Irrational System Behavior in a System of Systems

Conference Paper · June 2018

DOI: 10.1109/SYSOSE.2018.8428778

3 authors:

Douglas Lee Van Bossuyt
Naval Postgraduate School
135 PUBLICATIONS **936** CITATIONS

SEE PROFILE

Bryan O'Halloran
Naval Postgraduate School
64 PUBLICATIONS **476** CITATIONS

SEE PROFILE

Ryan Arlitt
Singapore University of Technology and Design
43 PUBLICATIONS **208** CITATIONS

SEE PROFILE

Irrational System Behavior in a System of Systems

Douglas L. Van Bossuyt, Bryan M. O'Halloran
Department of Systems Engineering
Naval Postgraduate School
Monterey, California 93943, USA
douglas.vanbossuyt@nps.edu, bmohallo@nps.edu

Ryan M. Arlitt
SUTD-MIT International Design Centre
Singapore University of Technology and Design, Singapore
arlitt.ryan@gmail.com

***Abstract*—System of systems (SoS) failures can sometimes be traced to a system within the SoS behaving in unexpected ways. Due to their emergent complexity, these types of failures are notoriously challenging to foresee. This paper presents a method to aid in predicting unknown unknowns in a SoS. Irrationality initiators – failure flows emanating from one system that serve as unexpected initiating events in another system – are introduced into quantitative risk analysis methods such as the Failure Flow Identification and Propagation framework and Probabilistic Risk Assessment. Analysis of models built using this approach yield a probability distribution of failure paths through a system within the SoS that are initiated by unexpected behaviors of other systems within the SoS. The method is demonstrated using an example of an autonomous vehicle network operating in a partially denied environment with hostile forces present. Using the concept of irrationality initiators, it is possible to identify and prioritize vulnerabilities in the system of interest in the SoS.**

I. Introduction

While emergent system behaviors have long been observed in engineered systems, concerted efforts to understand harmful emergent system behavior and protect against it have only been undertaken in recent decades. Much of the research and professional practice in emergent system behavior has focused on understanding predictable and expected events [1]. However, events can occur which have not been predicted or observed before [2], or which were ruled out as not likely enough to occur to not require further analysis [3]. For example, a series of unexpected interactions in the signaling system for Singapore's Mass Rapid Transit system led to a progression of undetected and degraded operating conditions. Eventually these conditions resulted in a collision that caused 38 injuries [4]. The authors propose considering these emergent behaviors to be irrational system behaviors – unexpected behaviors within a system that produce failure initiators in other systems within a System of Systems (SoS). Irrational system behaviors are those that have not been observed within the system previously and have not been predicted or analyzed through routine means of system simulation and failure analysis.

A logical and probabilistic approach to analyzing a system often fails to uncover potential failure scenarios that are initiated by an event that is seemingly irrational even with extensive guidance on searching for potentially overlooked initiating events [5]. Organizations conducting failure and risk analysis of complex systems sometimes overlook potential failure scenarios that have been identified – a failure scenario can be identified but discounted and not rigorously analyzed [6]. As systems are connected together to create a SoS, new behavioral properties can emerge from one or more systems behaving in unanticipated ways [7], [8]. In short, complex SoS have unexpected and irrational system behaviors that can result in severe consequences to the SoS.

In order to develop SoS that are more robust, resilient, interoperable, and survivable; and that are more able to complete their intended missions, it is important to seek out and better understand irrational failure events. While existing SoS can be analyzed, it is more cost effective to find potential problems early during the conceptual stage of design [9]. Modeling failure likelihoods of systems helps to better understand which failures are more likely to occur and thus more of a priority to address [10]. The architecture of a SoS can be iterated upon multiple times until the SoS has an acceptable probability of failure. Methods such as those developed around the Function Failure Identification and Propagation (FFIP) method [11]–[13] can assess failure propagation through systems. Similarly, Probabilistic Risk Assessment (PRA)-based methods of risk analysis can be used to assess failure propagation through a system and is especially well-suited for systems with multiple redundant subsystems and mitigating systems [1]. However, such analyses often miss failures that are a result of irrational initiating events. In SoS, irrational system behavior of one of the systems within the SoS currently cannot be well represented in failure and risk analysis methods. Practitioners currently lack a means to systematically identify potential irrational system behaviors and analyze the results of one system behaving irrationally on a SoS.

