly documented mission failures in the space industry including satellites, crewed spacecraft, and rockets [1]. Since the introduction of the commercial airline industry, there have been a reported 154,984 deaths as the result of 26,152 accidents [2]. According to [3], there have been a total of 25 major dam failures documented, 16 of which have occurred in the last 50 years. The nuclear power industry has observed over 200 significant failures, several of which have resulted mitigations exceeding one billion US dollars [4]. Recent events in the aviation industry [5] emphasize that failures occur even in newly designed systems with strong regulatory oversight. In short: systems routinely fail regardless of system type, purpose, age, design approach used, industry, or the era in which it was designed and built. Regardless of our best design and analysis of systems, we as practitioners and researchers continue to be surprised by emergent (unpredicted, unexpected, discounted, or seemingly illogical or irrational) system failures. While we would like to believe that the systems we design will behave exactly how we predicted and observed during the design and testing portion of the systems engineering process, the literature and the popular press show that this is often not true.

Within the context of system failures, harmful emergent system behaviors have been observed in engineered systems for many decades [6]. Over time, more simple-to-understand emergent system behaviors have been corrected for and are no longer a significant issue [7]. However, efforts to systematically understand the underlying causes of the emergent behaviors and design systems to minimize the potential for harm have only been undertaken in the last few decades [8, 9]. The majority of industry work and academic research has focused on events that have previously been observed, and expected and predictable events [10]. As a result of efforts to address such events, modern systems are much less likely to fail from single point failures or from commonly occurring failures caused by multiple component failures; such failures have largely been identified and corrected [11, 12]. The failures that are now observed in systems are often as a result of multiple failure events occurring together to develop an emergent system behavior that has previously not been predicted or observed [13], or which had been ruled out through previous analysis as unlikely to occur [14]. For example, a recent collision on Singapore's Mass Rapid Transit system resulted in 38 injuries and was a result of a series of unexpected interactions across multiple systems and subsystems in the signaling system which led to a series of undetected and progressively degraded operation conditions [15].

We believe that if the systems engineering community wishes to continue to increase the robustness, resilience, inter-operability, and survivability of SoSs in an effort to improve the probability of an SoS completing its mission successfully, a better understanding of how failures are initiated in systems of interest (Sols) by other members of a SoS that lead to Sol failure is needed. Already, there have been some efforts that begin to address the problem with most focusing on the more cost effective conceptual phase of system engineering when architectural trade-off studies of potential system architectures are conducted and before significant component design work has begun [16]. For instance, there is a proposed method to analyze SoS early on in the design cycle via SoS modeling [17]. There are many different ways to model SoS such as the Functional Basis for Engineering Design (FBED) [18] functional modeling taxonomy which can be used to make functional architectures. A better understanding of what potential failure events have a higher likelihood of occurrence can help determine priorities for mitigating such potential failure events [19]. One way to model potential failure events is to use functional models to predict the likelihood of failure of a system [20]. An SoS architecture can be iterated many times until an acceptable system failure probability has been reached [21]. Using methods such as the family of methods developed from FFIP [20, 22, 23, 24, 25], failure propagation can be assessed at a functional level through a system. PRA methods can also be useful to understand system failure

propagation, especially for systems with high redundancy and failure mitigation systems [10]. Many of the above mentioned methods and approaches fall under the umbrella of model-based systems engineering (MBSE) which has been heavily advocated by the International Council on Systems Engineering among others for several decades [26]. The INCOSE Systems Engineering Book of Knowledge (SEBOK) also includes several relevant sections on safety engineering (including several variations of hazard analysis) and reliability, availability, and maintainability which help to improve SOSs in those respective areas [26].

Within the systems engineering V model [27], the method presented in this paper is specifically meant to be used in the system architecture phase of design – near the front end of the V model. In specific, the conceptual phase of system architecture where functional models are being developed from requirements, design reference missions, and other similar information [28] is where the below introduced method is targeted for use. The early conceptual functional modeling stage of system architecture within the systems engineering process is an opportune time to uncover potential unexpected or unanticipated system behaviors, the corresponding initiating events in a Sol, and their impacts on Sols. Large changes to Sol system architecture can be made at this stage without significant adverse impact on schedule and budget [16]. The conceptual phase of system architecture also often precedes hazard analysis, FMECA, PRA, and other similar methods of failure and risk analysis although FFIP and UFFSR are conducted on functional models during conceptual design.

The method presented in this paper is intended for use on SoS and Sol typically used by the US Department of Defense (DoD) such as groups of autonomous vehicles operating in a SoS configuration; adaptive force packages that include surface vessels, underwater assets, airplanes, autonomous vehicles of a variety of types, and other related systems operating as an SoS (often in support of a mission objective and in relation to mission engineering [29]); forward operating base (FOB) complexes where ground vehicles, living quarters, maintenance depots, munitions storage, autonomous vehicles, and other systems are present and can be considered an SoS; and other similar systems. While the method we present below may be useful for other SoSs such as microgrids, cyber SoS (e.g.: fully software-based SoS – note: the previous examples that are within the scope of this method do include cyber-physical elements and are not excluded from consideration), and primarily human-based SoS (e.g.: a company of soldiers and their equipment) among other examples, this is not the primary focus of our presented method. Further, while irrational behaviors of humans can be incorporated to some extent in the method through the FBED flow set, our primary focus is explicitly not on human-system interaction but instead is primarily on the systems themselves.

In spite of the significant advances made in understanding how failures propagate through Sols and SoS, Sol failure events caused by other systems within the SoS are often still missed [30, 31, 32]. As far as we are aware, the analysis how one or more systems within a SoS can behave in unpredicted or unanticipated ways that result in initiating failure events in a Sol are not being well analyzed within existing failure analysis methods during conceptual functional modeling during the system architecture phase of systems design for the specific types of SoS mentioned above. Thus, there currently exists no practicable way for practitioners to identify and analyze potential system unanticipated or unpredicted system behaviors within a SoS that create failure initiating events in a Sol.

1.1 | Specific Contributions

This article contributes an analysis method that helps the practitioner consider irrational system behavior of member systems within a SoS and their impacts on a Sol in the form of "irrationality initiators" (failure initiating events caused by unanticipated or unpredicted system behaviors – described in detail in the methodology section). The method is intended to be used in early system modeling where conceptual functional architectures are developed. An analysis of potential effects (i.e., the *method*) caused by "irrational system behaviors" (system behaviors that are unanticipated

or unpredicted by the Sol systems engineers – described in detail in the methodology section) originating in one or more systems and adversely affecting the Sol through "irrationality initiators" is conducted using several techniques. The result of the analysis can then be used to further develop and refine the Sol system architecture to improve Sol robustness to irrational system behaviors.

2 | BACKGROUND AND RELATED WORK

The method developed in this article relies upon several bodies of work including systems modeling, failure analysis, and probability assessment. This section provides background and discusses related efforts of relevance to the method presented here.

Systems modeling is a family of techniques used to develop models of systems for the purposes of system representation and simulation. Many system modeling techniques are available to the practitioner such as the Integrated Computer Aided Manufacturing (Icam) DEFinition for Function Modeling (IDEF0) language [33] that has seen significant use in the systems engineering community. The Universal Modeling Language (UML) and its offspring, the System Modeling Language (SysML) [34, 35, 36], are seeing increased usage especially within the Department of Defense (DoD). Other modeling languages such as [37, 38] also are available and with varying levels of adoption. This article uses the Functional Basis for Engineering Design (FBED) [18, 39] functional hierarchical modeling language to represent systems. The FBED models system functions and flows where functions defined are the actions that a system can take (e.g., transport energy, convert rotational energy to electrical energy, etc.) and flows are defined as material, energy, or signal moving within the system (e.g., energy-chemical, signal-control-discrete, etc.), into or out of a system boundary, or between systems in the case of a SoS. The FBED function and flow taxonomies are each decomposed into primary, secondary, and tertiary categories where each deeper level has an increased level of specificity. System components are abstracted to a functional level to give engineers the freedom to consider functionality of a system without being locked into a specific component architecture. The abstraction of functions from components and the derivation of component solutions from functions is a well understood and established practice from the original and subsequent development of FBED [40]. FBED is an established National Institute of Standards (NIST) standard that helped to unify several disparate efforts in functional modeling for engineering design [18, 39]. This places FBED as a modeling language primarily suited for conceptual modeling. However, we have observed FBED being used to analyze existing designs as well.

Failure analysis is performed to understand how a system may degrade or fail primarily during operation although the analysis can also be performed for other phases of the system life cycle such as maintenance. Failure Modes and Effects Analysis (FMEA) [41] and its extension, Failure Modes Effects and Criticality Analysis (FMECA) [42], are heavily used in private industry [43] and in the DoD where MIL-STD-882E prescribes FMECA to conduct hazard analysis [44]. FMEA calculates a risk priority number (RPN) by multiplying the probability of a failure event happening, the ability to detect the event before it happens, and the severity of the event on 1-10 scales with the RPN being on a 1-1000 scale to prioritize the order in which potential failure events should be mitigated. However, FMEA and FMECA are ill-suited to identify emergent system behaviors such as multiple component failures that lead to a system-level failure and that have not been observed before in operating systems [45, 46].

PRA combines fault tree analysis (FTA) [47] and event tree analysis (ETA) [48] to produce failure event sequences that generally include multiple components or subsystem failures in sequence to cause a system-level failure. Initiating events are the probability of an event occurring that initiates a potential system failure [49]. However, valid initiating events can be erroneously discounted as being possible or are sufficiently beyond prior experience of engineers conducting the PRA that such initiating events can fail to be included in the analysis [30, 31].

Cut-sets are produced by PRA which can then be used to analyze failure events which require multiple events to occur to lead to failure (usually system failure although failure can be defined differently depending upon the application) [50]. The production of cut-sets is often truncated when the probability of an event occurring falls below a predetermined threshold [51]. This can occasionally lead to low probability but high consequence failure events from being missed in analysis conducted using PRA [52].

Unpredicted or unanticipated system behavior can occur in systems for many different reasons [53]. A significant body of research has been developed to understand unanticipated or unpredicted system behaviors [54, 55, 56] and address such system behaviors through increasing system robustness and resiliency to both external and internal failure initiating events [57, 58, 59]. However, we have found scant evidence of work being done to understand unexpected or unpredicted system behavior within the context of a SoS.

Several efforts have been made to combine functional modeling and failure analysis such as a method of developing FMEAs for functional models [60]. The FFIP method [20, 22] uses a probabilistic approach to analyze functional models for failure propagation. In recent years, FFIP has been extended with methods such as the Uncoupled Failure Flow State Reasoner (UFFSR) [24] that evaluates failure flows that do not travel along nominal flow pathways, and a functional Bayesian approach to developing prognostic and health monitoring subsystems that can detect incipient failures while they still can be corrected [23]. In the FFIP family of methods, initiating events are developed in a similar way to initiating events in PRA which leads to unexpected or unpredicted initiating events often not being considered. The FFIP family of methods produces cut-sets similar to those developed by PRA and handles truncation of analysis in a similar manner [59].

Nan et. al. developed a method of analyzing Supervisory Control and Data Acquisition (SCADA) systems and the systems under the SCADA's control in critical infrastructure to identify vulnerabilities. The method investigates interdependencies of four types including physical, cyber, geographic, and logic. However, Nan et. al. does not conduct the analysis at the functional level of system architecture nor is it an exhaustive flow-based method of identifying potential initiating events [61].

TABLE 1 Comparison of existing risk, reliability, safety, and related methods presented to identify gaps in existing methodologies. Note that instead of referencing multiple hazard analyses, the term "Hazard Analysis" encompasses the intent of the methods enclosed in MIL-STD-882E [44]. Examples methods include Preliminary Hazard Analysis (PHA), Functional Hazard Analysis (FHA), System Hazard Analysis (SHA), Subsystem Hazard Analysis (SSHA), Nan et. al., etc. Note that irrationality initiators are described in detail in the methodology section. In short: an irrationality initiator is an initiating event in a Sol that is induced by another system within the SoS behaving in an unpredicted or unanticipated manner.

Method capability	Proposed method	PRA	FMEA/FMECA	FFIP	Hazard analysis	Nan et. al.
Identifies all irrationality initiators	Y	N	N	N	N	N
Identify failure propagation paths within system	Y	Y	N	Y	N	Y
Quantifying failure probability outcomes	Y	Y	Y	N	Y	Y
Iterate functional model with results	Y	N	N	N	N	N
Propagate uncoupled failure flows	Y	N	N	N	N	N

3 | METHODOLOGY

The method presented in so that a this section has been developed specifically for use during conceptual design of a system that is part of a system of systems when architectural trade-off studies are performed. Significant alterations in a system's design at this stage of the design process are relatively inexpensive to perform and take relatively little time to implement. A high-level flow chart is provided in Figure 1 to graphically show the five steps of the methodology.

FIGURE 1 High-level overview of the proposed methodology.

Note that a demonstration of the method is provided in the Case Study and Results section of this article (Section 4). We have omitted examples within the Methodology section and instead direct the interested reader to the Case Study and Results section. It should also be noted that the case study is a fictionalized case study and is intended only to demonstrate the methodology presented in this section.

3.1 | Model the Systems within the System of Systems

The first step is to model the Sol within the SoS and their connections to one another. FBED [18] is our preferred functional modeling method and is used throughout this article. However, many equally valid methods are available [37, 38, 34, 35, 36]. In concert with developing functional models, physical component solutions to functions can be developed. Having component solutions to functions (either a one to one correlation or a many to one correlation if component solutions have not been down selected yet) at this stage in the design process allows mapping of what happens when a physical component fails to the function it fulfills. An example of a function to component mapping is the function *transfer liquid* which may be fulfilled by the component *pipe* or the component *canal* among other possible component solutions.

3.2 | Identify Potential Irrational System Behaviors

Based on observations in the literature and in our professional practices, we propose considering emergent system behaviors that were not previously predicted or were discounted as being highly unlikely to be what we term "irrational system behaviors." We further refine the definition of irrational system behaviors as unexpected behaviors within a system that emit potential failure initiating events that other systems within a System of Systems (SoS) may in turn receive as inputs and thus cause failures in the other systems. In short: irrational system behaviors are system behaviors that have not been previously observed or predicted (no prior knowledge or discounted as being a potential threat) by other systems within a SoS, and have not been analyzed through routine means of system simulation and hazard/failure analysis.

An example of potential irrational system behavior is a compressed air delivery network SoS with a compressor, an air cleaning system, a distribution system, and multiple compressed air use systems (e.g.: venturi chillers, pneumatic solenoid valves, pneumatic rotary motors, pneumatic cylinders, etc.). While the SoS is designed with the expectation that contaminants such as water and compressor oil may bypass the air cleaning system and would then be caught by filters on the compressed air use systems, the SoS is not designed for and does not expect the compressor to deliver corrosive gas. Such an irrational system behavior may occur, for instance, because of an acetylene tank unexpectedly venting near the compressor's air intake. This may lead to multiple failure events in the systems that receive compressed air in the SoS as unexpected corrosion occurs.

While an argument can be made that our definition of irrational system behavior could also be described as unexpected or unanticipated system behavior, we specifically use the word "irrational" to call practitioners' attentions to this phenomena. In our professional practice, on multiple occasions we have observed senior systems engineers and subject matter experts caught off guard by failure events caused by other systems than the Sol (system of interest – the system being designed to enter an existing or proposed SoS) within a SoS. Within the context of SoS, many practitioners we have discussed the concept of irrational system behavior with have their own stories of other systems within a SoS behaving completely irrationally and impacting their Sols when compared with the practitioners' understanding of how the other systems should behave. We have personally witnessed in several industries that, in spite of (1) excellent requirements, interface management strategies, and comprehensive work products, and (2) outstanding hazard, failure, reliability, and related analyses, irrational system behaviors continue to occur that impact Sols. While it may appear that irrational system behaviors occur with high frequency, it is important to be clear that these are in fact rare events. However, the consequences of irrational system behaviors are significant enough to warrant study and development of the methodology presented in this article.

Even when logical and probabilistic approaches for analyzing a Sol within the SoS are used, the approaches often fail to uncover potential emergent Sol system behaviors that are initiated by irrational system behaviors of other systems within the SoS. The aforementioned issue happens in spite of extensive guidance on searching for potentially overlooked initiating events [30, 62]. In cases where potential failure scenarios caused by irrational system behaviors have been identified, organizations that conduct system failure and risk analysis can sometimes discount such scenarios and not rigorously analyze the potential outcomes [63, 64]. The problem of not identifying or discounting identified emergent system behaviors is compounded as SoS are developed by connecting multiple systems together. As the number of systems in a SoS increases, the likelihood of irrational system behaviors increases. Irrational system behaviors can occur in one or more systems within an SoS [65, 66, 67]. In short, SoS can have irrational system behaviors that may result in severe negative outcomes to individual systems within an SoS or to the entire SoS.

One specific goal of this work is to identify irrational systems behaviors. In order to identify these behaviors, it is useful to understand how systems can behave irrationally. The study of irrational behavior began with investigating

irrational behavior of people such as in the context of economic models [68, 69]. Irrational behavior of people (also often called irrationality) can take different forms and have different causes such as visceral reactions [70] to events, psychosis [71], actions taken under duress [72], or even intentional irrationality [73, 74]. Engineers are no exception to irrational behavior; design engineers can appear to behave irrationally in their risk-based design decisions although such irrationality can sometimes be explained by the individual personal utility function of a specific engineer [75]. While some argue that humans are the only true source of irrational system behaviors, we are using the phrase "irrational system behavior" in a different context, as described above. However, examining the context of irrational behavior of humans is useful in conceptualizing how systems can appear to behave irrationally to an outside observer or even to the subject matter expert of the system behaving irrationally.

Decision theory and utility theory have been used to help understand how people can appear to behave irrationally [76, 77, 78, 79, 80, 81] including how neural systems work [82, 83]. Through the application of utility and decision theory, it is now possible to develop system models that deviate from the expected value theorem and instead match a specific utility function of either an individual or an organization [84]. We contend that (much like humans) while a system may appear to be behaving irrationally to an outside observer, the system's utility function may be different from the observer's expectation. In other words, the system is behaving normally based on its own internal utility function but appears to an external observer to be behaving irrationally.

From our proposed definition of irrational system behavior developed above, we further refine the definition of irrational system behavior to specifically refer to functional flows that exit a system boundary being unreasonable or illogical when compared to expected and previously experienced system behavior. We define unreasonable or illogical behavior as deviation from pre-programmed behaviors and rational expectations [85]; unresponsiveness to incentives [74]; and/or deviation from self-interest, self-preservation, and/or system-of-systems self interest and preservation [86]. We further refine the definition of irrational system behavior in the context of this article to specifically be a failure flow class [20, 22] that exits a system boundary and that would not normally be anticipated through common failure analysis techniques such as hazard analysis [87], failure modes and effects analysis and the related failure modes effects and criticality analysis (FMEA and FMECA) [41], probabilistic risk assessment (PRA) [10], function failure flow identification and propagation (FFIP) [20], uncoupled failure flow state reasoner (UFFSR) [24], and other similar methods. Thus, irrational system behavior produces potential initiating events for the Sol within a SoS. Another way to conceptualize irrational system behavior is that it is similar to Black Swan events as described by Taleb [13, 88] although irrational system behavior is focused specifically on failure initiating events while Black Swan events generally refer to system-level failure.

An initiating event is an event that initiates an incipient failure within a Sol that may propagate through the Sol until 1) the Sol has failed, 2) the Sol is operating in a stable but degraded condition, 3) the Sol recovers to a nominal operating state after a period of degraded performance, or 4) the Sol mitigates the incipient failure. Initiating events are used in PRA, FFIP, and other quantitative failure and risk analysis methods. While standard procedures are available to identify potential initiating events that may affect a system [49], a practitioner may discount initiating events that are outside of prior experiences with a system or that seem irrational [30, 31, 32].

We propose supplementing existing methods of identifying initiating events (e.g.: [49]) by introducing the concept of irrationality initiators which we define as irrational system behavior within a SoS that creates initiating events within a Sol. In other words, irrationality initiators are caused by irrational system behavior of one or more systems within a SoS that emit unexpected system boundary-crossing failure flows. The failure flows become irrationality initiators when they encounter the Sol in the SoS. Irrationality initiators may follow nominal flow paths between systems such as a data link between two systems or irrationality initiators may affect the Sol by propagating via uncoupled flow paths [24]. We distinguish irrationality initiators from failure flows that turn into ordinary initiating events to specifically

denote that irrationality initiators are initiating events originating outside of the Sol and caused by irrational system behavior of other systems within the SoS. Irrationality initiators are not caused by the environment and do not include expected and/or understood failure flows from other systems within the SoS that are already captured through existing methods of initiating event analysis. As is the case with ordinary initiating events, an irrationality initiator may also cause a failure to propagate through a Sol and may result in one of several system end states including partially failed or degraded performance of the Sol; failure of the Sol; after an initial period of disruption, the Sol recovers to a nominal state; or a nominal Sol state. To reiterate, in the context of a SoS, a Sol acquires irrationality initiators from **other** systems within the SoS. It should be noted that a Sol receiving irrational initiators may in turn generate its own irrational system behaviors which may turn into irrationality initiators in other systems within the SoS.

Based on the proposed definitions of irrational system behavior and irrationality initiators developed above, we propose the following approach, as shown in Figure 2, to identify irrationality initiators. The approach starts with all potential flows from the FBED flow set [18, 39] before reducing down to potential flows that may happen within a specific SoS.