A. Specific Contributions

The contribution of this research is the development of a method to consider the impact of irrational system behavior of a system on another system within a SoS environment. The method is applicable during the initial phase of systems design where functional and physical architectures are being developed. An analysis of potential effects (i.e., the *method*) caused by irrational system behaviors emanating from one system and impacting another system is conducted. Results of the analysis can be used to develop and refine individual system models to improve system and SoS robustness to irrational system behaviors.

II. Background and Related Work

The method developed in this research relies upon several bodies of work including systems modeling, failure analysis, and probability assessment. This section provides background to the research presented in this paper.

Systems modeling is a method of developing models that can be used to represent and to simulate systems. Many modeling techniques exist such as IDEF0 [14], System Modeling Language (SysML) [15], and others [16], [17]. This paper uses the Functional Basis for Engineering Design (FBED) [18] hierarchical taxonomy to represent systems. FBED models systems from a function and flow paradigm where functions are the actions that a system can take (e.g., transport material, convert electrical energy to rotational energy, etc.) and flows are material, energy, or signal moving around within the system (e.g., energy-chemical, signal-control-discrete, etc.) or between systems in a SoS. The function and flow taxonomies can each be broken down into primary, secondary, and tertiary categories with an increasing level of specificity. FBED intentionally abstracts system components from functions and flows to provide a means for system design and analysis that largely avoids pre-conceived notions of what components a system will use to meet functional requirements – this positions FBED to be used during the conceptual phase of system design.

Failure analysis is conducted on systems to understand how systems may fail during operation. One popular method of failure analysis is the Failure Modes and Effects Analysis (FMEA) [19] method and the related Failure Modes Effects and Criticality Analysis (FMECA) [20] method; both see extensive use in a variety of industries. FMEA uses a risk priority number (the multiplication of probability of occurrence of the failure, detectability of the failure, and severity of the failure each on a 0-10 scale) to provide guidance to systems engineers on the urgency of addressing potential system failures. However, FMEA often misses emergent system behaviors where multiple components or subsystems fail, or multiple conditions combine to cause a system-level failure.

Probabilistic Risk Assessment (PRA) combines fault tree analysis (FTA) [21] and event tree analysis (ETA) [22] to develop failures that progress through a system and require multiple components or subsystems to fail to lead to system failure. Initiating events are developed to provide the probability of an event occurring that initiates a potential system failure [23]. However, initiating events that are discounted as being possible or that are beyond prior experience with the system or similar systems can be missed [5].

Engineered systems can have episodes of irrational system behavior for a variety of reasons [24]. Much effort has been put into developing methods to understand and address detrimental emergent system behaviors [25], and making systems more robust and resilient to external and internal perturbations [26]. However, little work has been done to understand irrational system behavior within the context of a SoS.

Several efforts have been made to combine functional modeling of systems with failure analysis such as a method of developing FMEAs for functional models [27] and probabilistically assessing failure propagation through systems using functional models in the Failure Flow Identification and Propagation (FFIP) methodology [11]. Additional work has been done to extend FFIP such as with the Uncoupled Failure Flow State Reasoner (UFFSR) [13] that evaluates failure flows that do not follow nominal flow pathways in functional system models and a Bayesian approach to developing prognostic and health monitoring subsystems to detect incipient failures while they still can be corrected [12]. Initiating events are developed in much the same way as they are in PRA which leads to the same issue of unexpected or irrational initiating events not being considered in the analysis.

III. Methodology

The method presented here is useful during the early conceptual phase of system design when architectural trade-off studies are conducted. At this point in design, decisions that radically alter the design of a system are inexpensive in both time and resources. As such, this method can be re-applied until the practitioner is sufficiently satisfied with the results.

A. Model the Systems

The first step in the method is to model the systems and their connections with one another. FBED [18] is the authors' preferred functional modeling method, although many other methods exist [16], [17]. The models can be implemented in a variety of modeling languages such as SysML [15].

Following the development of system models, failure analysis models must then be created. The family of failure modeling techniques, based around the FFIP framework [11], is used in this research [12], [13]. However, using PRA [1] as the basis for failure analysis is equally acceptable. Regardless of the failure analysis method used, the method must be quantifiable to produce meaningful results that systems architecture decisions can be based upon.

B. Define Potential Irrational System Behaviors: Irrationality Initiators

Irrational behavior of people has long been studied in the context of economic models [28], [29]. Irrational behavior (also often called irrationality) can take different forms and have different causes such as visceral reactions [30] to events, psychosis [31], actions taken under duress [32], or even intentional irrationality [33]. Within an engineering context, design engineers can appear to behave irrationally although irrational behavior can sometimes be explained by an engineer's personal utility functions [34]. It is possible to develop system models that do not adhere to the expected value theorem, and instead match an individual's or an organization's utility function [35]. While it may appear to outside observers that a system is behaving irrationally, it is possible that the system's utility function is significantly different from external observer expectations – the system is behaving normally based on its utility function but abnormally based on external expectations.