- *Step 2, Part 1:* Start with **all** secondary and tertiary flow descriptions from FBED. Each flow may conceivably be an irrationality initiator coming from a generic black box system within a SoS. From a conceptual standpoint, it is irrelevant if a failure flow is being emitted by a function or a linked component within the models – in this step, the failure flows are considered to be emitted from a black box system model. Note that the use of the abstracted FBED flows is intentional; abstracting away from physical components and subsystems to the functional level can help practitioners to consider potential new initiating event sources that otherwise may be missed.
- *Step 2, Part 2:* Remove all flows from the list of potential irrationality initiators that are already modeled as initiating events through other failure analysis methods such as FFIP or PRA.
- *Step 2, Part 3:* Identify any potentially impossible candidate irrationality initiators that cannot be emitted by the generic black box system. Before eliminating a candidate irrationality initiator, the practitioner must attempt to identify ways that the irrationality initiator may be able to be generated even if it is highly implausible or unlikely. For instance, almost any material can produce spectral emissions that would normally be unexpected with sufficient energy applied to the material.
- *Step 2, Part 4:* Assign probabilities of occurrence to each of the irrationality initiators remaining on the list. We advocate that practitioners follow initiating event probability guidance from PRA such as [10] and [49].

Now that potential irrationality initiators within a SoS that may impact the Sol have been identified and probabilities assigned, the flow paths by which the irrationality initiators enter the system must be defined. Irrationality initiators may be introduced to a system along nominal flow paths or along non-nominal flow paths such as the uncoupled failure flow paths advanced in the UFFSR method [24]. Additions to or modifications of the failure model for a system may be necessary to sufficiently capture irrationality initiator entry points.

3.3 | Analysis of Potential Irrationality Initiators

The next step in the method is to conduct failure analysis on the Sol using the identified potential irrationality initiators. We advocate for and use in the case study the FFIP family of failure analysis tools to conduct failure analysis on the Sol. In order to produce a more accurate analysis of potential irrationality initiators using FFIP and related tools, we recommend that the analysis be performed using data collected from the proposed physical architecture that solves the functional architecture of the Sol.

FIGURE 2 Steps to Developing Irrationality Initiators.

The number of potential failure scenarios, often called "cut-sets" in PRA and sometimes in FFIP, resulting from the analysis of irrationality initiators is directly related to the number of irrationality initiators and the functional model of the Sol. Each irrationality initiator may proceed along many different flow paths in a Sol, causing functional failure along the way, which in turn may lead to system failure. The number of potential failure scenarios may further be expanded by having multiple potential component solutions available for specific functions before down-selection of component solutions has been conducted.

While probabilities for specific irrationality initiators were calculated in a prior step in the method, there are several options for how irrationality initiators are analyzed based on what type of analysis results a practitioner is interested in reviewing. These include an uninformative prior and an informative prior. Further, irrationality initiators that are either independent or dependent can be considered to provide additional insights into potential irrational failure scenarios such as when multiple irrationality initiators often occur together. Informative and uninformative priors, and independent and dependent irrationality initiators may be combined together. Further explanation immediately follows:

3.3.1 | Uninformative and Informative Priors

In order to understand the sensitivity of a Sol to irrationality initiators, the uninformative prior sets all irrationality initiators to the same probability of occurrence. It should be noted that using the uninformative prior approach does not allow for direct comparison of results with other FFIP results. The results are specifically useful to understand what high severity failure outcomes are present that otherwise may be truncated during computation. The uninformative prior method can also be used to perform a sensitivity analysis on the irrationality initiators by changing their probabilities

and comparing results. This may help to identify irrationality initiators that are not particularly sensitive to changes in their probabilities of occurrence and may also identify specific irrationality initiators that warrant extended scrutiny to ensure a higher degree of accuracy and realism in the probability statistics.

In contrast to the uninformative prior that uses arbitrarily assigned probabilities to determine potential low probability but very severe outcomes and to examine irrationality initiator probability sensitivity, the informative prior uses probabilities of occurrence that were already developed in a previous step of the methodology and that are based in reality. This allows for direct comparison of irrationality initiator-derived failure scenario probabilities with failure scenario probabilities produced from FFIP.

In the event that a probability was unable to be developed previously because of a lack of information, we suggest using a probability value that is 3x the highest probability of the highest known irrationality initiator probability. Using a 3x higher probability may help to ensure that any potential high consequence failure scenarios are identified and will help to motivate the development of a better estimation of the probability. If a failure scenario of a particular irrationality initiator that used the 3x higher probability is sufficiently probable, then this indicates the irrationality initiator probability needs to be better understood. However, if no failure scenarios are within a few orders of magnitude of the most likely failure scenario, then this is an indication that there is likely no further refinement of that irrationality initiator's probability. It is worth noting that we do not advocate for setting the multiplier higher than 3x for irrationality initiators without well-founded probabilities. While such an approach would almost certainly highlight every single potential failure scenario caused by the irrationality initiator in question, setting the irrationality initiator probability needlessly high without a rigorous analysis to back up the choice is likely to overwhelm a user of this method with many failure scenarios that masquerade as being of high likelihood while in reality being of vanishingly small probability. This in turn may lead to much wasted time and effort to disprove all of the failure scenarios. The suggestion of a 3x multiplier is based on our prior professional experience as risk analysts and reliability engineers and from examining the sensitivity of several failure models to which we have access to changing initiating event probabilities. While we believe the 3x multiplier is a good starting point, we recommend that systems engineering practitioners carefully examine the sensitivity of their own systems to initiating event probabilities and make adjustments as warranted and using their professional engineering judgment.

We recommend that both the informative and uninformative prior methods are used to analyze irrationality initiators in the Sol. The uninformative prior can shed light on potential high consequence failure scenarios that otherwise may be missed and can also be used to perform sensitivity studies on the irrationality initiators. The informative prior quantifies failure scenarios in a way that can be directly compared with standard FFIP results. This may help practitioners to prioritize where money and time is spent to mitigate potential issues.

3.3.2 | Independent and Dependent Irrationality Initiators

In almost every implementation of FFIP that we have encountered, initiating events are exclusively considered to be independent from each other. The same is true in many PRA analyses. However, we suspect based on our professional practice and observations that irrationality initiators may have a higher likelihood of being dependent upon one another to some extent. In other words, if one irrationality initiator occurs, then it is more likely that another will occur at the same time. We propose that irrationality initiators should be modeled both as independent and dependent events. By analyzing multiple irrationality initiators as single events, a practitioner can gain insight into scenarios where a system in a SoS begins emitting many irrationality initiators. This may help to identify "worst case scenarios" where completely unanticipated emergent system behaviors occur due to the Sol receiving several irrationality initiators at once. Recent research on external initiating events for autonomous robotic systems has indicated that unique emergent system

TABLE 2 Comparison of methods of analysis for dependent and independent and informative and uninformative approaches to irrationality initiators.

Method	Direct Comparison with FFIP Possible	Identify High Consequence but Potentially Low Probability Outcomes	Analyze Several Irrationality Initiators at Once
Independent Uninformative Prior Method	No	Yes	No
Dependent Uninformative Prior Method	No	Yes	Yes
Independent Informative Prior Method	Yes	No	No
Dependent Informative Prior Method	No	No	Yes

behaviors not predicted by other research methods can be caused several external initiating events simultaneously occurring and interacting with one another inside of a Sol [59, 89].

We suggest that all possible irrationality initiator dependent combinations be investigated. For example, in the case of three irrationality initiators [A, B, C], the following initiator dependent combinations should be investigated: [A & B], [A & C], [B & C], and [A, B, & C]. A generalized formula to determine the number of dependent combinations is shown by Equation 1. Note that the formula intentionally subtracts 1 to acknowledge that the baseline case of no irrationality initiators being present in the Sol is assumed to have been previously assessed.

$$2^n - n - 1$$

(1)

We recommend going through the the informative and uninformative prior analysis steps as described previously with the dependent irrationality initiators. In the case of the informative prior, we recommend conducting a thorough probability analysis of each dependent combination. However, we recognize that this may be very difficult to complete with any level of accuracy. In cases where analysis cannot or is not completed for dependent combinations, we suggest using the highest single probability of any of the irrationality initiators in the dependent combination. In effect, this approach uses an OR logic probability calculation which we believe is a conservative approach in this case. In many practical implementations of various types of risk analysis (e.g.: PRA, FFIP, etc.), the initiating event probabilities are often assumed to be independent from each other, and in scenarios where two initiating events occur simultaneously is generally considered extremely unlikely. However, observations of improbable events occurring with startling regularity suggests that perhaps the assumption that initiating events are almost always independent is not entirely valid [88]. Thus, without having a better understanding of the true likelihood of a specific dependent combination occurring, the highest probability of an irrationality initiator within the dependent combination is an appropriate and conservative approach. Table 2 shows an overview of the informative and uninformative, and the independent and dependent priors and the benefits and limitations of each.

3.3.3 | Specific Guidance on Modeling Implementation with FFIP

We envision the method presented in this article to be used in concert with a quantitative failure analysis technique such as FFIP or PRA. While conducting a failure analysis with either technique is well understood and documented in the literature, there are a few differences and caveats to be aware of when using irrationality initiators. As mentioned above, uninformative priors cannot be directly compared to FFIP failure scenarios. However, informative priors can. Dependent combinations of irrationality initiators can be compared with FFIP results as long as they are not using uninformative priors.

Within the family of tools developed around the FFIP methodology, each individual function's response to all potential failure flows is modeled. Results of a function receiving a failure flow can include: 1) reduction, increase, or stoppage of nominal flows leaving the function; 2) failure flows passing through the function and continuing on along nominal or non-nominal flow paths; 3) failure flows being arrested or rejected by the function and the function continuing to operate nominally; 4) new failure flows being output by the function; or 5) some combination of the above. A probability is developed for each potential outcome which is then used to develop and calculate the probability of specific failure sequences.

Irrationality initiators may enter a system through two flow path types that cross the system boundary: nominal flow paths and non-nominal flow paths. In the case of nominal flow paths, the core FFIP method can be used to model how an irrationality initiator initiates a failure that moves through a system. In the case of a non-nominal flow path, the UFFSR methodology [24] that extends FFIP is useful. UFFSR can model uncoupled failure flow paths where a failure flow "jumps" into or out of a system, or between functions in a system where no nominal flow path exists. Because of this ability, UFFSR is particularly useful when modeling irrationality initiators where they may enter a Sol through non-nominal flow paths. Recent events such as the Deepwater Horizon show that failure flows do not always travel along the nominal flow path and may jump between unconnected systems [90].

In summary, several analyses may be performed on irrationality initiators depending upon the needs of the practitioner. We suggest using all of the above approaches but acknowledge there will be specific instances where it may be appropriate to only use some of the approaches. Implementing the analysis can be done in FFIP and with the UFFSR extension to FFIP.