Irrational system behavior is defined for the purposes of this method as the functional flows that exit a system boundary being illogical or unreasonable compared to the expected and previously experienced system behavior. Illogical or unreasonable behavior is defined as deviation from preprogrammed behaviors and rational expectations [36]; unresponsiveness to incentives [33]; and deviation from self-interest, self-preservation, and/or system-of-systems self interest and preservation [37]. Within the context of this research, the definition of irrational system behavior is further refined as being a class of failure flow [11] that would not normally be anticipated through common failure analysis techniques (e.g., FMEA [19], PRA [1], FFIP [11], UFFSR [13], etc.). This is similar to the concept of Black Swan events as popularized by Taleb [2] although with focus on the initiating event aspect of Black Swan events rather than on the overall system failure.

Initiating events used in failure and risk analysis (e.g., PRA, FFIP, etc.) are events that initiate a failure within a system or a SoS. The failure then propagates through the system until either (1) the system fails, (2) the system is operating in a degraded but stable state, or (3) the system recovers from or mitigates the failure and continues operating normally. While there are well-defined procedures for identifying potential initiating events that can impact a system [23], initiating events that appear to be irrational or that are beyond prior experience can be missed or discounted [5].

The authors propose the concept of irrationality initiators (irrational behavior initiating events) to supplement existing methods of identifying initiating events for failure and risk analysis. Irrationality initiators are caused by a system within a SoS behaving irrationally and emitting failure flows beyond the system boundary that are unexpected within the realm of standard failure and risk analysis tools. The failure flows become irrationality initiators when they come into contact with other systems within the SoS. Irrationality initiators may follow nominal flow paths between systems or they may affect systems through propagation pathways that are not normally active or connected. The reason for distinguishing irrationality initiators from failure flows is to clearly denote that they are initiating events from outside of a system that impact the system. Irrationality initiators have the potential to cause a failure that propagates through a system and leads to a variety of failed, partially failed, or nominal system states. A system being analyzed within a SoS has irrationality initiators from **other** systems impacting it – this is an important distinction; however, a system being analyzed may in turn emit irrationality initiators that impact other systems during later analysis.

In order to identify potential irrationality initiators, the authors of this research propose the following approach, as shown in Figure 1.

Step 1: Begin with **all** of the secondary and tertiary flow descriptions from the FBED functional modeling methodology. Each flow could conceivably be an irrationality initiator coming from a generic black box system within the 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: Remove all flows from the list of potential irrationality initiators that are already modeled as initiating events through failure analysis methods such as FFIP or PRA.

Step 3: Identify any candidate irrationality initiators that appear to be impossible for the black box system that is behaving irrationally to emit. Attempt to identify ways that the irrationality initiators may be able to be generated. Some methods of generating irrationality initiators may be outlandish but should be noted regardless. For instance, almost any material can be forced to produce unexpected spectral emissions with sufficient energy being imparted to the material.

Step 4: Assign probabilities of occurrence to each item on the list of remaining irrationality initiators. Follow guidance on developing initiating event probabilities for PRA, such as provided by [1] and [23].

For irrationality initiators that have an understood and known probability of occurrence, the known value is used. Otherwise, a value of 3X the highest known probability of any irrationality initiator is assumed. This approach for choosing unknown values motivates the analyst to identify suitable probability data in situations where a high probability of occurrence significantly impacts the relative ranking of a failure path in comparison with other failure paths analyzed in a system – in many cases, the probability of the irrationality initiator will not greatly impact the relative ranking. While a much higher multiplier could be used to highlight potential irrationality initiators that require further investigation, setting the multiplier too high on irrationality initiators that do not have a rigorously analyzed and supported probability of occurrence may unduly burden systems engineers with unneeded analysis.

Note that additional failure model development may be necessary to implement analysis of the irrationality initiators. FFIP and UFFSR failure modeling techniques can be used to fully implement the necessary failure model additions in a function failure analysis. PRA has the flexibility to add appropriate FTAs and ETAs to implement assessment of irrationality initiators. Depending upon the complexity of the system and the extent of previous modeling efforts, potentially extensive failure model development may be needed.