3.4 | Analyze Results of Irrationality Initiator Failure Scenarios

After developing failure scenarios specific to irrationality initiators in the previous section, the results can now be analyzed to understand the potential impacts of irrational system behavior on a Sol in a SoS. The independent uninformative prior results can be used to identify high consequence failure scenarios and to identify irrationality initiators that are sensitive to changes in their probability values. These results can be used to identify potential areas to invest more effort in further developing knowledge of a specific irrationality initiator. Potential high consequence failure scenarios identified through the independent uninformative prior may also point to areas in the model where further development and scrutiny is warranted. The dependent uninformative prior failure scenario results provide similar information as described in the above paragraph but with focus on dependent combinations of irrationality initiators.

Failure scenario results from the independent informative prior can be compared directly with FFIP results (if using the FFIP methodology as the underlying failure analysis method). A comparison of the results of the independent informative prior with FFIP results can reveal if irrationality initiators are a significant or dominant contributor to probability of system failure. Failure scenario results of the independent informative prior can also be added to FFIP results to produce a combined analysis which gives a more holistic view of potential Sol system failure scenarios. It

is important to maintain a list of irrationality initiators that were assigned a generic probability of occurrence due to no realistic probability being available – failure scenarios associated with these specific irrationality initiators may be disproportionately represented high in the results.

For both the informative and uninformative priors, the dependent irrationality initiator combinations may provide insight into "worst case scenarios" where many irrationality initiators are emitted from one or more systems within a SoS and impact the Sol at the same time. An additional concern that may be uncovered from analyzing dependent irrationality initiator combinations is a situation where one or more irrationality initiators have passed into the Sol boundary but the Sol continues to function normally. A subsequent irrationality initiator that otherwise may have not caused the Sol to suffer a system failure could now cause the weakened Sol to fail. A practical example of this effect is a failed emergency brake in a car where the braking functionality is not available in an emergency stopping situation if the primary brakes have failed. The car can still brake under nominal operating conditions but the car will be unable to brake (excluding engine braking which may or may not be available depending on specifics of the car configuration) if the primary braking system also suffers a failure.

3.5 | System of Interest Design Iteration

The insights that the analyses provide can then be used by practitioners to help guide improvements to a Sol to increase its robustness to irrationality initiators. Improving robustness of the Sol within a SoS can help to improve the reliability of the overall SoS and the likelihood that the SoS will complete its mission. While a SoS will never be without risk of failure due to a member system behaving irrationally, failure risk can be sufficiently reduced to be manageable and acceptable through careful analysis and improvement of the constituent systems (e.g.: the Sol).

Following a thorough analysis of the results of the irrationality initiator analysis method presented in this article, changes to the Sol can be made to help prevent the irrational system behaviors of one system from adversely impacting the Sol and the SoS. There are many options to protect a Sol from irrational system behaviors of another system within the SoS. For instance, protection can be implemented against uncoupled irrationality initiators [91]. Redundant systems and subsystems [92, 93] can be added to provide higher reliability. Sacrificial subsystems or systems [59] can be added to route failure flows caused by irrationality initiators to a location where they can do the least harm. Robustness and resilience of the Sol [94, 95, 16] can be improved to better deal with inputs to the Sol that go beyond the design basis of the Sol. Changing Sol configuration or location can be used to decrease the likelihood of irrationality initiators from occurring [96].

After sufficient redesign of the Sol has been completed, the method can be iterated upon to verify that irrationality initiator risk of failure to the Sol and the SoS has decreased to an acceptable level. If the risk of Sol failure or SoS failure has not adequately decreased, further design iterations are necessary. Once a system engineer or designer is satisfied that the Sol architecture is sufficiently robust against irrationality initiators caused by another member of the SoS behaving irrationally, the Sol architecture can be locked and the system engineering process can continue to move forward in the systems engineering process.

4 | CASE STUDY AND RESULTS

In this section, we present a simplified and fictionalized case study to demonstrate the method. The case study SoS and Sol are representative of real systems in operation and/or being designed (such as [97] and others), although certain details and contexts have been changed to more clearly demonstrate the method and protect sensitive system

information. However, the system models and other details remain representative of several existing fielded SoS and Sol. The case study is only to be used for illustrative purposes and to demonstrate the method presented in this article; no engineering conclusions on existing or proposed SoS and Sol can be drawn from the case study without a practitioner first conducting their own thorough analysis.

In the case study, a SoS that delivers sensitive supplies from a logistics supply depot to a forward position has been operating for some time. The SoS uses a small fleet of autonomous vehicles to move supplies between the depot and the forward position through a partially denied environment [98] where there is poor and intermittent Global Positioning System (GPS) coverage, and other navigational aids such as way-point navigation beacons and celestial navigation are unavailable. The autonomous vehicles have limited ability to track their own positions internally and must receive regular position updates to prevent excessive drift. To overcome a lack of positioning data, the autonomous vehicles receive positioning information via a series of radar stations that are able to localize the autonomous vehicles. Two-way communications for command and control is also provided via the radar stations and links back to the logistics supply depot where a control station is located. The radar stations are configured in a nodal network. The location of the radar stations are not optimal due to the topography of the area and hostile forces active in the area. Radar and communications coverage does overlap in some areas and is desirable but much of the route the autonomous vehicles take only has single radar station coverage. The supplies are sufficiently sensitive that if positioning information is lost for more than 3 minutes or if a sufficiently large deviation from the expected path is detected, it is assumed that the autonomous vehicle may have been captured by hostile forces and both the vehicle and the supplies aboard are destroyed [99, 100]. As long as the autonomous vehicles remain on their intended path, there is no threat of capture by hostile forces.

The existing autonomous vehicles in the SoS are reaching the end of their service life and a defense contractor is developing a new fleet of autonomous vehicles (the Sol for the case study) to begin service. The radar stations and the control station are manufactured by other contractors. The SoS integration is handled by another contractor, as is often the case in defense SoS. While the defense contractor has an understanding of how the other systems in the SoS are supposed to operate and behave, the defense contractor desires to have a better picture of potential threats to the Sol from irrational system behaviors of the other systems. The defense contractor plans to use the knowledge gained from investigating irrationality initiators to improve robustness and reliability of the Sol during the conceptual phase of system architecture where functional models are being developed which will increase the likelihood of SoS mission success. Figure 3 shows the general SoS configuration.

4.1 | Model the Systems within the System of Systems

The defense contractor chose to use a functional modeling approach and FFIP as the underlying failure analysis tool for the irrationality initiator analysis. A model of the SoS is shown in Figure 4. Note that nominal flow paths are shown in the figure. A FFIP analysis of failure scenarios for the SoS and the system of interest have already been performed. Further, system solutions to functions within the SoS model have been chosen.

A simplified Sol (the new autonomous vehicles) system model developed with the FBED functional taxonomy is shown in Figure 5. A FFIP analysis was conducted and the five most likely failure scenarios are shown in Table 3. The FFIP family of methods examines how failures move through a system from an initiating event through to either failure of the system (often defined by failure of a critical function or functions and/or by a failure flow exiting the system boundary) or to termination of the failure flow without the failure flow causing system failure [20, 22, 24]. The probability of the failure outcome is calculated from the probability that the failure flow 1) initiates, 2) passes through each function, and 3) causes the system to fail and/or emit a failure flow. Each failure scenario that FFIP identifies is treated as an

FIGURE 3 SoS general physical configuration. Note that this figure is representative of a Department of Defense Architecture Framework (DoDAF) 2.0 High-Level Operational Concept Graphic (OV-1) [101].

FIGURE 4 SoS top level model with major systems and flows between systems identified. Note that Data is used in place of the Signal-Control-Discrete flow type. RADAR is used in place of the Energy-Electromagnetic flow type. Both substitutions have been made for ease of understanding for readers who are not intimately familiar with FBED. Two autonomous vehicle systems (the SolS) are shown in typical flow connection configurations where one or more radar stations are connected to the autonomous vehicles.

independent sequence of events from every other failure scenario similar to how many implementations of PRA treat cut-sets as independent from one another for the purposes of the associated probability statistics. Additionally, much like PRA, FFIP scenario outcome probabilities can be added together to understand the overall probability of system failure. The defense contractor has defined failure of the Sol (the autonomous vehicles) as the cargo not reaching its

FIGURE 5 Simplified Sol (the new autonomous vehicle systems) functional model. Many functions and flows have been excluded from or simplified in this functional model for brevity and ease of understanding the case study. The dashed border indicates the system boundary.

TABLE 3 Truncated list of highest probability of failure FFIP results from the Sol on a per unit basis.

Failure Propagation Pathway	Probability
Signal-Control-Discrete, Channel-Transmit, Signal-Process, Convert Electrical Energy to Mechanical Energy, Channel-Guide-Rotate	1.2E-3/day
Provision-Supply, Signal-Process, Signal-Control-Discrete, Channel-Export, Provision-Store-Contain	2.7E-3/day
Signal-Control-Discrete, Channel-Transmit, Signal-Process, Channel-Export, Provision-Store-Contain	3.7E-4/day
Energy-Mechanical, Channel-Guide-Rotate, Convert Electrical to Mechanical Energy, Provision-Supply, Signal-Process, Channel-Export, Provision-Store-Contain	1.4E-5/day
Signal-Process, Channel-Transmit, Provision-Supply, Convert Electrical to Mechanical Energy, Channel-Guide-Rotate	5.4E-5/day

final destination which may occur from the cargo being damaged, captured, destroyed, or lost.

4.2 | Identify Potential Irrational System Behaviors

The next step is to identify potential irrational system behaviors that other systems might undertake in the SoS. For the purposes of this case study, we are narrowing our focus specifically to the radar stations. In a full analysis using the method presented in this article, each member system in the SoS would be analyzed and all results would be used throughout the analysis.

First, the full list of secondary and tertiary flows from FBED is examined, as seen in Table 4. Next, the flows already represented in the FFIP analysis conducted previously (see Table 3) are struck from consideration (shown in Table 4 by a horizontal strike-through line). The third step is to validate that each remaining flow is somehow possible to occur and remove from consideration any flows that are absolutely impossible. This is represented in Table 4 by flows being crossed out. Validation of the flows can be conducted in a variety of different ways that the practitioner finds suitable to the task and with a variety of different levels of fidelity. For instance, a workbook could be developed for

each flow similar to what is done for individual initiating events in a nuclear PRA model [102]. The flows that remain are the irrationality initiators for the system. The final step is to develop probabilities of occurrence for each irrationality initiator. Again, there are a variety of ways these probabilities could be developed depending on the specific situation. For instance, guidance is provided in the nuclear power industry for the development of new initiating events [102], resources are available in MIL-STD-882E to estimate reliability of components and systems [44], and other methods are also available [32].