C. Analyses of Potential Irrational System Behaviors

In the case study in this paper, the analysis step uses the FFIP family of tools to develop failure paths and their associated probabilities. However, PRA can work equally well in developing failure models and associated probabilities. The rest of this paper uses the FFIP family of tools. In either case, practitioners shall use the established analysis method of choice with the addition of irrationality initiators, and with sufficient modifications to the models to allow the impact of irrationality initiators to be modeled in the system.

Specific Guidance on Modeling Implementation with FFIP: Within the FFIP family of tools, each function's response to a variety of failure inputs is modeled with potential results including failure of the function, cessation of nominal flows emanating from the function, failure flows being passed through the function, failure flows being rejected by the function, or new failure flows being generated by the function. Each possible outcome is assigned a probability which is used

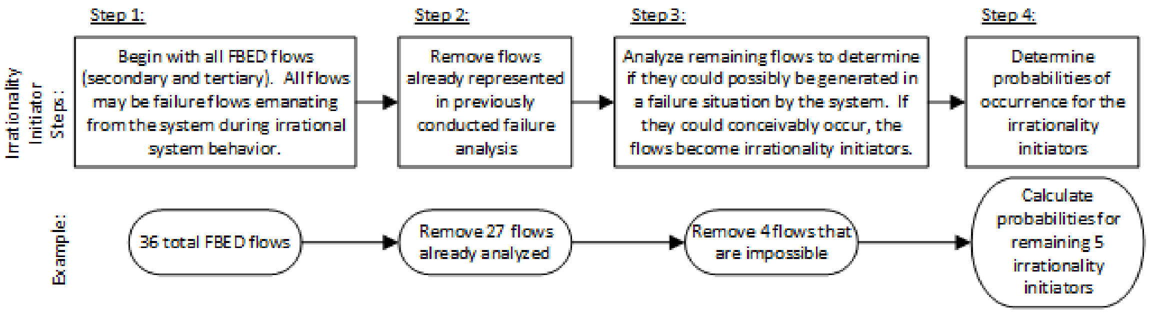
Irrationality Initiator Steps:
Step 1:
Begin with all FBED flows (secondary and tertiary). All flows may be failure flows emanating from the system during irrational system behavior.
Step 2:
Remove flows already represented in previously conducted failure analysis
Step 3:
Analyze remaining flows to determine if they could possibly be generated in a failure situation by the system. If they could conceivably occur, the flows become irrationality initiators.
Step 4:
Determine probabilities of occurrence for the irrationality initiators
Example:
36 total FBED flows
Remove 27 flows already analyzed
Remove 4 flows that are impossible
Calculate probabilities for remaining 5 irrationality initiators

Fig. 1. Steps to Developing Irrationality Initiators.

to calculate the probability of specific combinations of failure paths through a system.

Irrationality initiators may enter a system through nominal flow paths that cross the system boundary. In this case, the irrationality initiators can be modeled using FFIP. Irrationality initiators may also cross a system boundary and impact specific functions that do not have nominal flow paths entering or exiting the system. In this case, the UFFSR methodology is useful in modeling how irrationality initiators can make the jump into a system within a SoS. It is likely that failure flows will not always travel along nominal flow paths and instead will jump between systems through uncoupled paths, as has been seen in a variety of accidents such as the Deepwater Horizon. Further guidance on developing uncoupled flow propagation probabilities and analyses is provided in [13].

D. Analyze Results and Improve System Robustness

The results of the analysis provides specific insights into potential impacts of irrational system behavior on a system within a SoS. The analysis results are directly comparable to the results of a failure analysis conducted using FFIP or PRA as appropriate. It is useful to both compare the results of this analysis method with the results generated from FFIP and related methods, and sum the failure probabilities from both this method and the FFIP and related methods analysis (the same is true for PRA-based analyses). Comparing the results is useful in understanding if irrationality initiators are a potential large contributor to the probability of system failure. Summing the results provides a more complete overall picture of the system's risk of failure. Note that irrationality initiators that previously did not have a realistic probability assigned and were instead given a high probability of occurrence may disproportionately show up in the results of the analysis methods (e.g., FFIP or PRA, etc.). If they do rise to the top of the failure path list, then further effort can be expended on developing a more realistic probability for the irrationality initiators in question. However, if the probabilities of system failures caused by irrationality initiators assigned a default large probability of occurrence (3X the largest probability of occurrence of any irrationality initiator) are low relative to other failures within the system, the irrationality initiators in question likely do not need further analysis.