As an example of identifying irrationality initiators, we will now examine the Signal-Status-Auditory (i.e., noise) flow to determine if the flow should be carried forward as an irrationality initiator for further analysis. First, the flow is checked to verify that it was not already captured in the FFIP analysis (Table 3). Then, the flow is analyzed to determine its potential to reach the Sol (the new autonomous vehicles that are under development). However, based on the physical layout of the system, the Sol will never come close enough to the radar station for a noise generated by the radar station to impact the Sol. The distance is too great for the loudest sound possible to be generated by the radar station (e.g.: the radar station's fuel source being detonated) to reach the Sol with sufficient intensity to become an irrationality initiator. Thus, the Signal-Status-Auditory flow can be struck from the table of candidate irrationality initiators (Table 4). Had the flow not been struck from the list of candidate irrationality initiators, the next step would have been to quantify the likelihood of occurrence.

It should be noted that understanding if an irrational behavior-induced failure flow under consideration for an Sol irrationality initiator is not initially well understood, significant databases of information exist in a variety of industries which may aid systems engineers to better understand the situation. For instance, the nuclear power industry maintains significant databases of component part failures spanning many decades [103]. The petroleum industry maintains similar databases [104]. Mishap reports from similar systems to the Sol may also be useful [11].

After careful analysis, three flows remain as viable irrationality initiators including: Signal-Control-Analog, Material-Solid-Object, and Energy-Electromagnetic-Solar, as shown in Table 4. To illustrate how these three flows were determined to be irrationality initiators, We will briefly focus on the Material-Solid-Object flow. The Material-Solid-Object flow represents part or all of a radar station rolling off the side of a mountain where it was placed and physically impacting the Sol. While this may seem far-fetched, we have observed similar events in our own professional practices. Several causes of this irrationality initiator were identified such as a small landslide causing the radar station to fall down the mountain, hostile forces rolling the radar off the side of the mountain, abnormally high winds ripping one of the communications dishes off of the radar station and blowing it down the mountain, and other equally outlandish causes that nevertheless cannot be completely ruled out. Next, analysis was conducted to determine the probability of the Material-Solid-Object irrationality initiator occurring. It was found to have a relatively high probability of occurrence based on frequent windstorms observed in the area and topographic features in the area which may tend to funnel debris down the mountains and into the path of the Sol.

A simplified FFIP failure analysis model is shown in Figure 6 of the Sol (the new and under development autonomous vehicle system). The orange dashed lines indicate one failure flow sequence moving through the system at the functional level. The irrationality initiator initially crosses the system boundary the system along a non-nominal flow path before traveling along nominal flow paths to eventually cause system failure. Additional development work not shown here was completed on the system models to allow for analysis of the irrationality initiator failure scenarios to be conducted.

4.3 | Analyses of Potential Irrationality Initiators

After developing the Sol system models and identifying the irrationality initiators, analysis can be conducted on the Sol using the four approaches outlined in the methodology section above (independent and dependent irrationality

TABLE 4 Irrationality initiators are developed from the FBED functional taxonomy flow set. The three primary classes are material, signal, and energy. The secondary and tertiary classes have increasing levels of specificity. Note that not all flows are represented at the tertiary level in FBED and some flows may have several representations at the tertiary level. The flows that have been identified as irrationality initiators in the case study are not struck out. Probability of occurrence has been developed for the irrationality initiators as explained in the text above. Refer to [18, 39] for additional details regarding the FBED flow set.

Primary	Secondary	Tertiary	Probability
Material	Human		
	Gas		
	Liquid		
	Solid	Object	1E-2/day
		Particulate	
		Composite	
	Plasma		
	Mixture	Gas-gas	
		Liquid-liquid	
		Solid-solid	
		Solid-Liquid	
		Liquid-Gas	
		Solid-Gas	
		Solid-Liquid-Gas	
Signal		Colloidal	
	Status	Auditory	
		Olfactory	
		Tactile	
		Taste	
		Visual	
	Control	Analog	1.7E-3/day
		Discrete	
Energy	Human		
	Acoustic		
	Biological		
	Chemical		
	Electrical		
	Electromagnetic	Optical	
		Solar	8E-8/day
	Hydraulic		
	Magnetic		
	Mechanical	Rotational	
		Translational	
	Pneumatic		
	Radioactive/Nuclear		
	Thermal

FIGURE 6 Sol (the new autonomous vehicle system under development) high level functional model with a potential irrationality initiator failure scenario indicated by the dashed orange line. The Sol fails when the failure flow caused by the irrationality initiator reaches the Provision-Store-Contain (Cargo) function via the Channel-Export function which results in the cargo being destroyed.

TABLE 5 Irrationality initiator dependent combinations for the Sol (new autonomous vehicle system) on a per unit basis.

Dependent Grouping	Probability
Signal-Control-Analog AND Energy-Electromagnetic-Solar AND Material-Solid-Object	1E-2/day
Signal-Control-Analog AND Energy-Electromagnetic-Solar	1.7E-3/day
Signal-Control-Analog AND Material-Solid-Object	1E-2/day
Energy-Electromagnetic-Solar AND Material-Solid-Object	1.2E-2/day

initiators, uninformative and informative priors uninformative prior). The three irrationality initiators and their potential dependent combinations are shown in Table 5. A subset of the results of analysis conducted using the FFIP family of tools with the irrationality initiators is shown in Table 6.

4.4 | Analyze Results of Irrationality Initiator Failure Scenarios

Failure scenarios produced from the independent uninformative prior approach's sensitivity analysis show a high sensitivity to change in probability values for the Energy-Electromagnetic-Solar irrationality initiator. This is a strong indication that additional resources should be dedicated to evaluating the Energy-Electromagnetic-Solar irrationality initiator to ensure the probability of occurrence is realistic and conservative.

The dependent uninformative prior approach indicates that the three irrationality initiators occurring at the same time results in a much higher probability of system failure than other combinations of irrationality initiators produce. This indicates that a Sol system redesign may be needed to specifically protect against this scenario.

The Signal-Control-Analog irrationality initiator was identified as a significant contributor to Sol system failure probability based on the independent informative prior approach. This information can be used by a systems engineer to to make a decision on dedicating more resources toward mitigating potential Sol system failures caused by this

TABLE 6 A subset of failure scenarios caused by the irrationality initiators as developed through analysis using FFIP and related methods. The failure scenarios show below are the highest probability for each of the four methods (Independent Uninformative Prior, Dependent Uninformative Prior, Independent Informative Prior, Dependent Informative Prior).

Failure Propagation Pathway	Probability
Independent Uninformative Prior Method	
Energy-Electromagnetic-Solar, Provision-Supply	4.3E-3/day
Energy-Electromagnetic-Solar, Provision-Store-Contain	2.6E-3/day
Material-Solid-Object, Channel-Guide-Rotate	1.2E-3/day
Material-Solid-Object, Channel-Export, Provision-Store-Contain	5.2E-4/day
Signal-Control-Analog, Channel-Transmit, Signal-Process, Channel-Export, Provision-Store-Contain	1.3E-4/day
Dependent Uninformative Prior Method	
Signal-Control-Analog AND Energy-Electromagnetic-Solar AND Material-Solid-Object, Channel-Export, Provision-Store-Contain	4.2E-2/day
Signal-Control-Analog AND Energy-Electromagnetic-Solar, Channel-Guide-Rotate	3.6E-3/day
Signal-Control-Analog AND Energy-Electromagnetic-Solar, Channel-Export, Provision-Store-Contain	8.7E-4/day
Signal-Control-Analog AND Energy-Electromagnetic-Solar, Provision-Store-Contain	5.9E-5/day
Signal-Control-Analog AND Energy-Electromagnetic-Solar, Signal-Process, Convert Electrical to Mechanical Energy, Channel-Guide-Rotate	3.3E-5/day
Independent Informative Prior Method	
Signal-Control-Analog, Channel-Transmit, Signal-Process, Channel-Export, Provision-Store-Contain	4.2E-4/day
Signal-Control-Analog, Channel-Transmit, Signal-Process, Provision-Supply, Convert Electrical to mechanical Energy, Channel-Guide-Rotate	6.3E-4/day
Material-Solid-Object, Channel-Export, Provision-Store-Contain	7.2E-5/day
Material-Solid-Object, Channel-Guide-Rotate	8.4E-5/day
Signal-Control-Analog, Channel-Transmit, Provision-Supply, Convert Electrical to Mechanical Energy, Channel-Guide-Rotate	2.9E-6/day
Dependent Informative Prior Method	
Signal-Control-Analog AND Energy-Electromagnetic-Solar AND Material-Solid-Object, Channel-Export, Provision-Store-Contain	5.2E-4/day
Signal-Control-Analog AND Energy-Electromagnetic-Solar, Channel-Guide-Rotate	8.3E-5/day
Signal-Control-Analog AND Energy-Electromagnetic-Solar, Channel-Export, Provision-Store-Contain	7.2E-5/day
Signal-Control-Analog AND Energy-Electromagnetic-Solar, Provision-Store-Contain	5.1E-5/day
Signal-Control-Analog AND Energy-Electromagnetic-Solar, Signal-Process, Convert Electrical to Mechanical Energy, Channel-Guide-Rotate	3.3E-5/day

particular irrationality initiator.

As with the earlier dependent uninformative prior results, the dependent informative prior approach points toward the combination of all three irrationality initiators has the potential for significant Sol system failure events. However, because a more realistic probability of occurrence is being used as part of the calculations, the probability of the Sol system failure scenarios is lower than results from the original FFIP analysis. In spite of this, the failure scenario outcomes of the irrationality initiator combinations are significant enough that a systems engineer may still wish to verify the probabilities before deciding to discount the result.

Once this step has been completed, the systems engineer must evaluate the Sol and SoS requirements to determine if specific risk, reliability, and other relevant requirements have been satisfied in light of the above analysis. If the requirements have been satisfied, then the systems engineer can exist the method. However, if the requirements have not been satisfied as evidenced by the above analysis, then the systems engineer should proceed to the final step in the methodology.

4.5 | System of Interest Design Iteration

Now that the results of the method have been analyzed, a redesign of the Sol can be conducted to improve system reliability in the face of irrational system behaviors of other systems that the Sol interacts with within the SoS. Many resources exist for practitioners to conduct redesign efforts [16]. After a redesign has been completed, a new analysis should be conducted to verify the success of the redesign effort. We omit further redesign efforts from the case study here in the interest of brevity.

5 | DISCUSSION AND FUTURE WORK

The method we present in this article raises several topics worthy of discussion. This section reviews the benefits of the method as well as limitations and drawbacks. Several questions of philosophical importance to users of the method are also covered. Based on these discussion points, we present some potential future research directions to further improve upon the method.