IV. Case Study and Results

The following case study demonstrates the method proposed in this paper. An explicitly fictional SoS is used. Intentionally fictional probabilities are used in the analysis. The case study is for demonstration purposes only and cannot be used to draw conclusions on any existing or proposed SoS.

A group of autonomous vehicles is operating in a partially denied environment where Global Positioning System (GPS) coverage is not available and other broad-area navigational aids (e.g., celestial navigation, way-point navigation, etc.) are not available. The rovers receive their positioning information and communicate with a command and control station via a series of active radar stations that also contain two-way communications equipment in a nodal configuration. The active radar station locations are not optimal due to local topography considerations and hostile force actions.

The autonomous vehicles have poor internal positioning accuracy and require regular positioning updates from the radar stations to stay on course. Sensitive supplies are being carried aboard the autonomous vehicles from the command and control station to an outpost. If positioning information is lost for more than 3 minutes, it is assumed that the autonomous vehicle has sufficiently deviated from a safe path to warrant destruction of the autonomous vehicle to prevent them from falling into enemy hands.

A defense contractor is preparing a new version of the autonomous vehicle to enter service. The radar systems and the command and control system are manufactured by a separate contractor. Integration of the various systems into the large SoS is handled by a third contractor, as is often found in defense systems. The defense contractor in charge of the autonomous vehicles desires to understand how irrationality initiators caused by other systems can impact individual autonomous vehicles, and how the autonomous vehicles can be made more robust and reliable within the SoS to increase the likelihood of the SoS's mission success.

A. Model the Systems

Figure 2 shows the system functional model for an autonomous vehicle. Table I shows the top five failure pathways and their associated probabilities as located with FFIP. Note that failure in the context of the autonomous vehicle is defined as the cargo being damaged or lost and not reaching its intended destination. Further note that this analysis could be conducted using PRA but that it is conducted using FFIP here. Not shown here due to space limitations is the SoS model which includes the command center, the outpost, several autonomous vehicles, and several chained-together radar stations.

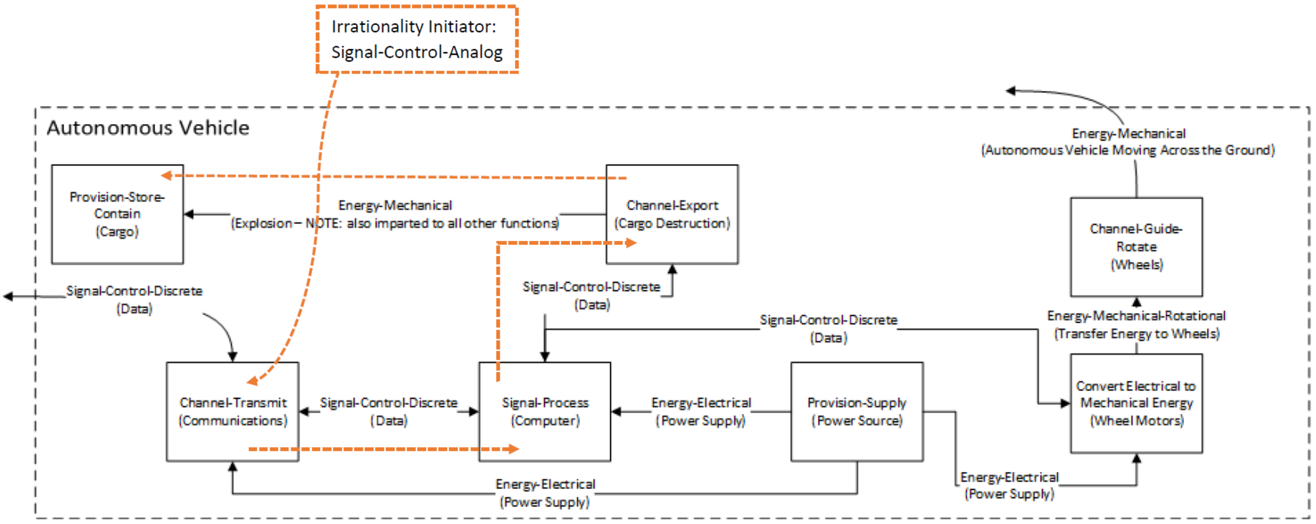

Fig. 2. High-level autonomous vehicle system functional model. Many functions and flows have been excluded from this functional model for brevity and ease of understanding the simplified case study. The dashed box indicates the system boundary. The orange dashed lines and boxes represent a potential irrationality initiator failure path. The system fails when the failure flows caused by the irrationality initiator reach the Provision-Store-Contain (Cargo) function which is assumed to lead to loss of the cargo.

TABLE I. TRUNCATED LIST OF HIGHEST PROBABILITY OF FAILURE FFIP RESULTS.

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

B. *Define Potential Irrational System Behaviors*

Next, potential irrational system behaviors of the other systems within the SoS are examined. In this case study, only potential irrational system behaviors from the radar stations are examined for demonstration purposes. In a complete analysis, all systems would be analyzed.

First, the entire list of flows from FBED is put into a table, as seen in Table II. Second, the flows already represented in FFIP and other analyses already conducted are removed, as indicated with horizontal lines through the particular flows. Third, each remaining flow is validated as being a potential failure flow and completely impossible flows are removed, as indicated by crossed out lines. The validation can be conducted in a variety of ways and with a varying fidelity levels depending upon the needs of the SoS being analyzed. For instance, a complete workbook similar to individual initiating events in a nuclear PRA model [38] can be generated or a very high level back of the envelope justification can be made. The remaining flows have now been identified as irrationality initiators. Fourth, probabilities are assigned to each irrationality initiator.

The Signal-Status-Auditory (i.e., noise) flow is used as an example to demonstrate how to determine if a flow should be included as an irrationality initiator. First, this flow was not analyzed in the FFIP analysis as being something that could affect the autonomous vehicle, and therefore, the flow is further considered as a potential irrationality initiator. When considering the flow as a candidate irrationality initiator, it is noted that the autonomous vehicles will never be close enough to the radar stations for any noise generated by the radar stations to be sufficiently intense to impact operations of the autonomous vehicles. The Signal-Status-Auditory flow is therefore struck from the list of potential irrationality initiators.

Three irrationality initiators are identified in Table II. The Material-Solid-Object flow is representative of part or all of a radar station becoming dislodged from atop a mountain, rolling down the mountain, and colliding with an autonomous vehicle. A variety of instigators of this irrationality initiator were found including the ledge on which the radar station was placed gives way, enemy forces push the radar station off the edge of the mountain, high winds dislodge a parabolic dish at the radar station and the dish hits autonomous vehicle, and other potential reasons for this irrational system behavior which previously was not considered. The Material-Solid-Object irrationality initiator was assigned a relatively high likelihood of occurrence on a per day basis based on regular high winds in the area and unique geographical features at two of the radar station sites that would cause a tumbling radar station to be funneled down a ravine and directly into the path of the autonomous vehicles. The other irrationality initiators were similarly studied and appropriate values were assigned.

While not shown here due to space limitations, additional failure model development was performed to implement potential failure flow paths and system behaviors as a result of the irrationality initiators. A limited graphical demonstration of added failure flow paths in the system is shown in Figure 2

TABLE II. Irrationality initiators are developed from the FBED flow set. Refer to [18] for FBED flow set details.

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~~	
		~~Colloidal~~	
Signal	~~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~~		

where the orange lines indicate a failure flow path caused by the Signal-Control-Analog irrationality initiator.

C. Analyses of Potential Irrational System Behaviors

Using the models and the irrationality initiators developed above, the analysis is conducted. Truncated results of the analysis methods are shown in Table III.

D. Analyze Results and Improve System Robustness

The analysis shows that the Signal-Control-Analog irrationality initiator is a significant contributor to probability of system failure. This may indicate that confirmation of the probability assigned to the irrationality initiator needs to be revisited or this may indicate that a redesign of the system to protect against this failure is necessary. Alternatively, this could indicate a need to protect the system against the irrationality initiator to prevent potential failures.

V. Discussion and Future Work

The method presented in this work presents several interesting points of discussion. This section discusses benefits and limitations of the method, philosophical questions of importance to practitioners, and future directions of research.

The insights provided by the method allows practitioners to compare irrationality initiator-derived failures to failures derived from FFIP or PRA in a quantitative manner, and provides a view of how important failures induced by irrationality initiators are compared to other failures within the system.

Identifying irrationality initiators and developing realistic probabilities can be a challenge for this method. It is difficult to say with a high degree of certainty whether each irrationality initiator is possible or not possible; it is more difficult to develop realistic probabilities that can be backed up with quantitative data. In spite of these limitations, the authors believe that the unique insights offered by this method are sufficiently beneficial to systems engineers that analysis using this method should be conducted on SoS.