As has been mentioned above, the four analyses (independent informative prior, dependent informative prior, independent uninformative prior, dependent uninformative prior) each yield unique insights that may be useful to a practitioner (see Table 2 for a summary). We advocate that each of the four analysis methods be used during an irrational system behavior analysis but we also recognize that such analysis may be too computationally expensive for very complex systems or may be too time intensive for very large SoS. The uninformative prior approaches help the practitioner to understand sensitivity to changes in probability of occurrence of irrationality initiators and to discover potentially significant failure scenarios which could otherwise be truncated as being of too low of probability. The informative prior approaches produce failure scenarios that can be compared directly to failure scenarios produced by FFIP. Further, the failure scenarios can be added to the list of failure scenarios that FFIP produces to give a more complete view of how a Sol may fail. Treating irrationality initiators as independent events versus dependent events allows a practitioner to investigate both situations where only one irrationality initiator occurs and cases where multiple occur simultaneously. Looking at combinations of irrationality initiators may help to find emergent system behaviors that otherwise would have been missed because any one irrationality initiator on its own might not have caused the failure to occur. Future work may include conducting a detailed comparison of existing risk, reliability, failure, safety, and other related analyses with the method presented in this paper from the specific perspective of dependent priors.

This may help to reveal gaps in the understanding of how emergent system behaviors occur due to several irrationality initiators being dependent upon one another in ways that have not previously been fully characterized.

The process of identifying and validating irrationality initiators as being possible, and of developing realistic probabilities can be very challenging. However, it is not too dissimilar to the process that is done for new nuclear power plant risk analysis to develop initiating events [102]. Where the method presented in this paper to develop irrationality initiators differs from other established methods of developing initiating events is that irrationality initiators by their very nature are rare events that either have not been seen before or have been discounted as being likely enough to occur to include in analysis. In this case, it can be a vexing problem to develop realistic probabilities that are validated with any sort of quantitative data. However, we believe that even with these limitations, the method is useful enough for practitioners to adopt. The insights gained could help to improve Sol and SoS reliability and robustness by improving constituent systems' (e.g.: the Sol) responses to irrationality initiators.

No explicit guidance has been provided in this method on how to deal with humans in an Sol or the SoS, other than the fact that they can explicitly be modeled into the functional and physical models in FFIP. While this might be sufficient, we recognize that in many situations humans are the most likely point of failure. Humans tend to behave in a manner that was not anticipated or expected [105, 64, 106, 107, 56]. The flows from FBED do contain human flows (e.g., *human energy*) that can be used to begin to develop irrationality initiators that are human-caused. However, acts of commission [108, 109], acts of malice [110, 31], and acts of irrationality, insanity, or calculated instability [111, 112] are not well represented in existing human reliability analysis methods. Further work is needed to more accurately assess potential irrationality initiators caused by humans and is beyond the scope of this article.

A potential computational benefit of the irrationality initiator approach is breaking potential loop-backs in the analysis of failure events. Loop-backs are a significant challenge in SoS that have a high number interconnections, and in systems with a large number of connections between functions [113]. Transforming irrational failure flows exiting a system into irrationality initiators entering a Sol helps to isolate the flows as a source of loop-backs in the analysis.

The FFIP family of methods is similar to PRA in that it can be very computationally expensive (taking many computational resources for long periods of time) to analyze large, complex Sols. Truncation, as discussed in the background and related work section, is heavily used in this situation to reduce computational expense and time requirements. The computational requirements for the method presented in this paper are on par with FFIP and PRA methods in computational expense. Further fundamental research in computational efficiency of probabilistic-based analysis methods is needed to reduce computational expense of the method presented in this paper and many other methods such as FFIP and PRA. It is beyond the scope of this research to quantitatively benchmark computational performance of the method presented above.

FFIP was used throughout the methodology and case study sections in this article. However, we have undertaken an initial proof-of-concept study which shows favorable results for implementing the method in PRA. The main challenge in a PRA implementation is building out event trees and failure trees that can accept irrationality initiators. Existing event trees and fault trees may not sufficiently capture what happens in a system when an irrationality initiator is introduced.

The issue of cost effectiveness of implementing this method is an open question that remains to be resolved not only for this method but for the larger world of risk and failure analysis, reliability analysis, and safety engineering [114]. Safety engineering is generally viewed as a cost center rather than a profit center. In our professional experience, and in the experience of safety engineers we have discussed this issue with, it quite very rare quantify savings from safety analysis. From our discussions with leading system safety engineers around the world who are affiliated with INCOSE [114], the issue of justifying safety engineering is largely driven by regulatory compliance. Certain well-known engineering ethics examples such as the Ford Pinto [115] highlight the situation that faces systems engineers and high level management and indicate a need for further study of this important topic.

Note that we have intentionally omitted discussion of uncertainty in this article. Methods of understanding and quantifying uncertainty in probabilistic-based methods [116, 117, 118, 119] are appropriate to implement in the method presented above. Including uncertainty in probabilistic calculations may present interesting decision points in an analysis conducted using the methodology presented above.

In summary, the method introduced in this article provides a way for practitioners to begin identifying "unknown unknowns" [120] that may result in Sol and/or SoS failure. While there are some challenges in implementing the method, especially with regards to down-selecting the irrationality initiators and developing realistic probability statistics, the method produces useful results that can influence the design of a Sol. As far as we are aware, analyzing all potential failure flows that enter a Sol within a SoS in the context of irrational system behavior is a novel approach.

One area of future expansion of the work is a validation of the method presented here that uses controlled experiments conducted with systems engineers to compare our method with existing methodologies to determine if our method improves Sol and SoS outcomes. This proposed effort is likely a significant undertaking requiring a large participant pool participating in lengthy controlled experiments to achieve sufficient replicates which will help to gain statistical significance of the results. Other potential validation methods for new system design methodologies such as the mechanical design theory and methodology (DTM) community's Validation Square method [121] are not yet universally accepted by either the DTM community or the systems engineering community.

The analysis technique presented above that uses the informative and uninformative priors, and the dependent and independent irrationality initiators may be a useful starting point to further investigate how emergent behaviors are initiated in complex systems. While many systems have significant analysis and resources dedicated to independent initiating events, less work is done to investigate dependent events. In our professional practice, we have observed this is because of the assumed relative rarity of dependent initiating events. However, as many of the independent initiating events are now being effectively addressed in design, dependent initiating events appear to us to be becoming more important. Informative and uninformative priors may also prove a very fruitful area of future research to help discover high consequence, low probability initiating events which warrant more attention and which may be discounted or truncated in analysis methods such as PRA.

Another potential fruitful line of future research is reversing the analysis to start with assuming the Sol is behaving irrationally. The analysis would then focus on the Sol's impact within the rest of the SoS. Insights from this approach may show system designers how to reduce the likelihood of specific irrational failure flows from exiting the Sol and potentially adversely affecting other systems. This may help to improve the probability of a SoS successfully completing its mission.

Focusing on human behaviors in a SoS that may cause irrationality initiators may be a productive area of future research. The human factors literature may provide a good starting point for further investigations [122, 123, 124, 125]. The psychology and related literature on irrational human behaviors may also be useful [126, 127]. There is also some research in the engineering design community regarding functional analysis of human errors [128].

Some may find the usage of the term "irrational" to be controversial. Other terms such as "unexpected" and "not accounted for" may be more palatable for some readers. However, we have specifically retained the term "irrational" to call attention to the issue that the method presented above addresses. Until methodologies are developed that can automatically identify an exhaustive list of potential system behaviors including behaviors that we have termed "irrational" and behaviors that no one has either hypothesized or observed, we believe that the use of the word "irrational" is appropriate.

6 | CONCLUSION

This article introduces the concept of irrationality initiators as a method to improve failure analysis while developing conceptual functional models of a Sol in a SoS. Irrationality initiators allow practitioners to closely examine potential irrational system behaviors by other systems within a SoS that could have negative effects on the Sol. This may result in discovering new and unexpected system vulnerabilities. Once a failure scenario has been found that poses a significant threat to the Sol's continued operation or an SoS completing its mission (due to the failure of the Sol), a practitioner can undertake a redesign of the Sol to make it more robust and reliable. Multiple iterations of the method can result in potentially significant improvements in the likelihood that a Sol remains functional in spite of irrational system behaviors from other systems in the SoS and also increases the likelihood that the SoS completes its intended mission. Four different irrationality initiator analysis techniques are introduced in this paper including dependent and independent irrationality initiators and uninformative and informative priors. Each analysis technique provides a unique and useful perspective on potential emergent system behaviors and potential consequences caused by the irrationality initiators. While we demonstrated the method using FFIP and its associated extensions, the method shows promise for being useful with PRA as well. Other quantitative risk analysis methods may also prove to be compatible with irrationality initiators.

ACKNOWLEDGEMENTS

This research is partially supported by the Naval Postgraduate School (NPS), the Singapore University of Technology and Design (SUTD)¹, and Danmarks Tekniske Universitet. Any opinions or findings of this work are the responsibility of the authors, and do not necessarily reflect the views of the sponsors or collaborators. Please address all correspondence to D.L. Van Bossuyt.

REFERENCES

- [1] Musgrave GE, Larsen A, Sgobba T. Safety design for space systems. Butterworth-Heinemann; 2009.
- [2] of Aircraft Accidents Archives (B3A) TB. Statistics; 2019.
- [3] of State Dam Safety Officials A. Lessons Learned from Dam Incidents and Failures, Case Studies. 239 S. Limestone Street Lexington, KY 40508;.
- [4] Wheatley S, Sovacool BK, Sornette D. Reassessing the safety of nuclear power. *Energy Research & Social Science* 2016;15:96 – 100. <http://www.sciencedirect.com/science/article/pii/S2214629615301067>.
- [5] Ragheb M. FAULT TREE ANALYSIS AND ALTERNATIVE CONFIGURATIONS OF ANGLE OF ATTACK (AOA) SENSORS AS PART OF MANEUVERING CHARACTERISTICS AUGMENTATION SYSTEM (MCAS) 2019; <https://mrageb.com/FaultTreeAnalysisandEnhancedConfigurationsofAngleofAttackAoASensorsasPartofMCAS.pdf>.
- [6] Thurston RH. A History of the Growth of the Steam-Engine. No. 24, Kegan Paul; 1887.
- [7] Frenken K, Nuvolari A. The early development of the steam engine: an evolutionary interpretation using complexity theory. *Industrial and Corporate Change* 2004;13(2):419–450.

¹Dr. Arlitt was at the Singapore University of Technology and Design during the initial development of this research prior to moving to Danmarks Tekniske Universitet.