The method presented here provides a tool for practitioners to begin to identify unknown unknowns within the context of failure and risk analysis. While a human is still needed at this point in the method's development to down-select potential irrationality initiators which can lead to biases and discounting potential irrationality initiators that in fact do play a role in system failures, the analysis of all potential flows crossing a system boundary as potential irrationality initiators is a novel way of generating irrationality initiators. In the future, an automated method of generating and assessing irrationality initiators may be developed which would allow less human bias to be introduced to the process.

A potential interesting future line of research may take the form of reversing the analysis and assuming that the system of interest within the SoS is behaving irrationally. The analysis then would focus on the impact the irrational system has on the rest of the SoS. This may provide insights on how to design individual systems to minimize irrational system behavior impacts on the SoS which in turn could improve the likelihood of a SoS completing its mission.

Another potential interesting expansion of this research is to investigate an uninformative prior distribution for irrationality initiators to examine outcomes of different scenarios that would not normally be given much attention do to the irrationality initiators being low probability. Further, investigating dependent irrationality initiators where multiple initiators happen at once may reveal interesting emergent system behaviors that otherwise are not analyzed with independent irrationality initiators. For example, there may be specific scenarios where irrationality initiators are known to be coupled outside the system boundary, and conducting comprehensive what-if studies based on combinatorial sets of initiators may reveal potential new failure scenarios.

VI. Conclusion

This paper introduces irrationality initiators as a tool to increase the power of failure analyses in the conceptual phase of system design. Using irrationality initiators within the FFIP analysis framework or with PRA provides a new means to uncover unexpected vulnerabilities in SoS. By treating a system model as a space over which to analyze a wide variety of potential initiating events, it is possible to assess the vulnerability of SoS concepts to unexpected or irrational inputs. This approach provides a way to uncover (1) vulnerable sections of a system and (2) the relative vulnerability of a system to a variety of unexpected initiating events.

Acknowledgment

This research is partially supported by the Naval Postgraduate School (NPS) and the Singapore University of Technology

TABLE III. TRUNCATED LIST OF FAILURE PATHS OF THE ANALYSIS METHODS.

Failure Propagation Pathway	**Probability**
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, Chanel-Guide-Rotate	2.9E-6/day

and Design (SUTD). 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] M. Stamatelatos, H. Dezfuli, G. Apostolakis, C. Everline, S. Guarro, D. Mathias, A. Mosleh, T. Paulos, D. Riha, C. Smith *et al.*, *Probabilistic risk assessment procedures guide for NASA managers and practitioners*. NASA, 2011.

[2] N. N. Taleb, *The Black Swan: The Impact of the Highly Improbable*. Random House, 2007.

[3] P. Y. Lipscy, K. E. Kushida, and T. Incerti, "The fukushima disaster and japans nuclear plant vulnerability in comparative perspective," *Environmental science & technology*, vol. 47, no. 12, pp. 6082–6088, 2013.

[4] "Executive summary of investigation report into train collision at joo koon station westbound platform on 15 november 2017 ("incident")," Land Transport Authority, Investigation Report, 2017.

[5] M. Knochenhauer and P. Louko, "Guidance for external events analysis," in *Probabilistic Safety Assessment and Management*. Springer, 2004, pp. 1498–1503.

[6] "The chernobyl accident: Updating of insag-1," International Nuclear Safety Advisory Group, International Atomic Energy Agency, Safety Series No. 75-INSAG-7, 1992.

[7] J.-W. Wang and L.-L. Rong, "Cascade-based attack vulnerability on the us power grid," *Safety Science*, vol. 47, no. 10, pp. 1332–1336, 2009.

[8] C. B. Keating, "Emergence in system of systems," *System of Systems Engineering*, pp. 169–190, 2008.

[9] D. Ullman, *The mechanical design process*. McGraw-Hill Science/Engineering/Math, 2009.

[10] J. J. Gertler, "Survey of model-based failure detection and isolation in complex plants," *IEEE Control systems magazine*, vol. 8, no. 6, pp. 3–11, 1988.

[11] T. Kurtoglu, I. Y. Tumer, and D. C. Jensen, "A functional failure reasoning methodology for evaluation of conceptual system architectures," *Research in Engineering Design*, vol. 21, no. 4, pp. 209–234, 2010.

[12] G. L'Her, D. L. Van Bossuyt, and B. M. O'Halloran, "Prognostic systems representation in a function-based bayesian model during engineering design," *International Journal of Prognostics and Health Management*, vol. 8, no. 2, p. 23, 2017.

[13] B. M. O'Halloran, N. Papakonstantinou, and D. L. Van Bossuyt, "Modeling of function failure propagation across uncoupled systems," in *Reliability and Maintainability Symposium (RAMS), 2015 Annual*. IEEE, 2015, pp. 1–6.

[14] "Icam architecture part ii-volume iv - function modeling manual (idef0)," 1981.

[15] S. Friedenthal, A. Moore, and R. Steiner, *A practical guide to SysML: the systems modeling language*. Morgan Kaufmann, 2014.

[16] D. C. Schmidt, "Model-driven engineering," *COMPUTER-IEEE COMPUTER SOCIETY-*, vol. 39, no. 2, p. 25, 2006.

[17] P. Derler, E. A. Lee, and A. S. Vincentelli, "Modeling cyber–physical systems," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 13–28, 2012.

[18] R. B. Stone and K. L. Wood, "Development of a functional basis for design," *Journal of Mechanical design*, vol. 122, no. 4, pp. 359–370, 2000.

[19] W. Gilchrist, "Modelling failure modes and effects analysis," *International Journal of Quality & Reliability Management*, vol. 10, no. 5, 1993.

[20] "System safety," Department of Defense, Standard Practices MIL-STD-882E, 2012.

[21] C. A. Ericson, "Fault tree analysis," *Hazard analysis techniques for system safety*, pp. 183–221, 2005.

[22] C. Ericson, "Event tree analysis," *Hazard Analysis Techniques for System Safety*, pp. 223–234, 2005.

[23] IAEA, "Defining initiating events for purposes of probabilistic safety assessment," International Atomic Energy Agency, Tech. Rep. IAEA-TECDOC-719, 1993.

[24] C. L. Bloebaum and A.-M. R. McGowan, "Design of complex engineered systems," *Journal of Mechanical Design*, vol. 132, no. 12, p. 120301, 2010.

[25] J. C. Mogul, "Emergent (mis) behavior vs. complex software systems," in *ACM SIGOPS Operating Systems Review*, vol. 40, no. 4. ACM, 2006, pp. 293–304.

[26] S. R. Hirshorn, L. D. Voss, and L. K. Bromley, "Nasa systems engineering handbook," National Aeronautics and Space Administration, Tech. Rep., 2017.

[27] P. G. Hawkins and D. J. Woollons, "Failure modes and effects analysis of complex engineering systems using functional models," *Artificial intelligence in engineering*, vol. 12, no. 4, pp. 375–397, 1998.

[28] G. S. Becker, "Irrational behavior and economic theory," *Journal of political economy*, vol. 70, no. 1, pp. 1–13, 1962.

[29] J. C. Harsanyi, "Morality and the theory of rational behavior," *Social research*, pp. 623–656, 1977.

[30] G. Loewenstein, "Out of control: Visceral influences on behavior," *Organizational behavior and human decision processes*, vol. 65, no. 3, pp. 272–292, 1996.

[31] B. G. Link and A. Stueve, "Psychotic symptoms and the violent/illegal behavior of mental patients compared to community controls," *Violence and mental disorder: Developments in risk assessment*, pp. 137–159, 1994.

[32] C. L. Carr, "Coercion and freedom," *American Philosophical Quarterly*, vol. 25, no. 1, pp. 59–67, 1988.

[33] B. Caplan, "Terrorism: The relevance of the rational choice model," *Public Choice*, vol. 128, no. 1, pp. 91–107, 2006.

[34] D. L. Van Bossuyt, A. Dong, I. Y. Tumer, and L. Carvalho, "On measuring engineering risk attitudes," *Journal of Mechanical Design*, vol. 135, no. 12, p. 121001, 2013.

[35] D. Van Bossuyt, C. Hoyle, I. Y. Tumer, and A. Dong, "Risk attitudes in risk-based design: Considering risk attitude using utility theory in risk-based design," *AI EDAM*, vol. 26, no. 4, pp. 393–406, 2012.

[36] P. Valckenaers, H. Van Brussel, O. Bochmann, B. Saint Germain, C. Zamfirescu *et al.*, "On the design of emergent systems: an investigation of integration and interoperability issues," *Engineering applications of artificial intelligence*, vol. 16, no. 4, pp. 377–393, 2003.

[37] K. Doya and E. Uchibe, "The cyber rodent project: Exploration of adaptive mechanisms for self-preservation and self-reproduction," *Adaptive Behavior*, vol. 13, no. 2, pp. 149–160, 2005.

[38] "Standard review plan for the review of safety analysis reports for nuclear power plants: Lwr edition," U.S. Nuclear Regulatory Commission, Chapter 19: Severe Accidents NUREG-0800, 2015.