- [8] NASA. Preparation for Flight, the Accident, and Investigation: March 16 through April 5, 1967. NASA; 1967.
- [9] Commission USNR, et al. Reactor safety study. An assessment of accident risks in US commercial nuclear power plants. Executive summary. United States Nuclear Regulatory Commission; 1975.
- [10] Stamatelatos M, Dezfuli H, Apostolakis G, Everline C, Guarro S, Mathias D, et al. Probabilistic risk assessment procedures guide for NASA managers and practitioners. NASA; 2011.
- [11] Dean J, A Probability Risk Assessment to Support A Defendable and Quantitative Safety Assessment of the Assault Amphibious Vehicle. Calhoun; 2018.
- [12] Sierla S, O'Halloran BM, Karhela T, Papakonstantinou N, Tumer IY. Common cause failure analysis of cyber-physical systems situated in constructed environments. *Research in Engineering Design* 2013;24(4):375–394.
- [13] Taleb NN. Black swans and the domains of statistics. *The American Statistician* 2007;61(3):198–200.
- [14] Lipsky PY, Kushida KE, Incerti T. The Fukushima disaster and Japan's nuclear plant vulnerability in comparative perspective. *Environmental science & technology* 2013;47(12):6082–6088.
- [15] Executive Summary of Investigation Report into Train Collision at Joo Koon Station Westbound Platform on 15 November 2017 ("Incident"). Land Transport Authority; 2017.
- [16] Ullman D. The mechanical design process. McGraw-Hill Science/Engineering/Math; 2009.
- [17] Wall SD. Model-based engineering design for space missions. In: *Aerospace Conference, 2004. Proceedings. 2004 IEEE*, vol. 6 IEEE; 2004. p. 3907–3915.
- [18] Stone RB, Wood KL. Development of a functional basis for design. *Journal of Mechanical design* 2000;122(4):359–370.
- [19] Gertler JJ. Survey of model-based failure detection and isolation in complex plants. *IEEE Control systems magazine* 1988;8(6):3–11.
- [20] Kurtoglu T, Tumer IY, Jensen DC. A functional failure reasoning methodology for evaluation of conceptual system architectures. *Research in Engineering Design* 2010;21(4):209–234.
- [21] Haskins C, Forsberg K, Krueger M, Walden D, Hamelin D. *Systems engineering handbook*. INCOSE; 2006.
- [22] Jensen D, Tumer IY, Kurtoglu T. Flow State Logic (FSL) for analysis of failure propagation in early design. In: *International Design Theory and Methodology Conference, IDETC/CIE, ASME, San Diego, CA; 2009*. .
- [23] L'Her G, Van Bossuyt DL, O'Halloran BM. Prognostic systems representation in a function-based Bayesian model during engineering design. *International Journal of Prognostics and Health Management* 2017;8(2):23.
- [24] O'Halloran BM, Papakonstantinou N, Van Bossuyt DL. Modeling of function failure propagation across uncoupled systems. In: *Reliability and Maintainability Symposium (RAMS), 2015 Annual IEEE; 2015*. p. 1–6.
- [25] O'Halloran BM, Papakonstantinou N, Van Bossuyt DL. Cable routing modeling in early system design to prevent cable failure propagation events. In: *Reliability and Maintainability Symposium (RAMS), 2016 Annual IEEE; 2016*. p. 1–6.
- [26] Board BE, in Chief RJCE, editor, *The Guide to the Systems Engineering Body of Knowledge (SEBoK)*, v. 2.0. Hoboken, NJ: The Trustees of the Stevens Institute of Technology;. Accessed August 2019. www.sebokwiki.org.
- [27] Blanchard BS, Fabrycky WJ. *Systems Engineering and Analysis*. Prentice-Hall international series in industrial and systems engineering, Prentice Hall; 2011.
- [28] Crawley E, Cameron B, Selva D. *System Architecture: Strategy and Product Development for Complex Systems*. Always learning, Pearson; 2016. <https://books.google.com/books?id=67TuoQEACAAJ>.

- [29] Van Bossuyt DL, Beery P, O'Halloran BM, Hernandez A, Paulo E. The Naval Postgraduate School's Department of Systems Engineering Approach to Mission Engineering Education through Capstone Projects. *Systems* 2019;7(3):38.
- [30] Knochenhauer M, Louko P. Guidance for External Events Analysis. In: *Probabilistic Safety Assessment and Management* Springer; 2004. p. 1498–1503.
- [31] Siu N, Mosleh A, Meacham B. September 11, Columbia, Davis-Besse, and PRA: Business as Usual? In: *Probabilistic Safety Assessment and Management* Springer; 2004. p. 1026–1031.
- [32] Zio E. Challenges in the vulnerability and risk analysis of critical infrastructures. *Reliability Engineering & System Safety* 2016;152:137–150.
- [33] ICAM Architecture Part II-Volume IV - Function Modeling Manual (IDEF0). Materials Laboratory, Air Force Wright Aeronautical Laboratories, Air Force Systems Command, Wright-Patterson Air Force Base; 1981.
- [34] Friedenthal S, Moore A, Steiner R. A practical guide to SysML: the systems modeling language. Morgan Kaufmann; 2014.
- [35] Huang E, Ramamurthy R, McGinnis LF. System and simulation modeling using SysML. In: *Proceedings of the 39th conference on Winter simulation: 40 years! The best is yet to come* IEEE Press; 2007. p. 796–803.
- [36] Fowler M. UML distilled: a brief guide to the standard object modeling language. Addison-Wesley Professional; 2004.
- [37] Schmidt DC. Model-driven engineering. *COMPUTER-IEEE COMPUTER SOCIETY-* 2006;39(2):25.
- [38] Derler P, Lee EA, Vincentelli AS. Modeling cyber-physical systems. *Proceedings of the IEEE* 2012;100(1):13–28.
- [39] Hirtz J, Stone RB, McAdams DA, Szykman S, Wood KL. A functional basis for engineering design: reconciling and evolving previous efforts. *Research in engineering Design* 2002;13(2):65–82.
- [40] Bohm MR, Stone RB, Szykman S. Enhancing virtual product representations for advanced design repository systems. *Journal of Computing and Information Science in Engineering* 2005;5(4):360–372.
- [41] Gilchrist W. Modelling failure modes and effects analysis. *International Journal of Quality & Reliability Management* 1993;10(5).
- [42] Borgovini R, Pemberton S, Rossi M. Failure Mode, Effects and Criticality Analysis (FMECA). Reliability Analysis Center; 1993.
- [43] Stamatis DH. Failure mode and effect analysis: FMEA from theory to execution. ASQ Quality Press; 2003.
- [44] Mil-std-882e, department of defense standard practice system safety. Department of Defense; 2012.
- [45] Rovito SM, Rhodes DH. Enabling better supply chain decisions through a generic model utilizing cause-effect mapping. In: *2016 Annual IEEE Systems Conference (SysCon)* IEEE; 2016. p. 1–7.
- [46] Hampl V, FMEA and FTA; 2010.
- [47] Ericson CA. Fault tree analysis. *Hazard analysis techniques for system safety* 2005;p. 183–221.
- [48] Ericson C. Event tree analysis. *Hazard Analysis Techniques for System Safety* 2005;p. 223–234.
- [49] IAEA. Defining initiating events for purposes of probabilistic safety assessment. International Atomic Energy Agency; 1993.
- [50] Bedford T, Cooke R, et al. Probabilistic risk analysis: foundations and methods. Cambridge University Press; 2001.

- [51] Modarres M, Dezfuli H. A truncation methodology for evaluating large fault trees. *IEEE transactions on reliability* 1984;33(4):325–328.
- [52] Čepin M. Analysis of truncation limit in probabilistic safety assessment. *Reliability Engineering & System Safety* 2005;87(3):395–403.
- [53] Bloebaum CL, McGowan AMR. Design of complex engineered systems. *Journal of Mechanical Design* 2010;132(12):120301.
- [54] Mogul JC. Emergent (mis) behavior vs. complex software systems. In: *ACM SIGOPS Operating Systems Review*, vol. 40 ACM; 2006. p. 293–304.
- [55] Scholl HJ. Agent-based and system dynamics modeling: a call for cross study and joint research. In: *System Sciences, 2001. Proceedings of the 34th Annual Hawaii International Conference on IEEE*; 2001. p. 8–pp.
- [56] Leveson N. A new accident model for engineering safer systems. *Safety science* 2004;42(4):237–270.
- [57] Hirshorn SR, Voss LD, Bromley LK. *NASA Systems Engineering Handbook*. National Aeronautics and Space Administration; 2017.
- [58] Andersson G, Donalek P, Farmer R, Hatziaargyriou N, Kamwa I, Kundur P, et al. Causes of the 2003 major grid blackouts in North America and Europe, and recommended means to improve system dynamic performance. *IEEE Transactions on Power Systems* 2005;20(4):1922–1928.
- [59] Short AR, Lai AD, Van Bossuyt DL. Conceptual design of sacrificial sub-systems: failure flow decision functions. *Research in Engineering Design* 2017;p. 1–16.
- [60] Hawkins PG, Woollons DJ. Failure modes and effects analysis of complex engineering systems using functional models. *Artificial intelligence in engineering* 1998;12(4):375–397.
- [61] Nan C, Eusgeld I, Kröger W. Analyzing vulnerabilities between SCADA system and SUC due to interdependencies. *Reliability Engineering & System Safety* 2013;113:76–93.
- [62] Kean T. The 9/11 Commission Report. National Commission on Terrorist Attacks Upon the United States; 2004.
- [63] Vaughan D. The Challenger launch decision: Risky technology, culture, and deviance at NASA. University of Chicago Press; 1997.
- [64] The Chernobyl Accident: Updating of INSAG-1. International Nuclear Safety Advisory Group, International Atomic Energy Agency; 1992.
- [65] Wang JW, Rong LL. Cascade-based attack vulnerability on the US power grid. *Safety Science* 2009;47(10):1332–1336.
- [66] Sridhar S, Hahn A, Govindarasu M. Cyber–physical system security for the electric power grid. *Proceedings of the IEEE* 2012;100(1):210–224.
- [67] Keating CB. Emergence in system of systems. *System of Systems Engineering* 2008;p. 169–190.
- [68] Becker GS. Irrational behavior and economic theory. *Journal of political economy* 1962;70(1):1–13.
- [69] Harsanyi JC. Morality and the theory of rational behavior. *Social research* 1977;p. 623–656.
- [70] Loewenstein G. Out of control: Visceral influences on behavior. *Organizational behavior and human decision processes* 1996;65(3):272–292.
- [71] Link BG, Stueve A. Psychotic symptoms and the violent/illegal behavior of mental patients compared to community controls. *Violence and mental disorder: Developments in risk assessment* 1994;p. 137–159.

- [72] Carr CL. Coercion and freedom. *American Philosophical Quarterly* 1988;25(1):59–67.
- [73] Breton A. Manifesto of surrealism. *Manifestoes of Surrealism* 1924;15.
- [74] Caplan B. Terrorism: The relevance of the rational choice model. *Public Choice* 2006;128(1):91–107.
- [75] Van Bossuyt DL, Dong A, Tumer IY, Carvalho L. On measuring engineering risk attitudes. *Journal of Mechanical Design* 2013;135(12):121001.
- [76] Tversky A. Utility theory and additivity analysis of risky choices. *Journal of experimental psychology* 1967;75(1):27.
- [77] Wang XT. Domain-specific rationality in human choices: Violations of utility axioms and social contexts. *Cognition* 1996;60(1):31–63.
- [78] Fischer GW. Utility models for multiple objective decisions: do they accurately represent human preferences? *Decision sciences* 1979;10(3):451–479.
- [79] Slovic P, Fischhoff B, Lichtenstein S. Behavioral decision theory. *Annual review of psychology* 1977;28(1):1–39.
- [80] Einhorn HJ, Hogarth RM. Behavioral decision theory: Processes of judgement and choice. *Annual review of psychology* 1981;32(1):53–88.
- [81] Blais AR, Weber EU. A domain-specific risk-taking (DOSPRT) scale for adult populations. *Judgment and Decision making* 2006;1(1).
- [82] Hsu M, Bhatt M, Adolphs R, Tranel D, Camerer CF. Neural systems responding to degrees of uncertainty in human decision-making. *Science* 2005;310(5754):1680–1683.
- [83] Pine A, Seymour B, Roiser JP, Bossaerts P, Friston KJ, Curran HV, et al. Encoding of marginal utility across time in the human brain. *Journal of Neuroscience* 2009;29(30):9575–9581.
- [84] Van Bossuyt D, Hoyle C, Tumer IY, Dong A. Risk attitudes in risk-based design: Considering risk attitude using utility theory in risk-based design. *AI EDAM* 2012;26(4):393–406.
- [85] Valckenaers P, Van Brussel H, Bochmann O, Saint Germain B, Zamfirescu C, et al. On the design of emergent systems: an investigation of integration and interoperability issues. *Engineering applications of artificial intelligence* 2003;16(4):377–393.
- [86] Doya K, Uchibe E. The cyber rodent project: Exploration of adaptive mechanisms for self-preservation and self-reproduction. *Adaptive Behavior* 2005;13(2):149–160.
- [87] Ericson CA, et al. Hazard analysis techniques for system safety. John Wiley & Sons; 2015.
- [88] Taleb NN. The Black Swan: The Impact of the Highly Improbable. Random House; 2007.
- [89] Short AR, Van Bossuyt DL. Active mission success estimation through PHM-informed probabilistic modelling. *International Journal of Prognostics and Health Management* 2016;.
- [90] Bly M. Deepwater Horizon accident investigation report. Diane Publishing; 2011.
- [91] Slater MR, Van Bossuyt DL. Toward a dedicated failure flow arrestor function methodology. In: ASME 2015 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference American Society of Mechanical Engineers; 2015. p. V02AT03A050–V02AT03A050.
- [92] Coit DW, Liu JC. System reliability optimization with k-out-of-n subsystems. *International Journal of Reliability, Quality and Safety Engineering* 2000;7(02):129–142.

- [93] Aggarwal K. Redundancy optimization in general systems. *IEEE Transactions on Reliability* 1976;25(5):330–332.
- [94] Mitra S, Seifert N, Zhang M, Shi Q, Kim KS. Robust system design with built-in soft-error resilience. *Computer* 2005;38(2):43–52.
- [95] Benjamin PC, Erraguntla M, Mayer RJ. Using simulation for robust system design. *Simulation* 1995;65(2):116–128.
- [96] Galyean WJ, Kelly DL. *External Events Risk Analysis*. Idaho National Laboratory; 2009.
- [97] Hunsaker L. ARSENL reaches its ultimate goal of 50 autonomous UAVs in flight 2015;.
- [98] He R, Prentice S, Roy N. Planning in information space for a quadrotor helicopter in a GPS-denied environment. In: *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on IEEE*; 2008. p. 1814–1820.
- [99] Timble S. What Killed the Polecat? The DEW Line 2007 March; <https://web.archive.org/web/20120118144853/http://www.flightglobal.com/blogs/the-dewline/2007/03/what-killed-the-polecat.html>.
- [100] Mizokami K. Yemeni Rebels Capture a U.S. Navy Drone. *Popular Mechanics* 2018 January; <http://www.popularmechanics.com/military/weapons/a14748393/yemeni-rebels-capture-a-us-navy-drone/>.
- [101] of Defense UD, The DoDAF Architecture Framework Version 2.02;. <https://dodcio.defense.gov/library/dod-architecture-framework/>.
- [102] Standard Review Plan for the Review of Safety Analysis Reports for Nuclear Power Plants: LWR Edition. U.S. Nuclear Regulatory Commission; 2015.
- [103] Drago J, Borkowski R, Pike D, Goldberg F. In-plant reliability data base for nuclear power plant components: data collection and methodology report. Oak Ridge National Lab.; 1982.
- [104] Prem KP, Ng D, Mannan MS. Harnessing database resources for understanding the profile of chemical process industry incidents. *Journal of Loss Prevention in the Process Industries* 2010;23(4):549–560.
- [105] Le Bot P. Human reliability data, human error and accident models—illustration through the Three Mile Island accident analysis. *Reliability Engineering & system safety* 2004;83(2):153–167.
- [106] English D, Branaghan RJ. An empirically derived taxonomy of pilot violation behavior. *Safety science* 2012;50(2):199–209.
- [107] Liang G, Weller SR, Zhao J, Luo F, Dong ZY. The 2015 ukraine blackout: Implications for false data injection attacks. *IEEE Transactions on Power Systems* 2017;32(4):3317–3318.
- [108] Swain AD. Accident sequence evaluation program: Human reliability analysis procedure. Sandia National Labs., Albuquerque, NM (USA); Nuclear Regulatory Commission, Washington, DC (USA). Office of Nuclear Regulatory Research; 1987.
- [109] Dougherty E. Context and human reliability analysis. *Reliability Engineering & System Safety* 1993;41(1):25–47.
- [110] Kirwan B. Human reliability assessment. *Encyclopedia of Quantitative Risk Analysis and Assessment* 2008;.
- [111] Stein A, Lewis J, The Donald and The Nuclear VI: Pushin' My Buttons; 2018. <https://www.armscontrolwonk.com/archive/1204606/the-donald-and-the-nuclear-vi-pushin-my-buttons/>.
- [112] Burr W, Kimball JP. *Nixon's Nuclear Specter: The Secret Alert of 1969, Madman Diplomacy, and the Vietnam War*. University Press of Kansas; 2015.
- [113] Kang HG, Kim MC, Lee SJ, Lee HJ, Eom HS, Choi JG, et al. An overview of risk quantification issues for digitalized nuclear power plants using a static fault tree. *Nuclear Engineering and Technology* 2009;41(6):849–858.

- [114] Kemp D, O'Neal M, INCOSE International Workshop: System Safety Working Group; 2019.
- [115] Birsch D, Fielder J. The Ford Pinto Case: A Study in Applied Ethics, Business, and Technology. State University of New York Press; 1994.
- [116] Winkler RL. Uncertainty in probabilistic risk assessment. *Reliability Engineering & System Safety* 1996;54(2-3):127–132.
- [117] Parry GW. The characterization of uncertainty in probabilistic risk assessments of complex systems. *Reliability Engineering & System Safety* 1996;54(2-3):119–126.
- [118] Aven T, Baraldi P, Flage R, Zio E. Uncertainty in risk assessment: the representation and treatment of uncertainties by probabilistic and non-probabilistic methods. John Wiley & Sons; 2013.
- [119] Iman RL, Helton JC. The repeatability of uncertainty and sensitivity analyses for complex probabilistic risk assessments. *Risk Analysis* 1991;11(4):591–606.
- [120] Jorion P. Risk management lessons from the credit crisis. *European Financial Management* 2009;15(5):923–933.
- [121] Pedersen K, Emblemavag J, Bailey R, Allen JK, Mistree F. Validating design methods and research: the validation square. In: *ASME Design Engineering Technical Conferences*; 2000. p. 1–12.
- [122] Gertman D, Blackman H, Marble J, Byers J, Smith C, et al. The SPAR-H human reliability analysis method. *US Nuclear Regulatory Commission* 2005;230.
- [123] Chang Y, Mosleh A. Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents: Part 1: Overview of the IDAC Model. *Reliability Engineering & System Safety* 2007;92(8):997–1013.
- [124] Sheridan TB. Risk, human error, and system resilience: fundamental ideas. *Human factors* 2008;50(3):418–426.
- [125] Vinnem J, Bye R, Gran B, Kongsvik T, Nyheim O, Okstad E, et al. Risk modelling of maintenance work on major process equipment on offshore petroleum installations. *Journal of Loss Prevention in the Process Industries* 2012;25(2):274–292.
- [126] Akerlof GA, Yellen JL. Rational models of irrational behavior. *The American economic review* 1987;77(2):137–142.
- [127] Haselton MG, Ketelaar T. Irrational emotions or emotional wisdom? The evolutionary psychology of affect and social behavior. *Affect in social thinking and behavior* 2006;8:21.
- [128] Zurita NFS, Stone RB, Demirel O, Tumer IY. The Function-Human Error Design Method (FHEDM). In: *ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference American Society of Mechanical Engineers*; 2018. p. V007T06A058–V007T06A058.

DR. DOUGLAS L. VAN BOSSUYT is an Assistant Professor in the Systems Engineering Department at the Naval Postgraduate School. His research focuses on understanding and mitigating deleterious emergent system behaviors from a risk analysis and failure modeling perspective through the development of system design methodologies targeted at the system architecture phase of the system design process. He holds an Honors Bachelor of Science in Mechanical Engineering, an Honor Bachelor of Arts in International Studies, a Masters' of Science in Mechanical Engineering, and a PhD in Mechanical Engineering all from Oregon State University.

DR. BRYAN M. O'HALLORAN is an Assistant Professor in the Systems Engineering Department at the Naval Postgraduate School (NPS) and the Academic Associate for the Reliability and Maintainability certificate program (curriculum 242). Previously he was a Senior Reliability and Systems Safety Engineer at Raytheon Missile Systems (RMS) and the Lead Reliability and Safety Engineer for hypersonic missile programs. He holds a Bachelor of Science degree in Engineering Physics and a Master of Science and Doctorate of Philosophy in Mechanical Engineering from Oregon State University. His current research interests include risk, reliability, safety, and failure modeling in the early design of complex systems.

DR. RYAN M. ARLITT is an Assistant Professor in the Department of Mechanical Engineering at the Technical University of Denmark. His research focus is on understanding (1) how successful designers solve complex conceptual design challenges, and (2) how computational support can improve the likelihood and quality of success in conceptual design and beyond. He holds a PhD in Mechanical Engineering from Oregon State University, and Bachelors and Masters Degrees in Interdisciplinary Engineering and Systems Engineering respectively from the Missouri University of Science and Technology.